Springer

Handbook of Nano-technology

Bhushan
Editor

4th Edition

Springer

Springer Handbook of Nanotechnology

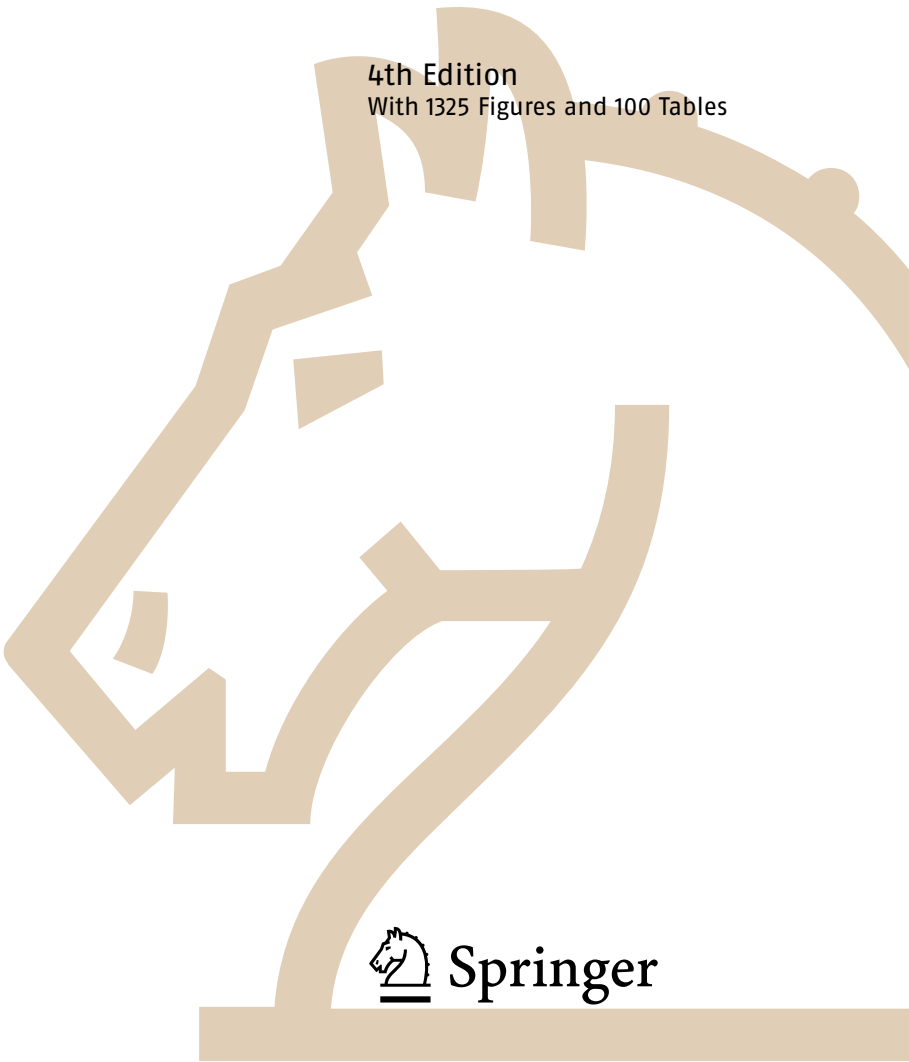
Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk reference book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook of Nanotechnology

Bharat Bhushan (Ed.)

4th Edition
With 1325 Figures and 100 Tables



Springer

Editor
Bharat Bhushan
The Ohio State University
Nanoprobe Lab for Bio- & Nanotechnology & Biomimetics
201 W. 19th Ave.
Columbus, OH, 43210-1107, USA
bhushan.2@osu.edu

ISBN: 978-3-662-54355-9 e-ISBN: 978-3-662-54357-3
DOI 10.1007/978-3-662-54357-3
Library of Congress Control Number: 2017954360

© Springer-Verlag GmbH Germany 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Production, typesetting and illustrations: le-tex publishing services GmbH, Leipzig
Typography and layout: schreiberVIS, Seeheim
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg
Printing and binding: Printer Trento s.r.l., Trento

Printed on acid free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH, DE
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Foreword

When President W. J. Clinton approved the National Nanotechnology Initiative in 2000, we could not really imagine its full impact on science, technology and the economy. Nanotechnology refers to technology at the nanoscale that has applications in the real world. It is a broad and highly interdisciplinary field that is still evolving. The field involves the discovery and development of novel materials, processes, and phenomena at the nanometer scale, as well as the development of new experimental and theoretical techniques for research. These discoveries provide the foundations for further development of innovative nanosystems and nanostructured materials. There is an increasing need for a multidisciplinary, systems-oriented approach to manufacturing micro/nanodevices that function reliably. This can only be achieved through the cross-fertilization of ideas from different disciplines and the systematic flow of information and people among research groups.

To encourage this multidisciplinary approach, Professor Bharat Bhushan has harnessed his own knowledge and experience, gained in several industries and universities, to assemble internationally recognized academic and industry authorities from around the world to write chapters for this handbook. The topics include the latest major advances in many fields where nanoscale science and engineering are being pursued, and illustrate how the field of nanotechnology has continued to emerge and blossom.

Covering even the most important aspects of nanotechnology in a single book that reaches readers ranging from students to active senior researchers in academia and industry is an enormous challenge. Adding to this challenge is the accelerating pace of discovery and applications in nanotechnology. However, as in earlier editions, Professor Bhushan does an admirable job bringing together disparate topics and areas of study and practice to create a cohesive and wide-ranging handbook, sure to serve as the bedrock for anyone interested in the field.

Professor Bhushan's comprehensive book is intended to serve as both a textbook for university courses and a reference for researchers. It is a testament to the quality of the handbook contents that continuing editions have been created. The three editions were timely additions to the literature on nanotechnology, and stimulated further interest in this important new field while serving as invaluable resources to members of the international scientific and industrial community. The increasing demand for up-to-date information on this fast moving field led to the third edition. It is increasingly important that scientists and engineers, whatever their specialty, have a solid grounding in the fundamentals and potential applications of nanotechnology. This fourth edition addresses that need by paying particular attention to the widening audience of readers, discussing nanotechnology and nanoscience education, and providing timely and useful information on new and burgeoning nanotechnology areas. It also includes a discussion of the societal, ethical, and political issues that tend to surround emerging technology.

The Editor-in-Chief and his team are to be warmly congratulated for bringing together this exclusive, opportune, and useful *Springer Handbook of Nanotechnology*.



Prof. Neal Lane

Malcolm Gillis University
Professor Emeritus
Department of Physics and Astronomy Emeritus
Senior Fellow, Rice University's Baker Institute for Public Policy
Rice University
Houston, Texas
September 2016

Prof. Lane served in the Clinton Administration as Assistant to the President for Science and Technology and Director of the White House Office of Science and Technology Policy (1998–2001) and, prior to that, as Director of the National Science Foundation (1993–1998). While at the White House, he was instrumental in creating the National Nanotechnology Initiative (NNI).

Preface for the Fourth Edition

On 29 December 1959, at the California Institute of Technology, Nobel Laureate *Richard P. Feynman* gave a talk at the Annual Meeting of the American Physical Society titled *There's Plenty of Room at the Bottom*, which has become one of the classic science lectures of the 20th century. He presented a technological vision of extreme miniaturization in 1959, several years before the word *chip* became part of the lexicon. He talked about the problem of manipulating and controlling things on a small scale. Extrapolating from known physical laws, Feynman envisioned a technology using the ultimate toolbox of nature: building nano-objects atom by atom or molecule by molecule. Since the 1980s, the many inventions and discoveries in the fabrication of nano-objects have been testaments to his vision.

In a January 2000 speech, also at the California Institute of Technology, President William J. Clinton talked about the exciting promise of nanotechnology and the importance of expanding research in nanoscale science and technology more broadly. Later that month, in the State of the Union Address, he announced an ambitious $ 497 million federal, multiagency National Nanotechnology Initiative (NNI) in the 2001 fiscal year (FY) budget, and made the NNI a top science and technology priority. The objective of this initiative was to form a broad-based coalition in which academia, the private sector, and local, state, and federal governments worked together to push the envelope of nanoscience and nanoengineering and to reap nanotechnology's potential social and economic benefits.

Nanotechnology refers to technology done at the nanoscale that has applications in the real world. The unique physical and chemical properties of nanomaterials can be exploited for applications that benefit society. The National Nanotechnology Initiative was formalized with the 21st Century Nanotechnology Research and Development Act in 2003. Since 2001, the US Congress has appropriated about $ 20 billion for nanotechnology research and development (R&D) through FY 2015, with about $ 1.5 billion spent annually. US private sector R&D funding is more than double that of federal and state funding with a focus on translating fundamental research into commercial products. The revenues from nanoenabled products continue growing, with over $ 200B in FY 2012 in the US alone, and over $ 700B worldwide. This represents an impressive return on investment.

The integration of nanotechnology into larger systems gives us breakthrough solutions to many current environmental, medical, and industrial problems, including smart materials, nanomanufacturing, electronics, targeted drug delivery, energy, water, biotechnology, information technology, and national security. Nanotechnology will have a profound impact on our economy and society; it is a modern industrial revolution. Nanotechnology represents a megatrend and has become a general purpose technology. It is now a household word and a buzzword often used in marketing products.

There is an increasing need for a multidisciplinary, system-oriented approach to design and manufacturing of micro/nanodevices that function reliably. This can only be achieved through the cross-fertilization of ideas from different disciplines and the systematic flow of information and people among research groups. Reliability is a critical technology for many micro- and nanosystems and nanostructured materials. A broad-based handbook was needed, and so the first edition of the *Springer Handbook of Nanotechnology* was published in April 2004. It presented an overview of nanomaterial synthesis, micro/nanofabrication, micro- and nanocomponents and systems, scanning probe microscopy, reliability issues for nanotechnology, and industrial applications. When the handbook went on sale in Europe, it was sold out in ten days. Reviews on the handbook were very flattering.

Given the explosive growth in nanoscience and nanotechnology, the publisher and the editor decided to develop a second edition only six months after the publication of the first edition. The second edition (2007) came out in December 2006. The publisher and the editor again decided to develop a third edition six months after the publication of the second edition, which came out in 2010.

There continue to be significant developments in bio- and nanotechnology, and many products have been commercialized. Now, politicians, investors, and industrialists can see return on their R&D investments. Societal, ethical, and political effects as well as environmental and health and safety issues need to be addressed. As nanoproducts become real and interest in this exploding field grows, there emerges a need to develop nanoscience and nanotechnology curricula and textbooks that can be used in academia. The need exists starting from K-12 through college and continuing education, as well as technical training. The present

fourth edition was launched after five years after the publication of the third editiom. The objective of this edition is to cover the latest developments and introduce new areas of nanoscience and nanotechnology. This edition integrates knowledge from nanostructures, nanofabrication, materials science, nanodevices, and reliability. It covers various industrial applications. It also addresses societal, ethical, and political issues, as well as environmental, health and safety aspects, and nanotechnology education. Given the significant interest in biomedical applications and biomimetics, a number of additional chapters in this arena were added.

Almost all chapters are either new or substantially revised with coverage of new areas. The fourth edition consists of 47 chapters. The chapters have been written by over 130 internationally recognized experts in the field, from academia, national research labs and industry from all over the world.

This handbook is intended for three types of readers: graduate students of nanotechnology, researchers in academia and industry who are active or intend to become active in this field, and practicing engineers and scientists who have encountered a problem and hope to solve it as expeditiously as possible. The handbook should serve as an excellent text for one- or two-semester graduate courses in nanotechnology in mechanical engineering, materials science, applied physics, or applied chemistry.

I wish to sincerely thank the authors for offering to write comprehensive chapters on a tight schedule. This is generally an added responsibility in the hectic work schedules of researchers today. I depended on a large number of reviewers who provided critical reviews. Last but not the least, I would like to thank my assistant Renée L. Ripley for various administrative duties; her tireless efforts are highly appreciated.

I hope that this handbook will stimulate further interest in this important new field, and that readers of this handbook will find it useful.

Bharat Bhushan
Editor
September 2017

About the Editor



Dr. Bharat Bhushan received an M.S. in mechanical engineering from the Massachusetts Institute of Technology in 1971, an M.S. in mechanics and a Ph.D. in mechanical engineering from the University of Colorado at Boulder in 1973 and 1976 respectively, an MBA from Rensselaer Polytechnic Institute at Troy, NY in 1980, Doctor Technicae from the University of Trondheim at Trondheim, Norway in 1990, a Doctor of Technical Sciences from the Warsaw University of Technology at Warsaw, Poland in 1996, and Doctor Honouris Causa from the National Academy of Sciences at Gomel, Belarus in 2000 and University of Kragujevac, Serbia in 2011. He is a registered professional engineer. He is presently an Ohio Eminent Scholar, The Howard D. Winbigler Professor, and the Director of the Nanoprobe Laboratory for Bio- & Nanotechnology and Biomimetics (NLB2) in the College of Engineering, and affiliated faculty in the John Glenn College of Public Affairs, at the Ohio State University, Columbus, Ohio. In 2013–14, he served as an ASME/AAAS (American Society of Mechanical Engineers/American Association for the Advancement of Science) Science & Technology Policy Fellow, House Committee on Science, Space & Technology, United States Congress, Washington, DC. His research interests include fundamental studies with a focus on scanning probe techniques in the interdisciplinary areas of bio/nanotribology, bio/nanomechanics and bio/nanomaterials characterization and applications in bio/nanotechnology and biomimetics. He is an internationally recognized expert in bio/nanotribology and bio/nanomechanics using scanning probe microscopy, and is one of the most prolific authors on the topic. He is considered by some a pioneer of the tribology and mechanics of magnetic storage devices. He has authored eight scientific books, 90+ handbook chapters, 800+ scientific papers (One of Google Scholar's 1612 highly cited researchers (h > 100) and h-index 110+ with 60k+ citations; Web of Science h-index 857+; ISI Highly Cited Researcher in Materials Science since 2007 and in Biology and Biochemistry in 2013; ISI Top 5% Cited Authors for Journals in Chemistry, 2011), and 60+ technical reports. He has also edited 50+ books and holds 20 US and foreign patents. He is coeditor of Springer NanoScience and Technology Series and also of Microsystem Technologies, and is a PNAS Editorial Board member. He has given more than 400 invited presentations on six continents and more than 200 keynote/plenary addresses at major international conferences.

Dr. Bhushan is an accomplished organizer. He organized the first Symposium on Tribology and Mechanics of Magnetic Storage Systems in 1984 and the first International Symposium on Advances in Information Storage Systems in 1990, both of which are now held annually. He organized two international NATO institutes in Europe. He is the founder of an ASME Information Storage and Processing Systems Division founded in 1993 and served as the founding chair from 1993–1998. His biography has been listed in over two dozen Who's Who books including Who's Who in the World. He has received more than two dozen awards for his contributions to science and technology from professional societies, industry, and US government agencies including the Life Achievement Tribology Award and the Institution of Chemical Engineers (UK) Global Award. His research was listed in the top ten science stories of 2015. He is also the recipient of various international fellowships including the Alexander von Humboldt Research Prize for Senior Scientists, the Max Planck Foundation Research Award for Outstanding Foreign Scientists, and the Fulbright Senior Scholar Award. He is a foreign member of the International Academy of Engineering (Russia), the Byelorussian Academy of Engineering and Technology and the Academy of Triboengineering of Ukraine, an honorary member of the Society of Tribologists of Belarus and STLE (Society of Tribologists and Lubrication Engineers), a fellow of ASME, IEEE (Institute of Electrical and Electronics Engineers) and the New York Academy of Sciences, and a member of ASEE (American Society for Engineering Education), Sigma Xi and Tau Beta Pi.

Dr. Bhushan has previously worked for Mechanical Technology Inc., Latham, NY; SKF Industries Inc., King of Prussia, PA; IBM, Tucson, AZ; and IBM Almaden Research Center, San Jose, CA. He has held visiting professorships at University of California at Berkeley, USA; University of Cambridge, UK; Technical University Vienna, Austria; University of Paris, Orsay, France; ETH Zurich, Switzerland; EPFL Lausanne, Switzerland; University of Southampton, UK; University of Kragujevac, Serbia; Tsinghua University; China, Harbin Institute, China; and KFUPM, Saudi Arabia.

https://nlbb.engineering.osu.edu/

List of Authors

Aditya Aryasomayajula
McMaster University
Dept. of Mechanical Engineering
1280 Main St. West
Hamilton, L8S 4L7, Canada
adityaarya83@gmail.com,
aryasoma@mcmaster.ca

W. Robert Ashurst
Auburn University
Dept. of Chemical Engineering
212 Ross Hall
Auburn, AL 36849-5127, USA
ashurwr@auburn.edu

Massood Z. Atashbar
Western Michigan University
Dept. of Electrical & Computer Engineering
4601 Campus Dr.
Kalamazoo, MI 49008-5329, USA
massood.atashbar@wmich.edu

Wolfgang Bacsa
Centre National de la Recherche Scientifique,
Université de Toulouse
Centre d'Elaboration de Matériaux et d'Etudes
Structurales (CEMES), UPR #8011
29 Rue Jeanne Marvig (BP 94347)
31055 Toulouse, France
wolfgang.bacsa@cemes.fr

William S. Bainbridge
National Science Foundation
Division of Information & Intelligent Systems
4201 Wilson Blvd., Suite 525
Arlington, VA 22230, USA
wbainbri@nsf.gov

Antoni Baldi
Barcelona Microelectronics Institute
(IMB-CNM-CSIC)
Chemical Transducers Group
Carrer dels Til.lers s/n, Campus UAB
08193 Bellaterra, Spain
antoni.baldi@imb-cnm.csic.es

Wilhelm Barthlott
University of Bonn
Nees-Institute for Biodiversity of Plants
Venusbergweg 22
53115 Bonn, Germany
barthlott@uni-bonn.de

Pouriya Bayat
York University
Dept. of Mechanical Engineering
4700 Keele St., BRG 433B
Toronto, M3J 1P3, Canada
pouriya@yorku.ca

Mehmet Z. Baykara
Bilkent University
Dept. of Mechanical Engineering & UNAM
EA 110
06800 Ankara, Turkey
mehmet.baykara@bilkent.edu.tr

Bharat Bhushan
The Ohio State University
Nanoprobe Lab for Bio- & Nanotechnology &
Biomimetics
201 W. 19th Ave.
Columbus, OH 43210-1107, USA
bhushan.2@osu.edu

Marcie R. Black
Advanced Silicon Group
173 Bedford Rd.
Lincoln, MA 01773, USA
mrkarty@yahoo.com

Maarten P. de Boer
Carnegie Mellon University
Dept. of Mechanical Engineering
5000 Forbes Ave., Scaife Hall 325
Pittsburgh, PA 15213, USA
mpdebo@andrew.cmu.edu

Donald W. Brenner
North Carolina State University
Dept. of Materials Science & Engineering
911 Partner's Way
Raleigh, NC 27695-7907, USA
brenner@ncsu.edu

Jean-Marc Broto
Centre National de la Recherche Scientifique,
Institut National des Sciences Appliquées de
Toulouse, Université de Toulouse
Laboratoire National des Champs Magnétiques
Intenses de Toulouse (LNCMI-T), UPR #3228
143 Avenue de Rangueil, UPR #3228 CNRS / UPS /
INSA
31400 Toulouse, France
jean-marc.broto@univ-tlse3.fr

Patrizio Candeloro
University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Viale Europa, loc. Germaneto
88100 Catanzaro, Italy
patrizio.candeloro@unicz.it

Brigitte Caussat
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire de Génie Chimique (LGC), UMR #5503
4 allée Emile Monso (BP 84234)
31432 Toulouse, France
brigitte.caussat@ensiacet.fr

Chunying Chen
National Center for Nanoscience & Technology of China
CAS Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety
No. 11, Beiyitiao, Zhongguancun
100190 Beijing, China
chenchy@nanoctr.cn

Rui Chen
National Center for Nanoscience & Technology of China
CAS Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety
No. 11, Beiyitiao, Zhongguancun
100190 Beijing, China
chenr@nanoctr.cn

Yu-Ting Cheng
National Chiao Tung University
Dept. of Electronics Engineering
1001 University Rd.
30010 Hsinchu, Taiwan
ytcheng@mail.nctu.edu.tw

Mu Chiao
University of British Columbia
Dept. of Mechanical Engineering
2054-6250 Applied Science Lane
Vancouver, V6T 1Z4, Canada
muchiao@mech.ubc.ca

Ciro Chiappini
King's College London
Craniofacial Development and Stem Cell Biology
Guy's Hospital, 27th Floor Tower Wing
London, SE1 9RT, UK
ciro.chiappini@kcl.ac.uk

Lilia A. Chtcheglova
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
lilia.chtcheglova@jku.at

Alessandro Coclite
Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Via Morego 30
16163 Genova, Italy
alessandro.coclite@iit.it

Maria Laura Coluccio
University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Viale Europa, loc. Germaneto
88100 Catanzaro, Italy
coluccio@unicz.it

Paolo Decuzzi
Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Via Morego 30
16163 Genova, Italy
paolo.decuzzi@iit.it

Frank W. DelRio
National Institute of Standards & Technology
Material Measurement Laboratory
325 Broadway St.
Boulder, CO 80305, USA
frank.delrio@nist.gov

Enzo Di Fabrizio
King Abdullah University of Science and Technology
Dept. of Physical Science and Engineering
Bldg. 2 (Ibn Al-Haytham) 4th floor, Office #4220
Red Sea side
23955-6900 Thuwal, Saudi Arabia
enzo.difabrizio@kaust.edu.sa

Daniele Di Mascolo
Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Via Morego 30
16163 Genova, Italy
daniele.dimascolo@iit.it

Lixin Dong
Michigan State University
Dept. of Electrical & Computer Engineering
2120 Engineering Building
East Lansing, MI 48824, USA
ldong@egr.msu.edu

Mildred S. Dresselhaus (deceased)

Andreas Ebner
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
andreas.ebner@jku.at

Toshiaki Enoki
Tokyo Institute of Technology
Dept. of Chemistry
2-12-1 Ookayama, Meguro-ku
152-8551 Tokyo, Japan
tenoki@chem.titech.ac.jp

Haig Alexander Eskandarian
Swiss Federal Institute of Technology Lausanne (EPFL)
Dept. of Bioengineering
Station 17, BM3109
1015 Lausanne, Switzerland
alex.eskandarian@epfl.ch

Georg E. Fantner
Swiss Federal Institute of Technology Lausanne (EPFL)
Dept. of Bioengineering
Station 17, BM3109
1015 Lausanne, Switzerland
georg.fantner@epfl.ch

Philip X.-L. Feng
Case Western Reserve University
Dept. of Electrical Engineering & Computer Science
10900 Euclid Ave., Glennan Building 713
Cleveland, OH 44106, USA
philip.feng@case.edu

Emmanuel Flahaut
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
118 route de Narbonne
31062 Toulouse, France
flahaut@chimie.ups-tlse.fr

Francesco Gentile
University Naples Federico II
Dept. of Electrical Engineering and Information Technology
Via Claudio 21
80125 Naples, Italy
francesco.gentile77@gmail.com

Christoph Gerber
University of Basel
Dept. of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
christoph.gerber@unibas.ch

Thilo Glatzel
University of Basel
Dept.of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
thilo.glatzel@unibas.ch

Enrico Gnecco
Friedrich Schiller University Jena
Otto Schott Institute of Materials Research
Löbdergraben 32
07743 Jena, Germany
enrico.gnecco@uni-jena.de

Miroslav Haluška
ETH Zürich
Dept. of Mechanical & Process Engineering
Tannenstr. 3
8092 Zurich, Switzerland
haluska@micro.mavt.ethz.ch

Judith A. Harrison
U.S. Naval Academy
Dept. of Chemistry
572 Holloway Rd., MS 9B
Annapolis, MD 21402, USA
jah@usna.edu

Martin Hegner
Trinity College Dublin
School of Physics, CRANN
Pearse St.
D2, Dublin, Ireland
martin.hegner@tcd.ie

Thomas Helbling
greenTEG AG
Technoparkstr. 1
8005 Zurich, Switzerland
thomas.helbling@greenteg.com

Seong-Jun Heo
Lam Research Corp.
4650 Cushing Parkway
Fremont, CA 94538, USA
heogyver@gmail.com

Barbara Herr Harthorn
University of California Santa Barbara
Dept. of Anthropology
HSSB 2020 MC:3210
Santa Barbara, CA 93106-3210, USA
barbara.harthorn@ucsb.edu

Christofer Hierold
ETH Zürich
Dept. of Mechanical & Process Engineering
Tannenstr. 3
8092 Zurich, Switzerland
hierold@micro.mavt.ethz.ch

Peter Hinterdorfer
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
peter.hinterdorfer@jku.at

Maria Iliut
University of Manchester
School of Materials
The Mill, Sackville St.
Manchester, M1 3BB, UK
maria.iliut@manchester.ac.uk

Douglas L. Irving
North Carolina State University
Dept. of Materials Science & Engineering
911 Partner's Way
Raleigh, NC 27695, USA
dlirving@ncsu.edu

Jacob N. Israelachvili
University of California Santa Barbara
Dept. of Chemical Engineering & Materials Dept.
Santa Barbara, CA 93106, USA
jacob@engineering.ucsb.edu

Esmaiel Jabbari
University of South Carolina
Dept. of Chemical Engineering
301 South Main St., Room 2C11
Columbia, SC 29208, USA
jabbari@cec.sc.edu

Anne Jourdain
IMEC
Kapeldreef. 75
3001 Leuven, Belgium
jourdain@imec.be

Harold Kahn
410 Altamont St.
Charlottesville, VA 22902, USA
hxkahn@gmail.com

Gopakumar Kamalakshakurup
University of California Irvine
Dept. of Biomedical Engineering
3120 Natural Sciences II
Irvine, CA 92697-2715, USA
gkamalak@uci.edu

Roger D. Kamm
Massachusetts Institute of Technology
Depts. of Mechanical Engineering & Biological Engineering
77 Massachusetts Ave., Room NE47-321
Cambridge, MA 02139, USA
rdkamm@mit.edu

Josef A. Käs
University of Leipzig
Peter Debye Institute
Linnéstr. 5
04103 Leipzig, Germany
jkaes@uni-leipzig.de

Jongbaeg Kim
Yonsei University
School of Mechanical Engineering
50 Yonsei-ro, Seodaemun-gu
120-749 Seoul, Republic of Korea
kimjb@yonsei.ac.kr

Marcin Kisiel
University of Basel
Dept. of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
marcin.kisiel@unibas.ch

Kerstin Koch
Rhine-Waal University of Applied Science
Dept. of Life Science, Biology & Nanobiotechnology
Marie-Curie-Str. 1, Gebäude 11, 2. Stock, Raum 17
47533 Kleve, Germany
koch@hochschule-rhein-waal.de

Melanie Koehler
Université Catholique de Louvain
Institute of Life Sciences
L7.07.06 (Carnoy C 461), Coix du Sud 4–5
1348 Louvain-la-Neuve, Belgium
melanie.kohler@uclouvain.be

Anders Kristensen
Technical University of Denmark
DTU Nanotech
Ørsteds Plads 345E
2800 Kongens Lyngby, Denmark
anders.kristensen@nanotech.dtu.dk

Jan Lammerding
Cornell University
Meinig School of Biomedical Engineering
Weill Institute for Cell & Molecular Biology, Weill Hall, Room 235
Ithaca, NY 14853, USA
jan.lammerding@cornell.edu

Constanze Lamprecht
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
constanze.lamprecht@gmx.de

Hans Peter Lang
University of Basel
Dept. of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
hans-peter.lang@unibas.ch

Christophe Laurent
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
118 route de Narbonne
31062 Toulouse, France
laurent@chimie.ups-tlse.fr

Abraham Lee
University of California Irvine
Dept. of Biomedical Engineering
3120 Natural Sciences II
Irvine, CA 92697, USA
aplee@uci.edu

Aeju Lee
Kumamoto University
International Research Organization for Advanced Science and Technology
2-31-1 Kurokami, Chuo-ku
860-8555 Kumamoto, Japan
aeju-lee@kumamoto-u.ac.jp

Dong Woog Lee
Ulsan National Institute of Science & Technology
School of Energy & Chemical Engineering
50, UNIST-gil
44919 Ulsan, Republic of Korea
dongwoog.lee@unist.ac.kr

Stephen C. Lee
The Ohio State University
Dept. of Biomedical Engineering
023 Hamilton Hall
Columbus, OH 43210, USA
lee.1996@osu.edu

Liwei Lin
University of California Berkeley
Dept. of Mechanical Engineering
5135 Etcheverry Hall
Berkely, CA 94720, USA
lwlin@me.berkeley.edu

Matthias Mail
University of Bonn
Nees-Institute for Biodiversity of Plants
Venusbergweg 22
53115 Bonn, Germany
matthias.mail@uni-bonn.de

Othmar Marti
Ulm University
Institute of Experimental Physics
Albert-Einstein-Allee 11
89081 Ulm, Germany
othmar.marti@uni-ulm.de

Shinji Matsui
University of Hyogo
Laboratory of Advanced Science & Technology for Industry
3-1-2 Koto, Kamigori, Ako
678-1205 Hyogo, Japan
matsui@lasti.u-hyogo.ac.jp

Etienne Menard
HeliosLite
Batiment Lama, 17 avenue du Lac Leman, Savoie Technolac
73370 Le Bourget du Lac, France
etienne.menard@helioslite.com

David Mesguich
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
118 route de Narbonne
31062 Toulouse, France
mesguich@chimie.ups-tlse.fr

Vincent Meunier
Rensselaer Polytechnic Institute
Dept. of Physics, Applied Physics, & Astronomy
110, 8th St.
Troy, NY 12180, USA
meuniv@rpi.edu

Ernst Meyer
University of Basel
Dept. of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
ernst.meyer@unibas.ch

Tom D. Milster
University of Arizona
College of Optical Sciences
1630 East University Blvd.
Tucson, AZ 85721, USA
milster@optics.arizona.edu

Hiroaki Misawa
Hokkaido University
Research Institute of Electronic Science
N21W10, Kita-Ward, Northern Campus Sousei Building, Room #04 303
001-0021 Sapporo, Japan
misawa@es.hokudai.ac.jp

Robert Modliński
EPCOS AG
Anzinger Str. 13
81671 Munich, Germany
robert.modlinski@epcos.com

Seyedsina Moeinzadeh
University of South Carolina
Dept. of Chemical Engineering
301 South Main St., Room 2B46
Columbia, SC 29208, USA
moeinzad@email.sc.edu

Mohammad R. K. Mofrad
University of California Berkeley
Depts. of Bioengineering & Mechanical Engineering
208A Stanley Hall, #1762
Berkeley, CA 94720, USA
mofrad@berkeley.edu

Marc Monthioux
Centre National de la Recherche Scientifique, Université de Toulouse
Centre d'Elaboration de Matériaux et d'Etudes Structurales (CEMES), UPR #8011
29 Rue Jeanne Marvig (BP 94347)
31055 Toulouse, France
marc.monthioux@cemes.fr

Markus Morgenstern
RWTH Aachen University
II. Institute of Physics B & JARA-FIT
Otto-Blumentahl-Str. 28
52074 Aachen, Germany
mmorgens@physik.rwth-aachen.de

Bradley J. Nelson
Swiss Federal Institute of Technology
Institute for Robotics & Intelligent Systems, ETH Zentrum, CLA H 1.1
Tannenstr. 3
8092 Zurich, Switzerland
bnelson@ethz.ch

T. Kieran Nunan
Kieran Nunan Consulting
854 Maple St.
Carlisle, MA 01741, USA
tknunan25@gmail.com

Pascal D. Odermatt
Swiss Federal Institute of Technology Lausanne (EPFL)
Dept. of Bioengineering
Station 17, BM3109
1015 Lausanne, Switzerland
pascal.odermatt@epfl.ch

Yoojin Oh
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
yoo_jin.oh@jku.at

Anna Lisa Palange
Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Via Morego 30
16163 Genova, Italy
annalisa.palange@iit.it

Rémy Pawlak
University of Basel
Dept. of Physics
Klingelbergstr. 82
4056 Basel, Switzerland
remy.pawlak@unibas.ch

Alain Peigney
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
118 route de Narbonne
31062 Toulouse, France
peigney@chimie.ups-tlse.fr

Gerardo Perozziello
University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Viale Europa, loc. Germaneto
88100 Catanzaro, Italy
gerardo.perozziello@unicz.it

Sandra Posch
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
sandra.posch@jku.at

Robert Puers
KU Leuven
ESAT/MICAS
Kasteelpark, Arenberg 10, Box 2443
3001 Leuven, Belgium
robert.puers@kuleuven.be

Oded Rabin
University of Maryland
Dept. of Materials Science & Engineering
1110B Chem Nuc Eng Bldg 090
College Park, MD 20742, USA
oded@umd.edu

Françisco M. Raymo
University of Miami
Dept. of Chemistry
1301 Memorial Drive
Coral Gables, FL 33146-0431, USA
fraymo@miami.edu

Manitra Razafinimanana
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire des Plasmas et de Conversion de l'Energie (LAPLACE), UMR #5002
118 route de Narbonne
31062 Toulouse, France
razafinimanana@laplace.univ-tlse.fr

Pouya Rezai
York University
Dept. of Mechanical Engineering
4700 Keele St., BRG 433B
Toronto, M3J 1P3, Canada
pouya.rezai@lassonde.yorku.ca,
prezai@yorku.ca

Gemma Rius
Barcelona Microelectronics Institute (IMB-CNM-CSIC)
NEMS & Nanofabrication Group
Carrer dels Til.lers s/n, Campus UAB
08193 Bellaterra, Spain
gemma.rius@imb-cnm.csic.es

John A. Rogers
Northwestern University
Dept. of Materials Science & Engineering
2145 Sheridan Road
Evanston, IL 60208-3109, USA
jrogers@northwestern.edu

Cosmin I. Roman
ETH Zürich
Dept. of Mechanical & Process Engineering
Tannenstr. 3
8092 Zurich, Switzerland
cosmin.roman@micro.mavt.ethz.ch

Marina Ruths
University of Massachusetts Lowell
Dept. of Chemistry
1 University Ave.
Lowell, MA 01854, USA
marina_ruths@uml.edu

Hélder A. Santos
University of Helsinki
Division of Pharmaceutical Chemistry and Technology
Viikinkaari 5E
Helsinki, 00790, Finland
helder.santos@helsinki.fi

Shintaro Sato
Fujitsu Laboratories Ltd.
Devices & Materials Laboratories
10-1 Morinosato-Wakamiya
243-0197 Atsugi, Kanagawa, Japan
sato.shintaro@jp.fujitsu.com

Helmut Schift
Paul Scherrer Institute
Laboratory for Micro- & Nanotechnology
ODRA/103
5232 Villigen PSI, Switzerland
helmut.schift@psi.ch

Jörg Schnauß
University of Leipzig
Peter Debye Institute
Linnéstr. 5
04103 Leipzig, Germany
joerg.schnauss@uni-leipzig.de

Alexander Schwarz
University of Hamburg
Institute of Nanostructure and Solid State Physics
Jungiusstr. 11
20355 Hamburg, Germany
aschwarz@physnet.uni-hamburg.de

Udo D. Schwarz
Yale University
Dept. of Mechanical Engineering & Materials Science
15 Prospect St.
New Haven, CT 06511, USA
udo.schwarz@yale.edu

P. Ravi Selvaganapathy
McMaster University
Dept. of Mechanical Engineering
1280 Main St. West
Hamilton, L8S 4L7, Canada
selvaga@mcmaster.ca

Philippe Serp
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire de Chimie de Coordination (LCC), UPR #8241
118 Route de Narbonne
31077 Toulouse, France
philippe.serp@ensiacet.fr

Sameer S. Shroff
Carnegie Mellon University
Dept. of Mechanical Engineering
5000 Forbes Ave.
Pittsburgh, PA 15213, USA
sshroff@alumni.cmu.edu

Mark G. da Silva
Analog Devices Inc.
High Performance Sensors
804 Woburn St. MS-112
Wilmington, MA 01887, USA
mark.dasilva@analog.com

Susan B. Sinnott
The Pennsylvania State University
Dept. of Materials Science & Engineering
221A Steidle Bldg., 1 Pollock Rd.
University Park, PA 16801, USA
sbs5563@psu.edu

David M. Smith
Fraunhofer Institute for Cell Therapy & Immunology (IZI)
DNA Nanodevices Group
Perlickstr. 1
04103 Leipzig, Germany
david.smith@izi.fraunhofer.de

W. Merlijn van Spengen
Falco Systems
Van Boshuizenstr. 12
1083BA, Amsterdam, The Netherlands
w.m.vanspengen@tudelft.nl

Jürgen Strasser
Center for Advanced Bioanalysis GmbH
Gruberstr. 40
4020 Linz, Austria
juergen.strasser@cbl.at

Yu-Chuan Su
National Tsing Hua University
Dept. of Engineering & System Science
101, Section 2, Kuang-Fu Rd.
30013 Hsinchu, Taiwan
ycsu@ess.nthu.edu.tw

Quan Sun
Hokkaido University
Research Institute of Electronic Science
N21W10, Kita-Ward, Northern Campus Sousei Building, Room #02 304
001-0021 Sapporo, Japan
quansun@es.hokudai.ac.jp

Flavien Valensi
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire des Plasmas et de Conversion de l'Energie (LAPLACE), UMR #5002
118 route de Narbonne
31062 Toulouse, France
valensi@laplace.univ-tlse.fr

Derek Vallejo
University of California Irvine
Dept. of Biomedical Engineering
3120 Natural Sciences II
Irvine, CA 92697-2715, USA
dhvallej@gmail.com

Aravind Vijayaraghavan
University of Manchester
School of Materials
The Mill, Sackville St.
Manchester, M1 3BB, UK
aravind@manchester.ac.uk

Alicia Weibel
Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
118 route de Narbonne
31062 Toulouse, France
weibel@chimie.ups-tlse.fr

Kurt Winkelmann
Florida Institute of Technology
Dept. of Chemistry
150 W. University Blvd., 127 Physical Sciences Bldg.
Melbourne, FL 32901, USA
kwinkel@fit.edu

Darrin J. Young
University of Utah
Dept. of Electrical & Computer Engineering
50 S. Central Campus Dr. Rm 3280, Joseph Merrill Engineering Bldg.
Salt Lake City, UT 84112, USA
darrin.young@utah.edu

Rong Zhu
Johannes Kepler University Linz
Institute of Biophysics
Gruberstr. 40
4020 Linz, Austria
rong.zhu@jku.at

Babak Ziaie
Purdue University
School of Electrical & Computer Engineering
1205 W. State St., Birck Nanotechnology Center
West Lafayette, IN 47907, USA
bziaie@purdue.edu; ziaiebabak@gmail.com

Christian A. Zorman
Case Western Reserve University
Dept. of Electrical Engineering & Computer Science
2123 Martin Luther King Jr. Blvd., 715A Glennan Building
Cleveland, OH 44106, USA
caz@case.edu

Contents

List of Abbreviations XXXIII

1 **Introduction to Nanotechnology**
Bharat Bhushan 1
1.1 Nanotechnology – Definitions and Examples 2
1.2 History and Early Research Expenditures 6
1.3 Governance of the National Nanotechnology Initiative 7
1.4 Nanotechnology R&D Funding Since 2001 8
1.5 Worldwide R&D Investments and Output 11
1.6 Translation of Basic Research to Technology Commercialization 15
1.7 Nanoscience and Nanotechnology Education 16
1.8 Summary and Outlook 17
1.9 Organization of the Handbook 18
References 18

Part A Micro/Nanofabrication Techniques

2 **Molecule-Based Devices**
Françisco M. Raymo 23
2.1 Chemical Approaches to Nanostructured Materials 24
2.2 Molecular Switches and Logic Gates 28
2.3 Solid-State Devices 35
2.4 Conclusions and Outlook 46
References 47

3 **Introduction to Micro-/Nanofabrication**
Gemma Rius, Antoni Baldi, Babak Ziaie, Massood Z. Atashbar 51
3.1 Basic Microfabrication Techniques 52
3.2 MEMS Fabrication Techniques 62
3.3 Nanofabrication Techniques 74
3.4 Summary and Conclusions 82
References 82

4 **3-D Nanostructure Fabrication by Focused-Ion Beam, Electron- and Laser Beam**
Shinji Matsui, Hiroaki Misawa, Quan Sun 87
4.1 Various 3-D Nanostructure Fabrication Techniques 88
4.2 3-D Nanostructure Fabrication by Focused Ion Beam 88
4.3 3-D Nanostructure Fabrication by Electron Beam 97
4.4 3-D Nanostructure Fabrication by Laser 103
References 109

5 **Nanoimprint Lithography**
Helmut Schift, Anders Kristensen 113
5.1 Emerging Nanopatterning Methods 115
5.2 Nanoimprint Process 119
5.3 Tools and Materials for Nanoimprint 125

5.4 Applications 132
5.5 Conclusion and Outlook 137
References 138

6 **Stamping Techniques for Micro- and Nanofabrication**
John A. Rogers, Etienne Menard 143
6.1 High-Resolution Stamps 144
6.2 Microcontact Printing 146
6.3 Nanotransfer Printing 148
6.4 Applications 152
6.5 Conclusions 159
References 159

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems**
Christian A. Zorman 163
7.1 Silicon and its Commonly-Used Derivatives 163
7.2 Germanium-Based Materials 170
7.3 Metals 171
7.4 Semiconductors for Harsh Environment Applications 173
7.5 GaAs, InP and Related III-V Materials 178
7.6 Piezoelectric Materials 180
7.7 Polymer Materials 183
7.8 Future Trends 185
References 185

Part B Nanomaterial and Nanostructures

8 **Carbon Nanotubes**
Marc Monthioux, Philippe Serp, Brigitte Caussat, Emmanuel Flahaut, Manitra Razafinimanana, Flavien Valensi, Christophe Laurent, Alain Peigney, David Mesguich, Alicia Weibel, Wolfgang Bacsa, Jean-Marc Broto 193
8.1 Structure of Carbon Nanotubes 194
8.2 Synthesis of Carbon Nanotubes 199
8.3 Growth Mechanisms of Carbon Nanotubes 210
8.4 Properties of Carbon Nanotubes 213
8.5 Carbon Nanotube-Based Nano-Objects (Carbon Meta-Nanotubes) 217
8.6 Carbon-Nanotube-Containing Materials (Composites) 223
8.7 Current Applications of Carbon Nanotubes (on the Market) 227
8.8 Toxicity and Environmental Impact of Carbon Nanotubes 231
8.9 Concluding Remarks 233
References 233

9 **Nanowires**
Mildred S. Dresselhaus, Marcie R. Black, Vincent Meunier, Oded Rabin 249
9.1 Synthesis 250
9.2 Characterization and Physical Properties of Nanowires 262
9.3 Applications 282
9.4 Concluding Remarks 290
References 290

10 **Nanoribbons**
Toshiaki Enoki, Shintaro Sato 303
10.1 Graphene Nanoribbons 303
10.2 Electronic and Magnetic Properties 305
10.3 Characterizations 310
10.4 Syntheses: Top-Down Methods and Bottom-Up Methods 317
10.5 Potential Applications 324
10.6 Conclusions 330
References 331

11 **Nanoparticles and Their Applications**
Seyedsina Moeinzadeh, Esmaiel Jabbari 335
11.1 Overview 335
11.2 NPs Classification and Synthesis 336
11.3 Properties of NPs 348
11.4 Applications of NPs 351
11.5 Summary 353
References 353

12 **Graphene**
Aravind Vijayaraghavan, Maria Iliut 363
12.1 Methods of Production 365
12.2 Properties 369
12.3 Characterization 372
12.4 Applications 376
12.5 Conclusions and Outlook 385
References 386

Part C MEMS/NEMS

13 **MEMS/NEMS Devices and Applications**
Philip X.-L. Feng, Darrin J. Young, Christian A. Zorman 395
13.1 MEMS Devices and Applications 397
13.2 NEMS Devices and Applications 415
13.3 Challenges and Perspectives 425
References 426

14 **Single-Walled Carbon Nanotube Sensor Concepts**
Cosmin I. Roman, Thomas Helbling, Miroslav Haluška, Christofer Hierold . 431
14.1 Sensor Design Considerations 432
14.2 Sensor Fabrication: SWNT Synthesis and Integration 439
14.3 Summary of State-of-the-Art, Applications Examples 445
14.4 Concluding Remarks 451
References 451

15 **Nanomechanical Cantilever Array Sensors**
Hans Peter Lang, Martin Hegner, Christoph Gerber 457
15.1 Technique 458
15.2 Cantilever Array Sensors 459
15.3 Modes of Operation 460
15.4 Microfabrication 464

15.5 Measurement Setup 465
15.6 Functionalization Techniques 468
15.7 Applications 470
15.8 Conclusions and Outlook 476
References 477

16 **Microfluidic Devices and Their Applications**
Aditya Aryasomayajula, Pouriya Bayat, Pouya Rezai, P. Ravi Selvaganapathy 487
16.1 Preface 487
16.2 Historical Developments 489
16.3 Dominant Phenomenon and Micro–Nanofluidic Design 491
16.4 Fabrication of Micro–Nanofluidic Devices 496
16.5 Applications 504
16.6 Outlook and Future Directions 521
References 522

17 **Microfluidic Micro/Nano Droplets**
Gopakumar Kamalakshakurup, Derek Vallejo, Abraham Lee 537
17.1 Introduction to Micro/Nano Droplet Microfluidic Technologies 537
17.2 Overview of Current Trends in Droplet Microfluidic Technologies 546
17.3 Fundamental Designs and Techniques for Microfluidic Generation of Droplets 547
17.4 Microfluidic Micro-/Nanodroplet Applications 550
17.5 Conclusion 556
References 556

18 **Nanorobotics**
Bradley J. Nelson, Lixin Dong 559
18.1 Overview of Nanorobotics 560
18.2 Actuation at Nanoscales 561
18.3 Nanorobotic Manipulation Systems 563
18.4 Nanorobotic Assembly 568
18.5 Applications 576
References 580

Part D BioMEMS/NEMS

19 **Applications of MEMS to Cell Biology**
Georg E. Fantner, Pascal D. Odermatt, Haig Alexander Eskandarian 587
19.1 Biological Background 588
19.2 High Resolution Microscopy Methods for Live Cell Imaging 592
19.3 Measuring Mechanical Properties on Living Cells 593
19.4 Cancer Detection Using Nanomechanical Sensors 596
19.5 Measurement of Cell Adhesion Using Cantilever Based Force Sensor 599
19.6 Manipulation of Cells Using MEMS Cantilevers 602
19.7 Mass Measurements of Single Cells Using Resonant Cantilevers 603
19.8 Characterization of Antibiotic Action Using MEMS Devices 606
19.9 Conclusions and Outlook 609
References 609

20 **Contact-free Mechanical Manipulation of Biological Materials**
Jörg Schnauß, Josef A. Käs, David M. Smith 617
20.1 Interaction of Laser Light with Biological Material 618
20.2 Optical and Magnetic Tweezers for Single-Molecule Manipulations 621
20.3 Optical and Electric Forces for the Manipulation of Whole Cells 629
20.4 Hydrodynamic Shear Forces for the Manipulation of Single Cells ... 636
20.5 Conclusion and Outlook 638
References 638

21 **Nano-Particles for Biomedical Applications**
Paolo Decuzzi, Alessandro Coclite, Aeju Lee, Anna Lisa Palange, Daniele Di Mascolo, Ciro Chiappini, Hélder A. Santos, Maria Laura Coluccio, Gerardo Perozziello, Patrizio Candeloro, Enzo Di Fabrizio, Francesco Gentile 643
21.1 Overview 644
21.2 Rational Design of Nanoconstructs 646
21.3 Multifunctional Polymeric Nanoconstructs 650
21.4 Sensing and Drug Delivery with Porous Silicon Nanomaterials 655
21.5 Nanomedicine Synthesis by Microfluidics Technology 662
21.6 Electroless Formation of Metal Nanoparticle Aggregates 669
21.7 2-D and 3-D Optical Nanostructures 677
References 681

22 **Biological Molecules in Therapeutic Nanodevices**
Stephen C. Lee, Bharat Bhushan 693
22.1 Definitions and Scope 694
22.2 Assembly Approaches 700
22.3 Sensing Devices 710
22.4 Concluding Remarks: Barriers to Practice 717
References 718

Part E Nanometrology

23 **Scanning Probe Microscopy – Principle of Operation, Instrumentation and Probes**
Bharat Bhushan, Othmar Marti 725
23.1 Scanning Tunneling Microscope 727
23.2 Atomic Force Microscope 731
23.3 AFM Instrumentation and Analyses 747
23.4 Conclusion 763
References 763

24 **Low-Temperature Scanning Probe Microscopy**
Mehmet Z. Baykara, Markus Morgenstern, Alexander Schwarz, Udo D. Schwarz 769
24.1 Microscope Operation at Low Temperatures 771
24.2 Instrumentation 772
24.3 Scanning Tunneling Microscopy and Spectroscopy 773
24.4 Scanning Force Microscopy and Spectroscopy 788
24.5 Summary 799
References 799

25 **Biomedical Sensing with the Atomic Force Microscope**
Constanze Lamprecht, Jürgen Strasser, Melanie Koehler, Sandra Posch, Yoojin Oh, Rong Zhu, Lilia A. Chtcheglova, Andreas Ebner, Peter Hinterdorfer 809
25.1 Topographical Imaging of Biological Samples 810
25.2 Single-Molecule Force Spectroscopy (SMFS) 814
25.3 Simultaneous Topography and Recognition Imaging (TREC) 821
25.4 AFM Biomedical Sensing – Examples 827
25.5 Perspectives and Concluding Remarks 836
References 837

26 **Superresolution Microscopy**
Tom D. Milster 845
26.1 Overview 845
26.2 Scanning Aperture Techniques 849
26.3 4-Pi Microscopy 853
26.4 Enhancement/Depletion Techniques 854
26.5 Photoactivated Localization 856
26.6 Structured Illumination 858
26.7 Light-Sheet Microscopy 862
26.8 Comparison of Techniques for Live-Cell Imaging 863
26.9 Summary 864
References 864

Part F Bio/Nanotribology and Bio/Nanomechanics

27 **Nanotribology, Nanomechanics and Materials Characterization**
Bharat Bhushan 869
27.1 Description of AFM/FFM and Various Measurement Techniques 871
27.2 Surface Imaging, Friction, and Adhesion 882
27.3 Micro/Nanoscale Wear and Scratching, Local Deformation, and Nanofabrication/Nanomachining 906
27.4 Indentation 915
27.5 Boundary Lubrication 919
27.6 Conclusion 928
References 929

28 **Surface Forces and Nanorheology of Molecularly Thin Films**
Dong Woog Lee, Marina Ruths, Jacob N. Israelachvili 935
28.1 Types of Surface Forces 936
28.2 Methods Used to Study Surface Forces 938
28.3 Normal Forces Between Dry (Unlubricated) Surfaces 941
28.4 Normal Forces Between Surfaces in Liquids 942
28.5 Adhesion and Capillary Forces 951
28.6 Introduction: Different Modes of Friction 956
28.7 Relationship Between Adhesion and Friction 957
28.8 Liquid Lubricated Surfaces 966
28.9 Effects of Nanoscale Texture on Friction 976
References 978

29 **Atomic Scale Friction Phenomena**
Enrico Gnecco, Rémy Pawlak, Marcin Kisiel, Thilo Glatzel, Ernst Meyer 987
29.1 Friction Force Microscopy in Selected Environments 988
29.2 The Prandtl–Tomlinson Model 993
29.3 Friction Experiments on the Atomic Scale 996
29.4 Thermal Effects on Atomic Friction 998
29.5 Friction on the Nanometer Scale 999
29.6 Wear on the Atomic Scale 1002
29.7 Noncontact Friction 1003
29.8 Single Molecule Friction 1005
29.9 Conclusion 1008
References 1008

30 **Computer Simulations of Nanometer-Scale Indentation and Friction**
Susan B. Sinnott, Seong-Jun Heo, Donald W. Brenner, Judith A. Harrison, Douglas L. Irving 1013
30.1 Computational Details 1014
30.2 Indentation 1019
30.3 Friction and Lubrication 1033
30.4 Conclusions 1058
References 1058

31 **Cellular Nanomechanics**
Roger D. Kamm, Jan Lammerding, Mohammad R. K. Mofrad 1069
31.1 Overview 1069
31.2 Structural Components of a Cell 1071
31.3 Experimental Methods 1077
31.4 Theoretical and Computational Descriptions 1082
31.5 Mechanics of Subcellular Structures 1086
31.6 Current Understanding and Future Needs 1095
References 1096

32 **Nanomechanical Properties of Nanostructures and Scale Effects**
Bharat Bhushan 1101
32.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures 1103
32.2 Experimental Results and Discussion 1109
32.3 Finite Element Analysis of Nanostructures with Roughness and Scratches 1125
32.4 Summary 1131
32.A Appendix: Fabrication Procedure for the Double-Anchored and Cantilever Beams 1132
References 1134

Part G Molecularly-Thick Films for Lubrication

33 **Nanotribology of Ultrathin and Hard Amorphous Carbon Films**
Bharat Bhushan 1141
33.1 Description of Commonly Used Deposition Techniques 1145
33.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties 1149
33.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques 1154
33.4 Conclusion 1174
References 1175

34 **Self-Assembled Monolayers for Nanotribology and Surface Protection**
Bharat Bhushan 1179
34.1 Background 1179
34.2 A Primer on Organic Chemistry 1182
34.3 Self-Assembled Monolayers: Substrates, Spacer Chains, and End Groups in the Molecular Chains..... 1186
34.4 Contact Angle and Nanotribological Properties of SAMs..... 1189
34.5 Conclusion 1209
References 1210

35 **Nanoscale Boundary Lubrication Studies**
Bharat Bhushan 1215
35.1 Nanodeformation, Molecular Conformation, Spreading, Nanotribological and Electrical Studies, and Environmental Effects of Commonly Used PFPE Lubricant Films 1216
35.2 Nanotribological, Electrical, and Chemical Degradations Studies and Environmental Effects in Novel PFPE Lubricant Films .. 1233
35.3 Nanotribological and Electrical Studies of Ionic Liquid Films..... 1242
35.4 Conclusion 1257
References 1258

Part H Biomimetics and Bioinspired Surfaces

36 **Plant Surfaces: Structures and Functions for Biomimetic Applications**
Wilhelm Barthlott, Matthias Mail, Bharat Bhushan, Kerstin Koch 1265
36.1 500 Million Years of Evolution for Innovative Technologies 1266
36.2 Chemistry of Plant Surfaces..... 1270
36.3 Structuring of Plant Surfaces: Hierarchical Architecture Between Nano- and Macrostructures..... 1279
36.4 Physical Basis of Surface Wetting 1285
36.5 Superhydrophilic and Superhydrophobic Plant Surfaces..... 1286
36.6 Functional Diversity of Plant Surfaces 1289
36.7 Biomimetic Application 1294
36.8 Living Prototypes: Evolution of Plant Surfaces and Biodiversity..... 1297
36.9 Conclusions 1298
References 1298

37 **Bioinspired Nanostructured Antibiofouling and Anti-inorganic Surfaces**
Bharat Bhushan 1307
37.1 Fields Susceptible to Fouling 1307
37.2 Biofouling and Inorganic Fouling Formation Mechanisms 1310
37.3 Antifouling Strategies from Living Nature 1313
37.4 Antifouling: Current Prevention and Cleaning Techniques 1315
37.5 Bioinspired Rice Leaf Surfaces for Antifouling 1318
37.6 Closure 1324
References 1324

Part I Micro/Nanodevice Reliability

38 **MEMS/NEMS and BioMEMS/BioNEMS: Tribology, Mechanics, Materials and Devices**
Bharat Bhushan 1331
38.1 MEMS/NEMS Basics 1332
38.2 Nanotribology and Nanomechanics Issues in MEMS/NEMS and BioMEMS/BioNEMS 1336
38.3 Nanotribology and Nanomechanics Studies of Silicon and Related Materials 1352
38.4 Lubrication Studies for MEMS/NEMS 1359
38.5 Nanoscale Friction, Wear, and Mechanical Behavior of Nano-Objects 1365
38.6 Nanotribological Studies of Biological Molecules on Polystyrene and Silicon Surfaces and Coated Polymer Surfaces 1376
38.7 Trajectory of Submicron Particles for Therapeutics and Diagnostics 1390
38.8 Component-Level Studies 1392
38.9 Conclusion 1403
38.A Appendix: Micro/Nanofabrication Techniques 1404
References 1407

39 **Friction and Wear in Micro- and Nanomachines**
Maarten P. de Boer, Sameer S. Shroff, Frank W. DelRio, W. Robert Ashurst 1417
39.1 From Single- to Multiple-Asperity Friction 1419
39.2 Rate-State Friction 1422
39.3 Putting MEMS Friction to Use: Making Motors and Friction Instruments 1429
39.4 Wear and Tribopolymer Evolution in Micro- and Nanoswitches 1430
39.5 Concluding Remarks 1432
References 1432

40 **Failure Mechanisms in MEMS/NEMS Devices**
W. Merlijn van Spengen, Robert Modliński, Robert Puers, Anne Jourdain 1437
40.1 Failure Modes and Failure Mechanisms 1438
40.2 Stiction and Charge-Related Failure Mechanisms 1439
40.3 Creep, Fatigue, Wear, and Packaging-Related Failures 1445
40.4 Conclusions 1454
References 1454

41 **Mechanical Properties of Micromachined Structures**
Harold Kahn 1459
41.1 Measuring Mechanical Properties of Films on Substrates 1459
41.2 Micromachined Structures for Measuring Mechanical Properties 1460
41.3 Measurements of Mechanical Properties 1471
41.4 Summary 1475
References 1476

42 **High Volume Manufacturing and Field Stability of MEMS Products**
T. Kieran Nunan, Mark G. da Silva 1481
42.1 High-Volume Manufacturing Strategy 1482
42.2 Robust Design for Volume Manufacturing 1491
42.3 Stable Field Performance 1496
42.4 Internet of Things (IoT) 1498
42.5 Conclusions and Outlook 1503
References 1503

43 **Packaging and Reliability Issues in Micro/Nano Systems**
Yu-Chuan Su, Jongbaeg Kim, Yu-Ting Cheng, Mu Chiao, Liwei Lin 1505
43.1 Introduction to MEMS Packaging 1505
43.2 Hermetic and Vacuum Packaging 1511
43.3 Emerging Packaging Approaches 1518
43.4 Thermal Issues and Packaging Reliability 1526
43.5 Future Trends and Summary 1533
References 1534

Part J Nanotechnology and Society and Education

44 **Nanotechnologies in Societal Context**
Barbara Herr Harthorn 1543
44.1 Assessing Technological Progress in Societal Terms 1545
44.2 Nanotechnologies and Upstream Societal Engagement 1547
44.3 Ethics 1549
44.4 Governance, Law and Regulation 1550
44.5 Public Perceptions and Participation in Decision-Making 1552
44.6 Integrating the Societal with the Technical 1553
44.7 Concluding Remarks 1554
References 1554

45 **Environment, Health and Safety Issues in Nanotechnology**
Rui Chen, Chunying Chen 1559
45.1 Impacts of the Development of Nanotechnology on Environment, Health, and Safety Issues 1559
45.2 Current Progress of the Most Important Nanomaterials 1562
45.3 Physicochemical Characteristics of Nanoparticles that Determine the Toxicity Impacts 1569
45.4 Novel Techniques and Biomarker Development in Nanotoxicology . 1574
45.5 Conclusion and Perspectives 1580
References 1580

46 **Nanoscience and Nanotechnology Convergence**
William S. Bainbridge 1587
46.1 Background 1587
46.2 Convergence with Society 1591
46.3 Concepts and Methods 1595
46.4 Convergence in Education 1598
46.5 Conclusions 1600
References 1600

47 **Global Perspectives of Nanotechnology Education**
Kurt Winkelmann, Bharat Bhushan 1603
47.1 Growth and Trends of Nanotechnology Education 1605
47.2 Primary and Secondary Education 1608
47.3 Vocational Education Training 1610
47.4 Undergraduate Education 1611
47.5 Graduate Education 1614
47.6 Teacher Professional Development 1616
47.7 Informal Education 1618
47.8 Summary and Outlook 1620
References 1622

About the Authors 1625
Detailed Contents 1645
Subject Index 1671

List of Abbreviations

μCP	microcontact printing
μTAS	micro total analysis system
0-D	zero-dimensional
1-D	one-dimensional
2-D	two-dimensional
2-DES	two-dimensional electron system
3-D	three-dimensional
5-FU	5-fluorouracil

A

a-C	amorphous carbon
ABP	actin binding protein
AC	alternating current
ADC	analog to digital converter
ADME	absorption, distribution, metabolism and excretion
AES	Auger electron spectroscopy
AF	acceleration factor
AFAM	atomic force acoustic microscopy
AFM	atomic force microscopy
AgNP	silver nanoparticle
AGNR	armchair nanoribbon
ALD	atomic layer deposition
AM	amplitude modulation
AM-SFM	amplitude modulation scanning force microscopy
AOD	acousto-optical deflector
AOM	acousto-optical modulator
AP	adhesion proteoglycan
APCVD	atmospheric-pressure chemical vapor deposition
APDMES	(3-aminopropyl)-dimethyl-ethoxysilane
APTES	(3-aminopropyl)-triethoxysilane
ARPES	angle-resolved photoemission spectroscopy
ASIC	application-specific integrated circuit
ASL	antistiction layer
ASNOM	apertureless scanning near-field optical microscopy
ATP	adenosine triphosphate
ATRP	atom transfer radical polymerization

B

BCB	benzocyclobutene
BCP	block copolymer
BCS	Bardeen–Cooper–Schrieffer
BDCS	biphenyldimethylchlorosilane
BECC	bottom effect cone correction
BEN	biased enhanced nucleation
BFg	fibrinogen
BGA	ball grid array
BHF	buffered HF
BHPET	(3,6,9,12,15-pentaoxapentadecane-1,15-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethanesulfonyl)imide]
BHPT	(pentane-1,5-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethanesulfonyl)imide]
BiCMOS	bipolar complementary metal oxide semiconductor
bioMEMS	biological microelectromechanical systems
bioNEMS	biological nanoelectromechanical systems
BiP	binding immunoglobulin protein
BIPY	bipyridinium
BMP-2	bone morphogenic protein-2
BOE	buffered oxide etch
BP	binding probability
BPAG1	bullous pemphigoid antigen 1
BPEI	branched polyethyleneimine
BPM	bit-patterned media
BPSG	borophosphosilicate glass
BPT	1,1′-biphenyl-4-thiol
BPTC	1,1′-biphenyl-4-thiol crosslinked
BSA	bovine serum albumin

C

c-side	cytosolic side
CA	contact angle
CAH	contact-angle hysteresis
CAIBE	chemically assisted ion beam etching
CAM	cell-adhesion molecule
CAS	Crk-associated substrate
CBA	cantilever beam array
CCD	charge-coupled device
CCVD	catalytic chemical vapor deposition
CDM	charge discharge model
cDNA	complementary DNA
CDR	complementarity determining region
CDW	charge density wave
CFM	chemical force microscopy
CFU	colony forming unit
CG	controlled geometry
CIC	cantilever-in-cantilever
CLSM	confocal laser scanning microscopy
CMA	cylindrical mirror analyzer
CMC	critical micelle concentration
CMOS	complementary metal-oxide-semiconductor
CMP	chemical mechanical polishing
CNFET	carbon nanotube field-effect transistor

CNH	carbon nanohorn
CNT	carbon nanotube
COC	cyclic olefin copolymer
COOH-PS	charged carboxylated polystyrene
CP	circularly permuted
CPA	chirped-pulsed amplifier
CPB	Al Cooper pair box
CPG	compute pattern generator
CRP	C-reactive protein
CS	circular dichroism spectroscopy
CSK	cytoskeleton
CSM	continuous stiffness measurement
CT	computed tomography
CTAB	hexadecyl trimethyl ammonium bromide
CTE	coefficient of thermal expansion
CU	control unit
CURC	curcumin
CVD	chemical vapor deposition
CW	continuous wave

D

dATP	deoxyadenosine triphosphate
DBBA	10,10-dibromo-9,9-bianthryl
DBR	distributed Bragg reflector
DBTF	dibromoterfluorene
DC	direct current
DC/RF	direct current/radio frequency
DCE	1,2-dichloroethane
dCTP	deoxycytidine triphosphate
dDAT	*Drosophila melanogaster* dopamine transporter
DDMS	dimethyldichlorosilane
ddNTP	dideoxynucleotide triphosphate
ddPCR	droplet digital polymerase chain reaction
DDT	dichlorodiphenyltrichloroethane
DDV	dichlorvos
DEP	dielectrophoresis
DFB	distributed feedback
DFS	dynamic force spectroscopy
DFT	density functional theory
dGTP	deoxyguanosine triphosphate
DGU	density gradient ultracentrifugation
DI	deionized
DIMP	diisopropyl methylphosphonate
DIP	dual-in-line packaging
DLA	diffusion-limited aggregation
DLC	diamond-like carbon
DLD	deterministic lateral displacement
DLP	digital light processing
DLS	dynamic light scattering
DLVO	Derjaguin–Landau–Verwey–Overbeek
DLW	direct laser writing
DMA	*N*,*N*-dimethylamide
DMAc	dimethylacetamide
DMD	digital micromirror device
DMF	dimethylformamide
DMMP	dimethylmethylphosphonate
DMSO	dimethyl sulfoxide
DMT	Derjaguin–Muller–Toporov
DNA	deoxyribonucleic acid
DNS	direct numerical simulation
DOE	design of experiments
DOF	degree of freedom
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DOS	density of states
DOTA	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
Dox	doxorubicin
DP	*n*-decylphosphonate
DPN	dip-pen nanolithography
DPPC	1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine
DPW	die per wafer
DRC	design rule check
DRIE	deep reactive-ion etching
DSA	directed self-assembly
DSC	differential scanning calorimetry
DSP	digital signal processor
DSPE-PEG	1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[amino(polyethyleneglycol)-2000] (carboxylic acid)
dSTORM	direct stochastic optical reconstruction microscopy
DTR	discrete track recording
dTTP	deoxythymidine triphosphate
DTXL	docetaxel
DUV	deep ultraviolet
DWNT	double-wall nanotube

E

EAM	embedded atom method
EB	electron beam
EBID	electron-beam-induced deposition
EBL	electron-beam lithography
EC	erythrocyte
ECD	electrochemical deposition
ECM	extracellular matrix
ECR	electron cyclotron resonance
ECR-CVD	electron cyclotron resonance chemical vapor deposition
ED	electron diffraction
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDL	electrostatic double layer
EDLC	electrical double-layer capacitor
EDP	ethylenediamine pyrocatechol
EDS	energy-dispersive x-ray spectroscopy
EELS	electron energy loss spectroscopy
EFM	electric field gradient microscopy
EHL	elastohydrodynamic lubrication
EIP	emulsion inversion point
EL	electroluminescence
ELISA	enzyme-linked immunosorbent assay
ELNES	electron energy loss near edge structure

ELP	elastin-like polypeptide
EMI	electromagnetic interference
EMT	epithelial–mesenchymal transition
ENM	engineered nanomaterial
ENP	engineered nanoparticle
EPR	enhanced permeability and retention
EPS	extracellular polymeric substance
ER	endoplasmic reticulum
ERK	extracellular signal-regulated kinase
ESD	electrostatic discharge
EUV	extreme ultraviolet
EWOD	electrowetting-on-dielectric

F

FADS	fluorescence-activated droplet sorting
FAK	focal adhesion kinase
FB	fluidized bed
FBAR	film bulk acoustic resonator
FBGC	foreign body giant cell
FBS	fetal bovine serum
FCA	filtered cathodic arc
fcc	face-centered cubic
FCP	force calibration plot
FDM	fused deposition modeling
FDTD	finite-difference time-domain
FDTS	1*H*,1*H*,2*H*,2*H*-perfluoro-decyltrichlorosilane
FE-SEM	field electron SEM
FEA	finite element analysis
FEM	finite element modeling
FESEM	field-emission scanning electron microscopy
FET	field-effect transistor
FFM	friction force microscope
FFT	fast Fourier transform
FIB	focused ion beam
FK	Frenkel–Kontorova
FKT	Frenkel–Kontorova–Tomlinson
FLG	few-layer graphene
FluidFM	fluid force microscopy
FM	frequency modulation
FM-SFM	frequency-modulation scanning force microscopy
FMEA	failure modes and effects analysis
fPALM	fluorescent photoactivated localization microscopy
FPGA	field-programmable logic arrays
FRET	fluorescence (or Förster) resonant energy transfer
FRS	forward recoil spectrometry
fs-laser	femtosecond laser
FWHM	full-width at half-maximum

G

GF	gauge factor
GFP	green fluorescent protein
GIC	graphite intercalation compounds
GLAD	glancing angle deposition
GLUT4	glucose transporter 4
GMR	giant magnetoresistive
GNP	gold nanoparticle
GNR	graphene nanoribbon
GO	graphene oxide
GSD	ground-state depletion
GSH-Px	glutathione peroxidase
GTP	guanosine triphosphate
GW	Greenwood and Williamson

H

HAADF-STEM	high-angle annular dark-field-scanning transmission electron microscope
HARMEMS	high-aspect-ratio MEMS
HARPSS	high-aspect-ratio combined poly- and single-crystal silicon
HART	high aspect ratio tip
HAS-CtW	hybrid self-assembly-based chip-to-wafer bonding
HBM	human body model
hBN	hexagonal boron nitride
hcp	hexagonal close-packed
HCS	high content screening
HDI	high density interconnect
HDL	high density lipoprotein
HDT	hexadecane thiol
HEL	hot embossing lithography
HEMT	high electron mobility transistor
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
HER	hydrogen evolution recovery
HEXSIL	hexagonal honeycomb polysilicon
HFCVD	hot filament chemical vapor deposition
hIL	human interleukin
HMDS	hexamethyldisilazane
HOMO	highest occupied molecular orbital
HOPG	highly oriented pyrolytic graphite
HOT	holographic optical tweezers
HP	hot pressing
HPLC	high-pressure liquid chromatography
HRTEM	high-resolution transmission electron microscopy
HSA	human serum albumin
HSQ	hydrogen silsesquioxane
HTS	high-throughput screening
HTSC	high-temperature superconductor
HVM	high-volume manufacturing

I

I/O	input/output
IA	isogeometric analysis
IB	ion beam
IBD	ion beam deposition
IBM	immersed boundary method
IC	integrated circuit

IC3D	integrated comprehensive droplet digital detection
ICA	independent component analysis
ICP	inductively coupled plasma
ICP-MS	inductively coupled plasma mass spectrometry
ICP-OES	inductively coupled plasma optical emission spectroscopy
IFEM	immersed finite element method
IgG	immunoglobuline G
IKVAV	isoleucine–lysine–valine–alanine–valine
IMU	inertial measurement unit
IoT	internet of things
IP	ion pump
IPA	isopropanol alcohol
IPTG	isopropyl-β-D-thiogalactoside
IR	infrared
IRE1	inositol-requiring protein 1
IT	isotopic tracing
ITO	indium tin oxide

J

JC	jump-to-contact
JFIL	jet-and-flash imprint lithography
JKR	Johnson–Kendall–Roberts

K

KPFM	Kelvin probe force microscopy

L

L–B	Langmuir–Blodgett
LAMP	loop-mediated isothermal amplification
LBL	layer-by-layer
LBM	lattice Boltzmann method
LCD	liquid-crystal display
LCoS	liquid crystal on silicon
LCPD	local contact potential difference
LDA	local density approximation
LDL	low-density lipoproteins
LDOS	local density of states
LED	light-emitting diode
LEED	low-energy electron diffraction
LFM	lateral force microscope
LIF	light-induced fluorescent
LIGA	lithography, electroforming and molding
LINC	linker of nucleoskeleton and cytoskeleton
LING	laterally integrated nanogenerator
LJ	Lennard-Jones
LNA	locked nuclei acid
LoC	lab-on-chip
LoD	limit-of-detection
LOEC	lowest observable effect concentration
LPCVD	low-pressure chemical vapor deposition
LSI	large-scale integrated circuit
LSP	localized surface plasmon
LT	low-temperature
LT-STM	low-temperature scanning tunneling microscopy
LTM	laser tracking microrheology
LTO	low-temperature oxide
LTRS	laser tweezers Raman spectroscopy
LUMO	lowest unoccupied molecular orbital
LVDT	linear variable differential transformer

M

m-SWNT	metallic single-walled carbon nanotube
MACE	metal-assisted chemical etching
MALDI	matrix-assisted desorption or ionization
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight mass spectrometer
MAPK	mitogen-activated protein kinase
MBE	molecular beam epitaxy
MC	microcantilever
MCBJ	mechanically controlled break junction
MCM	multichip module
MD	molecular dynamics
MDA	malondialdehyde
ME	metal evaporated
MEAM	modified EAM
MED	monolithic elastomer device
MEMS	microelectromechanical system
MExFM	magnetic exchange force microscopy
MExFS	magnetic exchange force spectroscopy
MFC	mass flow controller
MFFD	microflow-focusing device
MFM	magnetic force microscopy
MHD	magnetohydrodynamic
MIC	minimum inhibitory concentration
micro-CT	micro-computer tomography
MIM	metal/insulator/metal
MLE	maximum likelihood estimator
MLG	multilayer graphene
MM	machine model
MMS	methylmethanesulfonate
MOCVD	metalorganic chemical vapor deposition
MOEMS	microoptoelectromechanical systems
MOS	metal–oxide–semiconductor
MOSFET	metal–oxide–semiconductor field-effect transistor
MP	metal particle
MPCVD	microwave plasma CVD
MPECVD	microwave-PECVD
MPPD	multiple path particle dosimetry
MPTMS	3-mercaptopropyltrimethoxysilane
MR	magnetoresistance
MRAM	magnetoresistive random access memory
MRFM	magnetic resonance force microscopy
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MRP	molecular recognition phases

mSBF	modified simulated body fluid
MSCK	magnetic shell cross-linked knedel-like nanoparticles
MscL	mechanosensitive channel of large conductance
MSS	membrane-type surface stress sensor
MST	microsystem technology
MTTF	mean time to failure
MTX	methotrexate
MUMP	multi-user MEMS process
MWNT	multiwalled (carbon) nanotube

N

NAA	neutron activation analysis
NAC	*N*-acetyl-cysteine
NAT	nuclear analytical technique
NCG	glass nanochannel
NE	nuclear envelope
NEMS	nanoelectromechanical system
NEXAFS	near-edge x-ray absorption fine structure
NGL	next-generation lithography
NHS	*N*-hydroxysuccinimide
NIL	nanoimprint lithography
NIR	near-infrared
NMP	nitroxide mediated polymerization
NMR	nuclear magnetic resonance
NOEMS	nanooptoelectromechanical systems
NP	nanoparticle
NPC	nuclear pore complex
NR	nanorod
NRAM	nonvolatile random access memory
NRM	nanorobotic manipulator
NSOM	near-field scanning optical microscopy
NSS	neurotransmitter sodium symporter
NST	nanosystem technology
NT	nanotube
NTA	nitrile-tri-acetic acid
nTP	nanotransfer printing
NW	nanowire

O

O-E-O	optical-electronic-optical switches
ODDMS	*n*-octadecylmethyl (dimethylamino) silane
ODMS	*n*-octyldimethyl (dimethylamino) silane
ODP	*n*-octadecylphosphonate
ODTS	octadecyltrichlorosilane
OLED	organic light-emitting device
OM	optical microscope
OMCTS	octamethylcyclotetrasiloxane
OMVPE	organometallic vapor phase epitaxy
OS	optical stretcher
OTF	optical transfer function
OTRS	optical tweezers Raman spectroscopy
OTS	octadecyltrichlorosilane

P

PA	peptide amphiphile
PA-GFP	photoactivatable green fluorescent protein
PA-PF	photoactivatable protein fluorophores
PAH	polycyclic aromatic hydrocarbon
PALM	photoactivated localization microscopy
PAMAM	polyamidoamine
PANI	polyaniline
PBC	periodic boundary condition
PBD	polybutadiene
PBG	photonic band gap
PBS	phosphate buffered saline
PCL	polycaprolactone
PCR	polymerase chain reaction
PCR-CE	polymerase chain reaction-capillary electrophoresis
PDF	probability density function
PDGF	platelet-derived growth factor
PDK	process design kit
PDMS	poly(dimethylsiloxane)
PE	polyethylene
PECVD	plasma-enhanced chemical vapor deposition
PEDOT	polyethylene dioxythiophene
PEEK	polyether ether ketone
PEG	polyethylene glycol
PEGMA	poly(ethylene glycol) methacrylate
PEI	polyethylenimine
PERK	PKR-like endoplasmic reticulum kinase
PES	photoemission spectroscopy
PET	poly(ethylene terephthalate)
PETN	pentaerythritol tetranitrate
PEUT	poly(ether urethane)
PFDA	perfluorodecanoic acid
PFDP	perfluorodecylphosphonate
PFDTES	perfluorodecyltriethoxysilane
PFPE	perfluoropolyether
PFTS	perfluorodecyltricholorosilane
PGA	polyglycolic acid
PhC	photonic crystal
PI	polyimide
PI3K	phosphatidylinositol-3-kinase
PIN	p-type/intrinsic/n-type
PIT	phase inversion temperature
PKA	protein kinase
PKC	protein kinase C
PKI	protein kinase inhibitor
PL	photoluminescence
PLA	polylactic acid
PLAF	poly(lactide fumarate)
PLD	pulsed laser deposition
PLGA	poly(lactic-*co*-glycolic acid)
PLGF	poly(lactide-*co*-glycolide fumarate)
PLL	phase lock loop
PMAA	poly(methacrylic acid)
PMMA	poly(methyl methacrylate)
PMN	polymorphonuclear

PMN–PT	lead magnesium niobate-lead titanate
PmPV	poly[(*m*-phenylenevinylene)-*co*-(2,5-dioctoxy-*p*-phenylenevinylene)]
PMT	photomultiplier tube
PN	purine nucleotide
PNA	peptide nucleic acid
PNIPAm	poly(*N*-isopropylacrylamide)
poly-Si	polysilicon
POM	polyoxymethylene
PP	plasmon polariton
PPDDAC	poly(diallyldimethylammonium chloride)
PPGMA	poly(propylene glycol) methacrylate
PPMA	poly(propyl methacrylate)
PPNF	polypropylene nonwoven fabric
PS	polystyrene
PS-b-PDMS	poly(styrene-*b*-dimethylsiloxane)
PSA	prostate-specific antigen
PSD	position-sensitive detector
PSF	point spread function
PSG	phosphosilicate glass
PT	Prandtl–Tomlinson
PTCDA	perylenetetracarboxylic dianhydride
PTFE	polytetrafluoroethylene
PTR	pole-tip recession
PTX	paclitaxel
PVD	physical vapor deposition
PZT	lead zirconate titanate

Q

QB	quantum boxes
QCM	quartz crystal microbalance
QD	quantum dot
QEM	quantum electromechanics
QHE	quantum Hall effect
QPD	quadrant photodiode
QSAR	quantitative structure–activity relationship
QWR	quantum wire

R

RAFT	reversible addition and fragmentation transfer
RAM	random access memory
RBC	red blood cell
RDL	redistribution layer
RESLOFT	reversible saturable optical fluorescence transition
RF	radio frequency
RF MEMS	radio frequency MEMS
RGD	arginine–glycine–aspartic acid
RH	relative humidity
RHEED	reflection high-energy electron diffraction
RIE	reactive ion etching
RKKY	Ruderman–Kittel–Kasuya–Yoshida
RMS	root mean square
RNA	ribonucleic acid
RO	reverse osmosis
ROS	reactive oxygen species
RSF	rate-and-state friction
RT	room temperature
RT-DC	real-time deformability cytometry
RT-PCR	reverse-transcription PCR
RTA	rapid thermal annealing
RTP	rapid thermal processing

S

s-SWNT	semiconducting single-walled carbon nanotube
SA-MCtW	self-assembly-based multi-chip-to-wafer bonding
SAED	selected area electron diffraction
SAIL	self-aligned imprint lithography
SAM	self-assembled monolayer
SAW	surface acoustic wave
SB	Schottky barrier
SCC	stress corrosion cracking
SCFv	single-chain fragment variable
SCIL	substrate conformal imprint lithography
SCM	scanning capacitance microscopy
SCPM	scanning chemical potential microscopy
SCREAM	single-crystal reactive etching and metallization
SDA	scratch drive actuator
SEcM	scanning electrochemical microscopy
SEFM	scanning electrostatic force microscopy
SEM	scanning electron microscopy
SERT	serotonin transporter
SET	single electron transistor
SFA	surface force apparatus
SFAM	scanning force acoustic microscopy
SFM	scanning force microscopy
SHG	second harmonic generation
SHiMMeR	Sandia high-volume measurement of micromachine reliability
SHO	simple harmonic oscillator
SICM	scanning ion conductance microscopy
SKPM	scanning Kelvin probe microscopy
SL	soft lithography
SLIPS	slippery liquid-infused surfaces
SLS	solution–liquid–solid
SMA	shape memory alloy
SMFS	single molecule force spectroscopy
SMLM	single molecule localization microscopy
SMM	scanning magnetic microscopy
SMR	suspended micro-channel resonator
SMRT	single-molecule real-time
SNOM	scanning near-field optical microscopy
SNR	signal-to-noise ratio
SNS	superconductor–normal metal–superconductor
SOD	superoxide dismutase
SOFI	superresolution optical fluctuation imaging

SOG	spin-on-glass
SOI	silicon-on-insulator
SOIMEMS	silicon-on-insulator microelectromechanical systems
SoS	silicon-on-sapphire
SP	sputtering
SP-STM	spin-polarized scanning tunneling microscopy
SPECT	single-photon emission computed tomography
SPIM	selective plane illumination microscopy
SPM	scanning probe microscopy
SPN	spherical polymeric nanoconstruct
SPR	surface plasmon resonance
SPS	spark plasma sintering
sptPALM	single-particle tracking PALM
SQL	standard quantum limit
SRAM	static random access memory
SRCD	synchrotron radiation circular dichroism spectroscopy
SSIL	step-and-stamp imprinting lithography
SSRM	scanning spreading resistance microscopy
STED	stimulated emission depletion
STEM	scanning transmission electron microscope
SThM	scanning thermal microscopy
STM	scanning tunneling microscopy
STORM	stochastic optical reconstruction microscopy
STP	standard temperature and pressure
STS	scanning tunneling spectroscopy
SUN	Sad1p/UNC-84
SVR	surface-to-volume ratio
SWNT	single-walled (carbon) nanotube

T

TASA	template assisted self-assembly
TB-BGA	top–bottom ball grid array
TCP	tricresyl phosphate
TEM	transmission electron microscopy
TEP	thermoelectric power
TFET	tunnel field-effect transistor
TFM	traction force microscopy
TGA	thermogravimetric analysis
TGF	transforming growth factor
THC	tetrahydrocurcumin
THF	tetrahydrofuran
THG	third harmonic generation
TIRF	total internal reflection fluorescence
TM	tapping mode
TMA	trimethylaluminum
TMP	turbo-molecular pump
TNT	trinitrotoluene
TP	tribopolymer
TR	torsional resonance
TREC	topography and recognition imaging
TRIM	transport of ions in matter
tRNA	transfer ribonucleic acid
TSV	through-silicon via
TTF	tetrathiafulvalene
TTS	time-temperature superposition
TXM	transmission x-ray microscopy

U

UAA	unnatural AA
UFP	ultrafine particle
UHF	ultra high frequency
UHV	ultrahigh vacuum
ULSI	ultra large scale integration
UML	unified modeling language
UNCD	ultrananocrystalline diamond
UPR	unfolded protein response
UV	ultraviolet
UV-Vis	ultraviolet-visible

V

VBS	vinculin binding site
VCO	voltage-controlled oscillator
VCSEL	vertical-cavity surface-emitting laser
vdW	van der Waals
VE	vascular endothelial
VEGF	vascular endothelial growth factor
VGCNF	vapor-grown carbon nanofilament
VHH	variable heavy–heavy
VING	vertical nanowire array integrated nanogenerator
VLDL	very low density lipoprotein
VLS	vapor–liquid–solid
VLSI	very large-scale integration
VOC	volatile organic compound
VPE	vapor-phase epitaxy
VSM	vital-sign monitoring
VSS	vapor–solid–solid

W

WAT	wafer acceptance test
WLCSP	wafer-level chip-scale packaging

X

XAFS	x-ray absorption fine structure
XANES	x-ray absorption near-edge structure
XPS	x-ray photoelectron spectroscopy
XRD	x-ray diffraction
XRF	x-ray fluorescence

Z

ZMW	zero-mode waveguide

1. Introduction to Nanotechnology

Bharat Bhushan

Nanotechnology refers to technology that is implemented at the nanoscale and has applications in the real world. Unique physical and chemical properties of nanomaterials can be exploited for applications that benefit society. Nanotechnology represents a *megatrend* and has become a *general purpose* technology. An executive action of 2000, the National Nanotechnology Initiative was formalized with the 21st Century Nanotechnology Research and Development Act in 2003. Through FY 2015 in the United States, federal R&D investment has been about $20 billion, with annual investment in FY 2015 of about $1.5 B, and more than double that by the private sector. The revenues from nanoenabled products continue growing, with over $200 B in FY 2012 in the US alone, and over $700 B worldwide. This represents an impressive return on investment. Basic research and development as well as research on potential safety issues of nanotechnology, workforce development, and education and curriculum should be continued.

> *A biological system can be exceedingly small. Many of the cells are very tiny, but they are very active; they manufacture various substances; they walk around; they wiggle; and they do all kinds of marvelous things – all on a very small scale. Also, they store information. Consider the possibility that we too can make a thing very small which does what we want – that we can manufacture an object that maneuvers at that level.*
>
> R.P. Feynman: There's Plenty of Room at the Bottom. In: *Annu. Meet. Am. Phys. Soc. California Institute of Technology, Pasadena* (1959)

1.1 **Nanotechnology – Definitions and Examples** 2
1.1.1 Nanomanufacturing 2
1.1.2 MEMS/NEMS 2
1.1.3 Convergence 4
1.1.4 Industrial Applications 4
1.2 **History and Early Research Expenditures** 6
1.3 **Governance of the National Nanotechnology Initiative** 7
1.4 **Nanotechnology R&D Funding Since 2001** 8
1.5 **Worldwide R&D Investments and Output** 11
1.5.1 R&D Investments 11
1.5.2 Output 11
1.6 **Translation of Basic Research to Technology Commercialization** 15
1.7 **Nanoscience and Nanotechnology Education** 16
1.7.1 Interdisciplinary Education 16
1.7.2 K–12 Education, Excitement, and Creativity 16
1.7.3 Vocational, Undergraduate, Graduate, and Continuing Education Programs 16
1.7.4 Educator Training 17
1.7.5 Informal Science Education 17
1.8 **Summary and Outlook** 17
1.9 **Organization of the Handbook** 18
References 18

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_1

1.1 Nanotechnology – Definitions and Examples

Nanotechnology refers to any technology that is implemented at the nanoscale and has applications in the real world. It is defined as the control or restructuring of matter at the atomic and molecular levels in the size range of about 1 to 100 nm [1.1, 2]. The underlying science is referred to as nanoscience. The properties of matter at the nanoscale are different from those at a larger scale. When the dimensions of a material are reduced from a large size, the properties remain the same at first, then small changes occur. Finally, when the size drops below 100 nm, dramatic changes in properties can occur. The unique physical and chemical properties of nanomaterials can be exploited for commercial applications and for novel performance that benefits society. The discovery of novel materials, processes, and phenomena at the nanoscale and the development of new experimental and theoretical techniques for research at the end of the 20th century provide fresh opportunities for the development of innovative nanosystems and nanomaterials. This field is opening new avenues in science and technology [1.3–5].

Nanotechnology encompasses the nanomanufacturing, nanomaterials, and application of physical, chemical, and biological systems at scales ranging from individual atoms or molecules to submicron dimensions, as well as the integration of the resulting nanostructures into larger systems. It spans scientific fields, including chemistry, physics, material science, engineering, and manufacturing. Its impact on our society and economy in the 21st century is comparable to that of semiconductor technology, information technology, or cellular and molecular biology in the 20th century. The integration of nanotechnology into larger systems has provided breakthrough solutions to many current environmental, medical, and industrial problems, including smart materials, nanomanufacturing, electronics, drug delivery, energy and water, biotechnology, information technology, and national security. Nanotechnology will have a profound impact on our economy and society; it is a modern industrial revolution. Nanotechnology represents a megatrend, bringing disruptive innovation. It has become a general purpose technology, being applicable across various industrial sectors [1.3, 4].

1.1.1 Nanomanufacturing

Manufacturing at the nanoscale, referred to as nano-manufacturing, is accomplished by using either a *bottom–up* or *top–down* approach to the production of nanomaterials, structures, devices, and systems [1.4–6].

In the bottom–up approach, nanoscale features are mainly built up from their elemental constituents. Elemental constituents are combined using various nanomanufacturing processes including self-assembly, chemical synthesis, molecular beam epitaxy, nanoimprint lithography, roll-to-roll processing, and dip pen lithography (http://www.nano.gov). Molecular self-assembly, the spontaneous self-assembly of molecular clusters, can occur from simple reagents in solution or from biological molecules (e.g., DNA) used as building blocks for the production of three-dimensional nanostructures. Chemical synthesis is carried out using gaseous precursors and solvents. These chemicals react to produce nanostructures. A variety of vacuum deposition and nonequilibrium plasma chemistry techniques are then used to produce layered nanocomposites and nanotubes. Molecular beam epitaxy is a method for deposition of thin films with a thickness of one atom on a surface. Nanoprint lithography is used to fabricate nanostructures by stamping or printing them on a surface. Roll-to-roll processing is a high volume process used to produce nanodevices on a flexible substrate. Dip pen lithography uses an atomic force microscope tip in a fluid to produce nanoscale features on a surface.

The top–down approach uses lithographic and nonlithographic fabrication technologies. Lithographic technology is an underlying technology to make semiconductor chips and components. Continued improvements in lithography for use in the production of nanocomponents have resulted in line widths as small as 10 nanometers. The top–down approach is commonly used to fabricate micro- and nanosystem components, which range in size from micro- to nanometers.

1.1.2 MEMS/NEMS

Micro- & nanosystems include Micro/NanoElectroMechanical Systems (MEMS/NEMS). The acronym MEMS originated in the US. The term commonly used in Europe is micro/nanosystem technology (MST/NST), and in Japan it is micro/nanomachines. Another term generally used is micro/nanodevices. Advances in silicon photolithographic process technology since the 1960s led to the development of MEMS in the early 1980s. MEMS combine electrical and mechanical components to create microscopic devices that have a characteristic length between 100 nm and 1 mm. In the early 2000s, with the development of nanofabrication techniques, NEMS were fabricated. NEMS combine electrical and mechanical components to create nanoscopic devices that have a characteristic length of less than 100 nm. In mesoscale devices, if the functional components are on the micro- or nano-

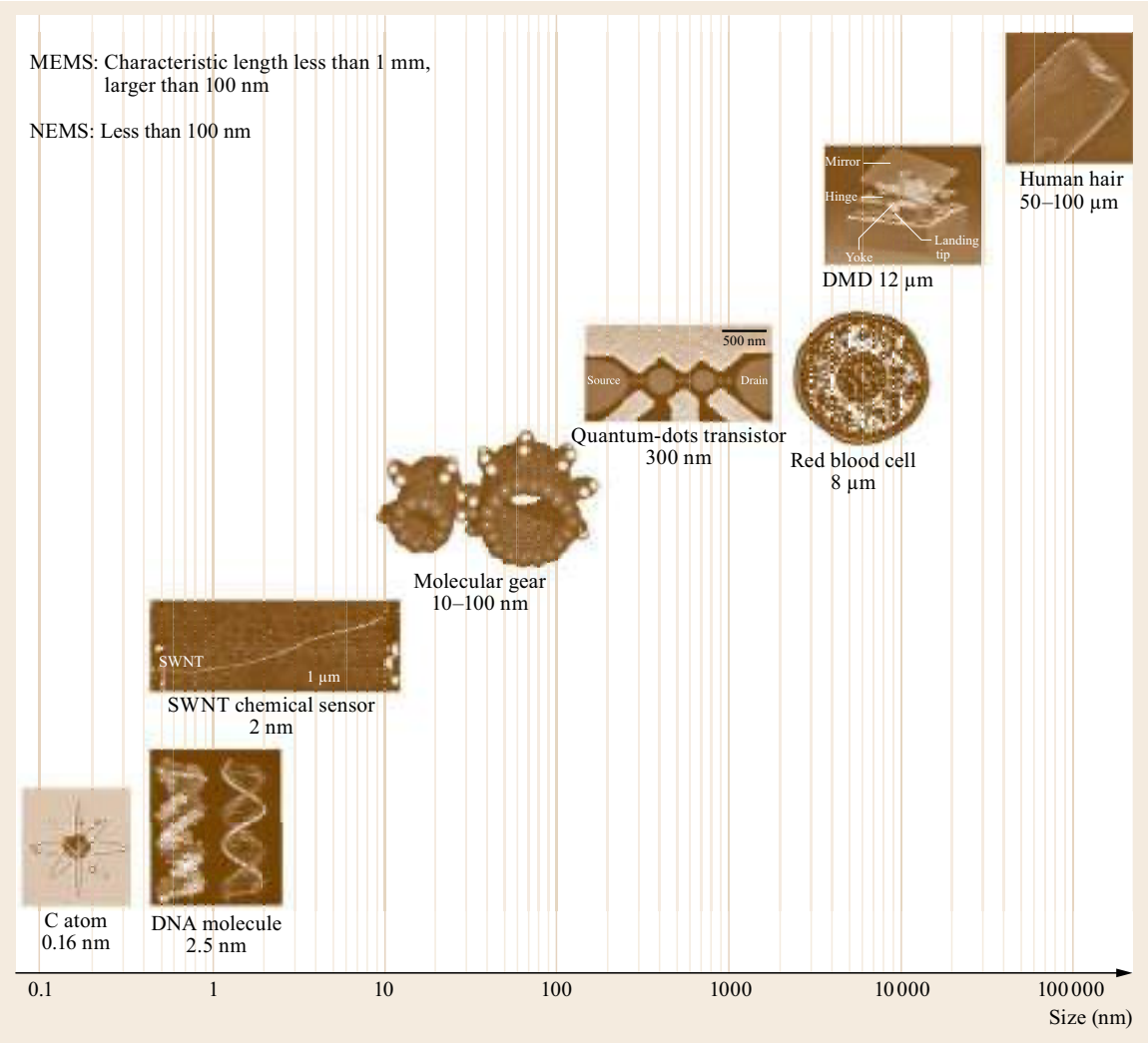


Fig. 1.1 Characteristic dimensions of nanomaterials, MEMS/NEMS, and BioNEMS in perspective (after [1.3]). Examples shown are a single-walled carbon nanotube (SWNT) chemical sensor (after [1.7]), molecular dynamic simulations of carbon nanotube-based gears (after [1.8]), a quantum-dot transistor (after [1.9]), and a digital micromirror device (DMD) (after [1.10])

scale, they may be referred to as MEMS or NEMS, respectively. These are intelligent miniaturized systems comprised of sensing, processing, and/or actuating functions that combine electrical and mechanical components [1.4]. The MEMS/NEMS terms are also now used in a broad sense and include electrical, mechanical, fluidic, optical, and/or biological function. MEMS/NEMS for optical applications are referred to as micro/anooptoelectromechanical systems (MOEMS/NOEMS). MEMS/NEMS for electronic applications are referred to as radio-frequency-MEMS/NEMS or RF-MEMS/RF-NEMS. MEMS/NEMS for biological applications are referred to as bioMEMS/bioNEMS.

To put the dimensions and weights of nanomaterials, MEMS/NEMS, and BioNEMS in perspective, see Fig. 1.1 and Table 1.1. Examples shown are a single-walled carbon nanotube (SWNT) chemical sensor (adapted from [1.7]), molecular dynamic simulations of carbon nanotube-based gears (adapted from [1.8]), a quantum-dot transistor (adapted from [1.9]), and a digital micromirror device (adapted from [1.10]). Individual atoms are typically a fraction of a nanometer in diameter, DNA molecules are about 2.5 nm wide, biological cells are in the range of thousands of nm in diameter, and human hair is about 75 000 nm (75 µm) in diameter. The smaller length of a BioNEMS shown

Table 1.1 Characteristic dimensions and weights of various objects in perspective (after [1.3])

Characteristic dimensions in perspective	
Fabricated	
NEMS characteristic length	< 100 nm
MEMS characteristic length	< 1 mm and > 100 nm
SWNT chemical sensor	≈ 2 nm
Molecular gear	≈ 10 nm
Quantum-dots transistor	300 nm
Digital micromirror	12 000 nm
Natural	
Individual atoms	Typically a fraction of a nm in diameter
DNA molecules	≈ 2.5 nm wide
Biological cells	In the range of thousands of nm in diameter
Human hair	$\approx 75\,000$ nm in diameter
Weight in perspective	
Fabricated	
NEMS built with cross-sections of about 10 nm	As low as 10^{-20} N
Micromachine silicon structure	As low as 1 nN
Natural	
Eyelash	≈ 100 nN
Water droplet	$\approx 10\,\mu$N

in the figure is about 2 nm. NEMS shown range in size from 10 to 300 nm, and the size of MEMS is on the order of 12 000 nm (12 μm). The weight of a micromachined silicon structure can be as low as 1 nN, and NEMS can be built with weights as low as 10^{-20} N with cross-sections of about 10 nm. In comparison, the weight of a drop of water is about 10 μN, and the weight of an eyelash is about 100 nN.

1.1.3 Convergence

Roco and *Bainbridge* [1.11] have discussed convergence of nanotechnology, biotechnology, information, and cognitive science for emerging technologies. Convergence is the process of bringing disparate technologies together into a unified field. Convergence is progressing in stages, beginning with foundational interdisciplinary research at the nanoscale between 2001 and 2010, followed by integration of nanoscience and nanotechnology for general purpose technology from approximately 2010 onwards.

1.1.4 Industrial Applications

MEMS/NEMS and BioMEMS/BioNEMS are used for applications in the industrial, consumer, defense, and biomedical fields and include electromechanical, electronics, information/communication, chemical, and biological applications, to name a few. Growth of Si-based MEMS/NEMS appears to have slowed down, while growth of nonsilicon MEMS/NEMS and BioMEMS/BioNEMS has been picking up since the first decade of the 21st century [1.4]. It is expected that nanomaterials and biomedical applications as well as nanoelectronics or molecular electronics would continue to expand. Because of the enabling nature of these systems and the significant impact they can have on both commercial and defense applications, both industry and national governments have taken a special interest in nurturing growth in this field.

The first MEMS were commercially introduced around 1990. These were accelerometers for the deployment of an air bag in a car in the event of a crash, piezoresistive pressure sensors for manifold absolute pressure sensing in car engines, and thermal inkjet print heads. In 2015, additional examples of MEMS with large and established markets with production volumes of several hundred million units per year include piezoresistive pressure sensors for disposable blood pressure sensors, capacitive pressure sensors for tire pressure measurements, digital micromirror arrays for digital projection display, and optical cross-connections in telecommunications. Other applications of MEMS devices include chemical/biosensors and gas sensors, microresonators, infrared detectors and focal plane arrays for earth observations, space science and missile defense applications, picosatellites for space applications, fuel cells, and many hydraulic, pneumatic, and other consumer products.

NEMS applications include both biological and nonbiological fields. Examples include: microcantilevers with integrated sharp nanotips for scanning tunneling microscopy (STM) and atomic force microscopy (AFM), quantum corral formed by one-by-one placing of atoms using STM, AFM tips for nanolithography, dip pen lithography for printing molecules, quantum dot transistors, nanotube-based sensors, biological (DNA) motors, molecular gears made by attaching benzene molecules to the outer walls of carbon nanotubes, devices incorporating nm-thick films (e.g., in giant magnetoresistive or GMR read/write magnetic heads and magnetic media) for magnetic rigid disk and magnetic tape drives, and nanopatterned magnetic rigid disks.

Nanomaterials include nanoparticles (aggregate of $10-10^5$ atoms bonded together), nanowires, nanotubes, and quantum wires. Nanomaterials have many applications. For example, nanoparticles are used in magnetic coatings for information storage devices and drug delivery, and nanotubes and nanowires are used in various sensor applications. Nanoelectronics are being developed from these nanomaterials. They are used to build computer memory using individual molecules or nano-

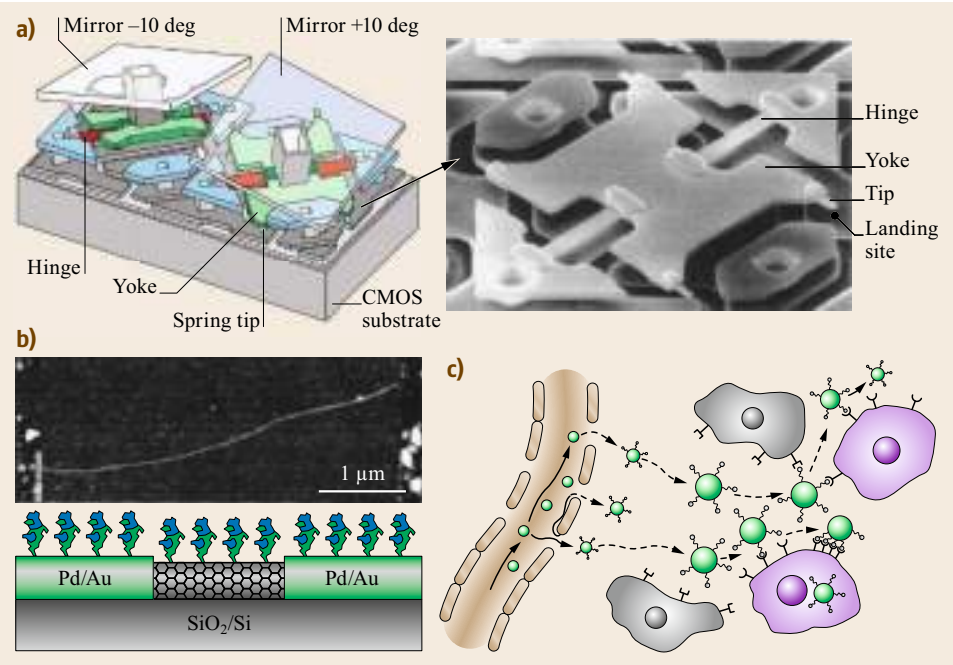


Fig. 1.2a–c Examples of **(a)** MEMS – digital micromirror device used for digital projection display (after [1.10] reprinted with permission, courtesy of Texas Instruments), **(b)** NEMS – SWNT chemical sensor (after [1.7]), and **(c)** functionalized nanoparticles for cancer detection and drug delivery (after [1.12])

tubes to store bits of information, molecular switches, molecular or nanotube transistors, nanotube flat-panel displays, nanotube integrated circuits, fast logic gates, switches, nanoscopic lasers, and nanotubes as electrodes in fuel cells.

BioMEMS/BioNEMS are increasingly being used in commercial and defense applications. They are used for chemical and biochemical analyses (biosensors) in medical diagnostics (e.g., DNA, RNA, proteins, cells, blood pressure and assays, and toxin identification), tissue engineering, and implantable pharmaceutical drug delivery. Biosensors, also referred to as biochips, deal with liquids and gases. There are two types of biosensors. The first type of biosensors are micro/nanofluidic devices. A large variety of biosensors are based on micro/nanofluidics. Micro/nanofluidic devices offer the ability to work with smaller reagent volumes and shorter reaction times, and perform analyses multiple times at once. The second type of biosensors are micro/nanoarrays, which perform one type of analysis thousands of times. Micro/nanoarrays are a tool used in biotechnology research to analyze DNA or proteins, to diagnose diseases, or discover new drugs. One microarray of silicon nanowires is used to selectively bind to and detect even a single biological molecule, such as DNA or protein. It is roughly a few nanometers in size and uses nanoelectronics to detect the slight electrical charge caused by the selective binding. Another type is a microarray of carbon nanotubes used to electrically detect glucose.

After the tragedy of 11 Sept. 2001, concern surrounding biological and chemical warfare led to the development of handheld units with bio- and chemical sensors for detection of biological germs, chemical or nerve agents, and mustard agents, as well as chemical precursors to protect subways, airports, water supplies, and the population at large. BioMEMS/BioNEMS are also being developed for minimally invasive surgery, including endoscopic surgery, laser angioplasty, and microscopic surgery. Other applications include implantable drug-delivery devices: micro/nanoparticles with drug molecules encapsulated in functionalized shells for site-specific targeting applications, and a silicon capsule with a nanoporous membrane filled with drugs for long-term delivery.

Figure 1.2 shows an example of (a) MEMS – a digital micromirror device (DMD) used for digital projection display (adapted from [1.10]), (b) NEMS – a single-walled nanotube (SWNT) chemical sensor (adapted from [1.7]), and (c) functionalized nanoparticles for cancer detection and drug delivery (adapted from [1.12]).

In this chapter, we will first discuss history and early research expenditures in relation to nanotechnology followed by governance of the National Nanotechnology Initiative, nanotechnology research and development (R&D) funding since 2001, worldwide R&D investments, nanoscience and nanotechnology education, and a summary and outlook. Finally, we will present the organization of the Handbook.

1.2 History and Early Research Expenditures

On 29 Dec. 1959 at the California Institute of Technology, Nobel Laureate Richard P. Feynman gave a talk at the Annual Meeting of the American Physical Society that has become one of the classic science lectures of the 20th century, titled, *There's Plenty of Room at the Bottom* [1.13]. He presented a technological vision of extreme miniaturization, several years before the word *chip* became part of the lexicon. He talked about the problem of manipulating and controlling things on a small scale. Extrapolating from known physical laws, Feynman envisioned a technology using the ultimate toolbox of nature, building nano-objects atom-by-atom or molecule-by-molecule. Since the 1980s, many inventions and discoveries in the fabrication of nano-objects have been testaments to his vision.

In 1998, the National Science and Technology Council (NSTC) of the White House Office of Science and Technology Policy (OSTP) created the Interagency Working Group on Nanoscience, Engineering and Technology (IWGN), recognizing the reality of the

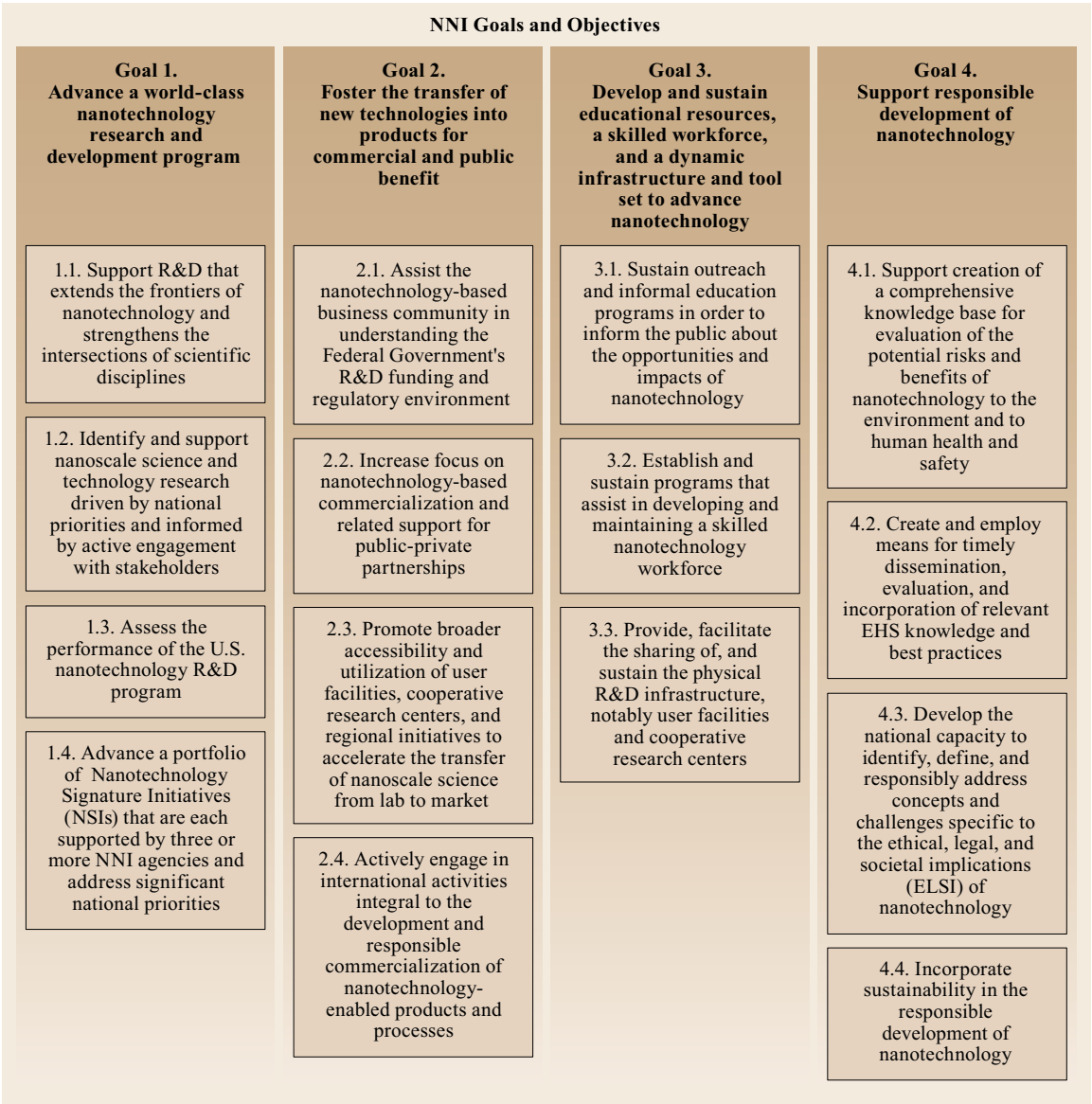


Fig. 1.3 A summary of NNI goals and objectives from the 2014 NNI Strategic Plan (after [1.14])

proliferation of nanotechnology. In 1999, the IWGN was elevated to the Subcommittee on Nanoscale Science, Engineering, & Technology (NSET). In a January 2000 speech also at the California Institute of Technology, President William J. Clinton talked about the exciting promise of nanotechnology and the importance of expanding research in nanoscale science and technology more broadly. Later that month in his State of the Union Address, he announced an ambitious $497 million federal, multiagency National Nanotechnology Initiative (NNI) in the FY (fiscal year) 2001 budget, and made the NNI a top science and technology priority [1.2]. The objective of this initiative was to form a broad-based coalition in which academia, the private sector, and local, state, and federal governments work together to push the envelope of nanoscience and nanoengineering to reap nanotechnology's potential social and economic benefits.

In Jan. 2003, following the creation of the NNI, the US Senate in the 108th Congress introduced a bill to establish a National Nanotechnology Program. On 3 Dec. 2003, President George W. Bush signed into law the 21st Century Nanotechnology Research and Development Act (NRDA) (Public Law 108–153). The legislation put into law programs and activities supported by the NNI. The bill gave nanotechnology a permanent home in the federal government and authorized $3.7 billion to be spent in the four year period beginning in October 2005. This allocation was for nanotechnology initiatives at six federal agencies: National Science Foundation (NSF), Department of Defense (DOD), Department of Energy (DOE), National Aeronautics and Space Administration (NASA), National Institute of Standards and Technology (NIST), and the National Institutes of Health (NIH). The funds were used to provide grants to researchers, coordinate R&D across the six agencies, establish interdisciplinary research centers, and accelerate technology transfer into the private sector.

Nanotechnology R&D is directed towards understanding and controlling matter at the nanoscale. NNI has four goals [1.14]:

1. Advance world-class nanotechnology research and development
2. Foster the transfer of new technologies into products for commercial and public benefit
3. Develop and sustain educational resources, a skilled workforce, and the supporting infrastructure and tools to advance nanotechnology; and
4. Support the responsible development of nanotechnology.

Further details are presented in Fig. 1.3.

Regarding international activities in early stages, in 2002, the European Union (EU) made nanosciences and nanotechnologies a priority in the Sixth Framework Program (FP6) for the period of 2003–2006. They had also dedicated small funds in the FP4 and FP5 programs before. FP6 was tailored to help better structure European research and to cope with the strategic objectives set out in the Lisbon Strategy Development Plan of 2000. Japan identified nanotechnology as one of its main research priorities in 2001. The funding levels increased from $400 million in 2001 to around $950 million in 2004. In 2003, South Korea embarked upon a ten-year program with around $2 billion in public funding, and Taiwan had committed around $600 million in public funding over six years. Russia, Singapore, and China also started to invest on a large scale.

1.3 Governance of the National Nanotechnology Initiative

In 2001, OSTP established a National Nanotechnology Coordination Office (NNCO), which was funded by NSET. The NNCO serves as a central point of contact for Federal nanotechnology R&D activities, and provides public outreach on behalf of the NNI. The NNCO Director and Deputy Director are appointed by the White House Co-Chair of the NSTC Committee on Technology. The NNCO coordinates the preparation and publication of interagency planning, budget, and various documents, such as the NNI supplement to the President's budget. The 21st Century Nanotechnology Research and Development Act called for a National Nanotechnology Advisory Panel (NNAP), to review the NNI triennially. In 2004, PCAST was designated as that panel. In addition, the Act had called for a triennial review by the National Academies.

NNI investments are guided by an annual strategic plan published by the OSTP. Figure 1.4 shows the organizational structure for management of NNI and OSTP coordination and assessment of NNI efforts. The NNI reporting requirements overseen by OSTP include:

1. The triennial strategic plan (Year 1)
2. The triennial assessment by PCAST (Year 2)
3. The triennial assessment of the program by the National Academies (Year 3); and
4. The annual supplemental report submitted with the President's budget.

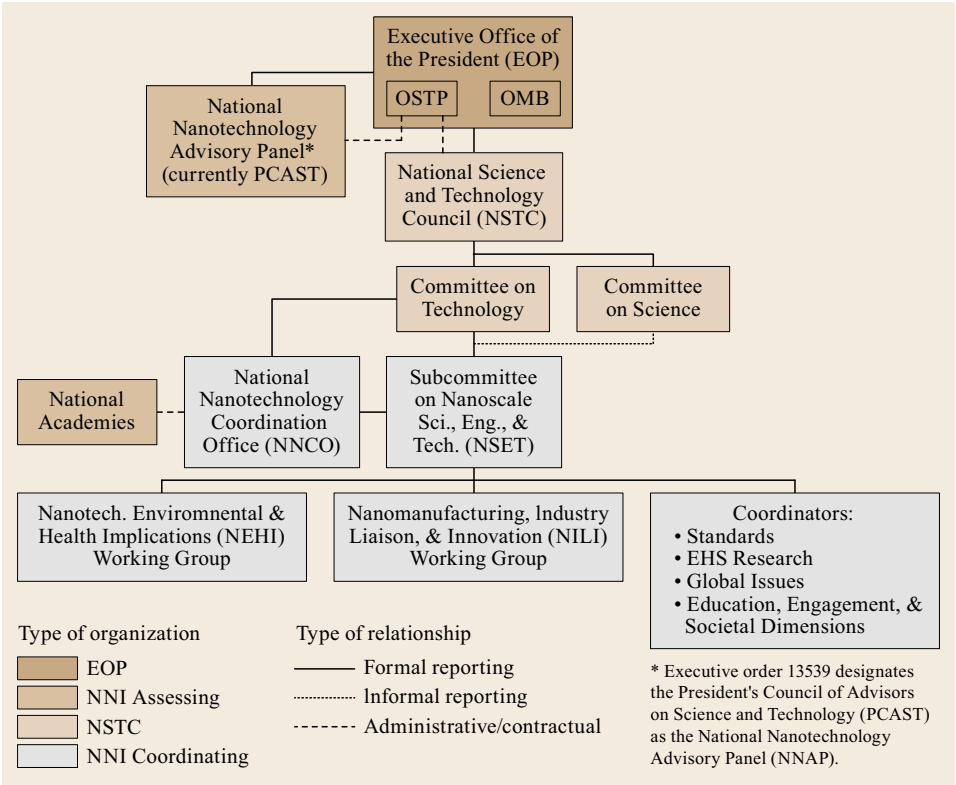


Fig. 1.4 Organizational structure for management of NNI (after [1.14])

It should be noted that since the enactment of the Nanotechnology Act on 3 Dec. 2003, it has not been reauthorized by US Congress. Reauthorization sets the policy for how funding should be spent by the various agencies that carry out the vision of the Nanotechnology Act. While the various funding agencies continue to fund under their current authority, no new direction for that funding has been set.

1.4 Nanotechnology R&D Funding Since 2001

Since 2001, the US Congress has appropriated about $20 billion for nanotechnology R&D through FY 2015, with about $1.5 billion spent in FY 2015 alone. Figure 1.5 shows NNI funding growth from FY 2001 through FY 2015. Twenty top-level federal agencies (27 including all subsidiary subagencies) fund NNI-related research, as shown in Table 1.2. Table 1.3 shows the breakdown of the NNI budget by agency for the period of FY 2001 to FY 2015. About 93% of the funding is spent by five agencies –NIH (28.7%), NSF (26.8%), DOE (22.3%), DOD (9.4%), and NIST (5.4%). US private sector R&D funding, presented in Table 1.4, is more than double that of federal and state funding with a focus on translating fundamental research into commercial products.

The NNI strategic plan for FY 2015 was published in February 2014 by OSTP [1.14]. In this plan, priorities for NNI funding by Program Component Area were:

1. Foundational Research (35% of NNI funding)
2. Applications, Devices, and Systems (24%)
3. Signature Initiatives (19%)
4. Infrastructure & Implementation (16%); and
5. Environment, Health, and Safety (EHS) (7%).

Five areas, called Nanotechnology Signature Initiatives (NSIs), were identified for targeted program-level interagency collaboration. These NSIs were intended to enable the rapid advancement of selected science and technology in the service of national security, economic, and environmental goals by focusing resources on critical challenges and R&D gaps [1.14, 16]. These areas were:

- *Nanotechnology for Solar Energy Collection and Conversion: Contributing to Energy Solutions for the Future*: This NSI will utilize the unique physical phenomena that occur on the nanoscale to help

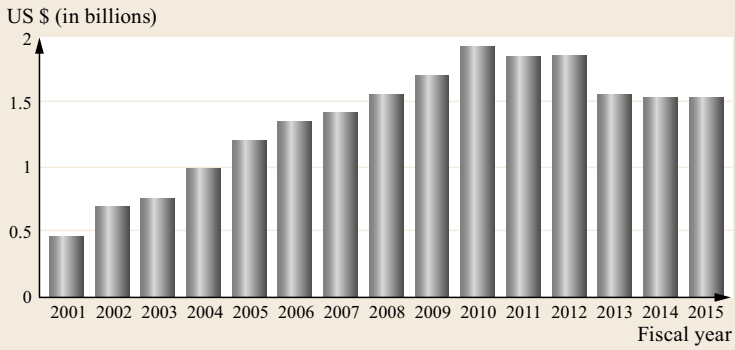


Fig. 1.5 US National Nanotechnology Initiative Funding, FY 2001-2015 (after [1.3, 15])

Table 1.2 Federal Departments and Agencies Participating in the NNI (after [1.14])

A: Eleven Federal Departments and Independent Agencies and Commissions with Nanotechnology R&D Budgets:
Consumer Product Safety Commission (CPSC)
Department of Commerce (DOC)
National Institute of Standards and Technology (NIST)
Department of Defense (DOD)
Department of Energy (DOE)
Department of Health and Human Services (DHHS)
Food and Drug Administration (FDA)
National Institute for Occupational Safety and Health (NIOSH)
National Institute for Health (NIH)
Department of Homeland Security (DHS)
Department of Transportation (DOT)
Federal Highway Administration (FHWA)
Environmental Protection Agency (EPA)
National Aeronautics and Space Administration (NASA)
National Science Foundation (NSF)
US Department of Agriculture (USDA)
Agricultural Research Service (ARS)
Forest Service (FS)
National Institute of Food and Agriculture (NIFA)
B: Nine Other Participating Departments and Independent Agencies and Commissions:
Department of Education (DOEd)
Department of the Interior (DOI)
US Geological Survey (USGS)
Department of Justice (DOJ)
National Institute of Justice (NIJ)
Department of Labor (DOL)
Occupational Safety and Health Administration (OSHA)
Department of State (DOS)
Department of the Treasury (DoTreas)
Intelligence Community (IC)
Office of the Director of National Intelligence (ODNI)
Nuclear Regulatory Commission (NRC)
US International Trade Commission (USITC)
C: Also participating from the Department of Commerce (DOC), listed above
Bureau of Industry and Security (BIS)
Economic Development Administration (EDA)
US Patent and Trademark Office (USPTO)

Table 1.3 NNI Investments for Various Funding Agencies from FY 2001 to FY 2015 (request). About 93% of the funding is spent by five agencies shown in bold. Data provided by NNCO (after [1.15])

National Nanotechnology Initiative Investments by Agency FY 2001–2015 (dollars in millions)																
FY	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009[b]**	**2010**	**2011**	**2012**	**2013**	**2014[c]**	**2015[d]**	**Total[a]**
CPSC	0	0	0	0	0	0	0	0	0.2	0.5	1.8	2.0	1.3	2.0	2.0	9.7
DHS	0	2	1	1	1	1.5	2.0	3.2	9.1	21.9	9.0	18.7	14.0	24.0	32.4	140.7
DOC/NIST	**33**	**77**	**64**	**77**	**79**	**77.9**	**87.6**	**85.6**	**93.4**	**114.7**	**95.9**	**95.4**	**91.4**	**97.8**	**82.6**	**1252.3**
DOD	**125**	**224**	**220**	**291**	**352**	**423.9**	**450.2**	**460.4**	**459**	**439.6**	**425.3**	**426.1**	**170.1**	**175.9**	**144**	**4786.4**
DOE	**88**	**89**	**134**	**202**	**208**	**231**	**236**	**244.7**	**332.6**	**373.8**	**346.2**	**313.8**	**314.2**	**303.3**	**343.1**	**3759.7**
DOJ	1	1	1	2	2	0.3	1.7	0.1	1.2	0.2	0	0	0	0	0	10.5
DOT	0	0	0	0	0	0.9	0.9	0.9	0.9	3.2	1.0	1.0	2.4	2.0	1.5	14.7
EPA	5	6	5	5	7	4.5	7.6	12.1	11.6	17.7	17.4	17.5	14.6	15.5	16.8	163.3
DHHS (tot)	40	59	78	106	168	195.4	222.7	311.4	356.0	472.6	428.6	479.6	485.4	469.5	469.6	4341.7
FDA	0	0	0	0	0	0	0	0	6.5	7.3	9.9	13.6	16.1	17.0	17.0	87.4
NIH	**40**	**59**	**78**	**106**	**165**	**191.6**	**215.4**	**304.5**	**342.8**	**456.8**	**408.6**	**456.0**	**458.8**	**441.5**	**441.5**	**4165.4**
NIOSH	0	0	0	0	3	3.8	7.3	6.9	6.7	8.5	10.0	10.0	10.5	11.0	11.1	88.8
NASA	22	35	36	47	45	50	19.8	17.4	13.7	19.7	17	18.6	16.4	17.9	13.7	389.2
NSF	**150**	**204**	**221**	**256**	**335**	**359.7**	**388.8**	**408.6**	**408.6**	**428.7**	**485.1**	**466.3**	**421.0**	**410.6**	**412.4**	**5355.8**
USDA (tot)	0	0	0	2	3	6.2	6.8	10.1	15.3	20.3	20.0	18.3	19.5	19.1	18.8	159.4
ARS	0	0	0	0	0	0	0	0	0	0	0	2.0	2.0	2.0	2.0	8.0
FS	0	0	0	0	0	2.3	2.9	4.6	5.4	7.1	10.0	5.0	5.0	4.0	4.0	50.3
NIFA	0	0	0	2	3	3.6	3.9	5.5	9.9	13.2	10.0	11.3	12.5	13.1	12.8	101.1
Total[b]	**464**	**697**	**760**	**989**	**1200**	**1351.3**	**1424.1**	**1554.5**	**1701.5**	**1912.8**	**1847.3**	**1857.3**	**1550.2**	**1537.5**	**1536.9**	**20 383.4**

[a] NOTE: Totals may not add due to rounding
[b] NOT including American Recovery and Reinvestment Act funds for NSF ($101 M), DOE ($239 M), NIST ($43 M), and NIH ($73 M)
[c] FY '14 estimated based on 2014 enacted levels and may shift as operating plans are finalized
[d] FY '15 Request

overcome current performance barriers and substantially improve the collection and conversion of solar energy. Participating agencies include: DOE, NIST, NSF, DOD, Intelligence Community (IC), US Department of Agriculture (USDA), and the National Institute of Food and Agriculture (NIFA).

- *Sustainable Nanomanufacturing: Creating the Industries of the Future*: This NSI will establish manufacturing technologies for economical and sustainable integration of nanoscale building blocks into complex, large-scale systems by supporting product, tool, and process design informed by and adhering to the overall constraints of safety, sustainability, and scalability. Participating agencies include: NIST, NSF, DOE, Environmental Protection Agency (EPA), IC, NIH, National Institute for Occupational Safety and Health (NIOSH), Occupational Safety and Health Administration (OSHA), and USDA/Forestry Service (FS).
- *Nanoelectronics for 2020 and Beyond*: This NSI is aimed at discovering and using novel nanoscale fabrication processes and innovative concepts to produce revolutionary materials, devices, systems, and architectures to advance the field of nanoelectronics. Participating agencies include: NSF, DOD, NIST, DOE, and IC.
- *Nanotechnology Knowledge Infrastructure (NKI): Enabling National Leadership in Sustainable Design*: This NSI will coordinate the nanoscale science, engineering, and technology communities around the fundamental, interconnected elements of collaborative modeling, a cyber-toolbox, and data infrastructure that will capitalize on American strengths in innovation, shorten the time from research to new product development, and maintain US leadership in sustainable design of engineered nanoscale materials. Participating agencies include: Consumer Product Safety Commission (CPSC), DOD, DOE, EPA, Food and Drug Administration (FDA), NASA, NIOSH, NIST, NSF, OSHA.
- *Nanotechnology for Sensors and Sensors for Nanotechnology: Improving and Protecting Health, Safety, and the Environment*: This NSI has considerable potential for nanotechnology to open the door to the development of inexpensive, portable devices that can rapidly detect, identify, and quantify biological and chemical substances. Participating agencies include: CPSC, DOD, EPA, FDA, NASA, NIH, NIOSH, NIST, NSF, USDA.

In the international context, Federal investment in FY 2014 and 2015 by the US of about $1.5 B per year is slightly higher than that by the EU or Japan. In 2014, the EU launched Horizon 2020 (2014-2020) after the conclusion of the EU FP7, with a total budget of €77 B. In Horizon 2020, the EU identified six key areas: nanotechnology, advanced manufacturing, advanced materials, nanoelectronics, photonics, and biotechnology, and allocated €6.6 B for these areas over a six year period; about €1.1 B per year. An additional €5 B was planned for public-private partnerships. It should be noted that, in addition to EU expenditure, individual countries within the EU invest additional money on research. In 2014, Japan planned to spend ¥550 B over the following five year period on nanotechnology, which translates to about $1 B per year. Japanese companies are reported to invest about 90% of total R&D in nanotechnology.

1.5 Worldwide R&D Investments and Output

1.5.1 R&D Investments

Figure 1.6 shows Federal R&D funding from FY 1997 to FY 2009 for the US, EU, Japan, and others. The funding has steadily grown for all countries. The US and the EU have made comparable investments, followed by Japan, China, Korea, and Taiwan. A substantial change in global investment rate is observed around 2000 after introduction of the NNI and around 2006 due to the introduction of second-generation nanotechnology products. The figure also shows the specific Nano R&D 2008 in dollars per capita. The numbers for the US, EU, and Korea are comparable, with Japan being higher. Figure 1.7 shows industrial spending for various countries in 2012. The US spent far more than other countries. However, the rest of the world, particularly emerging economies such as China, have been increasing their funding both at the national government and industrial levels at a rapid rate.

1.5.2 Output

Limited current US data exists on the return on investment in nanotechnology as it relates to scientific output, the number of jobs created, and product revenues [1.17–22]). The European Commission Research Directorate General, Brussels, and the US Organization for Economic Cooperation and Development (OECD) have published some data on these subjects, though individual US federal agencies have not published much

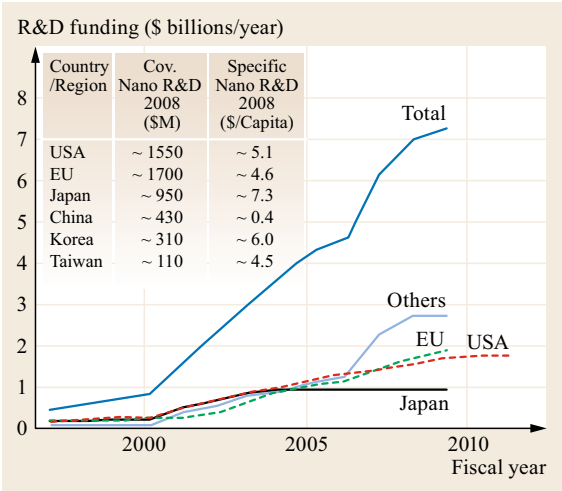


Fig. 1.6 FY 1997-2009 Federal/national government R&D funding. Specific nanotechnology R&D per capita is based on the national nanotechnology expenditures and effective expenditure for all other R&D programs (after [1.3, 17])

data themselves. These data are critical for tracking performance and for planning purposes. They are also critical for Congress to make educated decisions on reprioritization and total investments in nanotechnology R&D. In this section, we present an overview of the US data that exists in the public domain.

Figure 1.8a shows the number of nanotechnology publications in archival journals in the SCI (Science Citation Index) database between 1990 and 2013 for the US and other countries, and Fig. 1.8b,c shows the number of nanotechnology patents published for the US and other countries. There is a steady growth in the number of publications for each country, although unevenly distributed among the various countries. The metrics of patents published in three or more countries is a more representative indicator of significance compared to those patents published in just one or two countries. It should be noted that although the rate of patents published has decreased over the past decade, the cumulative growth has continued. Additional data are provided by *Roco* [1.17], and *Chen* et al. [1.19]. Figure 1.9 gives the breakdown of the number of patents issued by the US Patent and Trademark Office (USPTO) by technology area as of 31 December 2012, type of invention in 2012, and by country of residence of the first-named inventor taken from patent applications.

R&D funding breakdown and nanoenabled product revenues for various regions in 2012 are presented in Tables 1.4 and 1.5. In 2012, for the US alone, nanoenabled product revenues were $236 billion for a modest annual federal R&D investment of $1.5 billion, or a total public/private investment of $6.6 billion. This represents an impressive return on investment.

Table 1.6 presents a summary of six key indicators of nanotechnology R&D investment and output in the world and the US between 2000 and 2010. The key indicators include:

1. R&D funding (public and private)
2. Venture capital
3. SCI papers
4. Patent applications
5. People – primary workforce, and
6. Final products market.

A growth rate of the key nanotechnology indicators of 16–33% is observed. It is noted that the global market for nanomaterials has been reported to be tens of billions of dollars in FY 2015 with a rapid growth rate. The global revenues of nanoenabled products and nanomaterials are approaching $1 trillion.

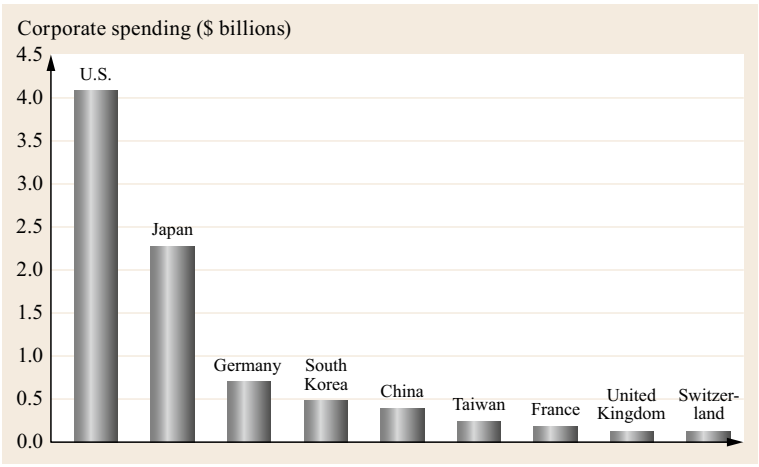


Fig. 1.7 Industrial R&D spending in 2012 for industry focused on applied R&D for first & second generations of nanotechnology products (after [1.3, 18])

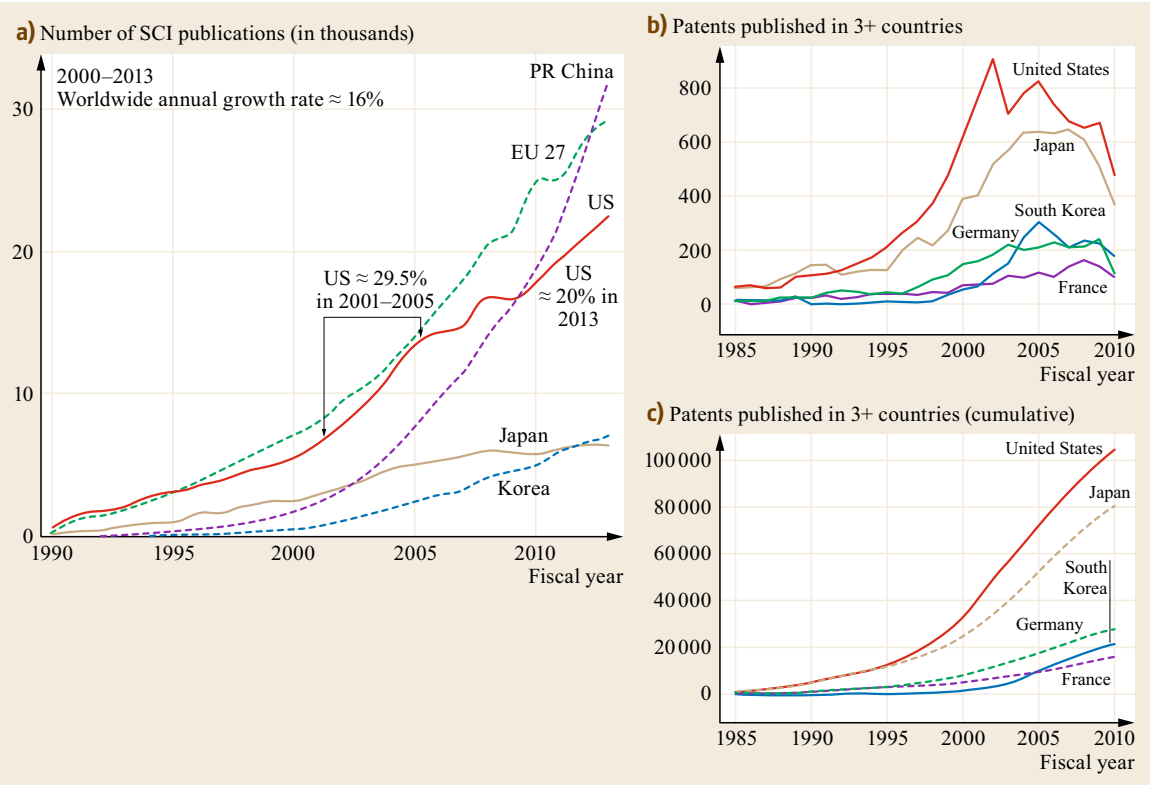


Fig. 1.8 **(a)** Number of nanotechnology SCI publications from 1990 to 2013. Data was generated from an online search in the Web of Science using a *title-abstract* search in the SCI database for nanotechnology by keywords (after [1.3, 18]). **(b,c)** Number of nanotechnology patents published from 1985 to 2010 for the US and other countries. The metrics of patents published in three or more countries is a more representative indicator of significance (after [1.3, 15])

Table 1.4 R&D funding data across all industrial sectors in various regions in 2012, data collected by Lux Research (after [1.3, 15, 21])

US	Federal and state	$ 2.1 billion
	Industry	$ 4.0 billion
	Venture capital investors	$ 0.5 billion
	Total	$ 6.6 billion
Worldwide	Federal, state, industry, and venture capital investors	$18.5 billion

Table 1.5 Nanoenabled Product Revenues data across all industrial sectors in various regions in 2012, data collected by Lux Research (after [1.3, 21])

US	$236 billion
Europe	$235 billion
Asia	$213 billion
Rest of the world	$ 47 billion
Total	$731 billion

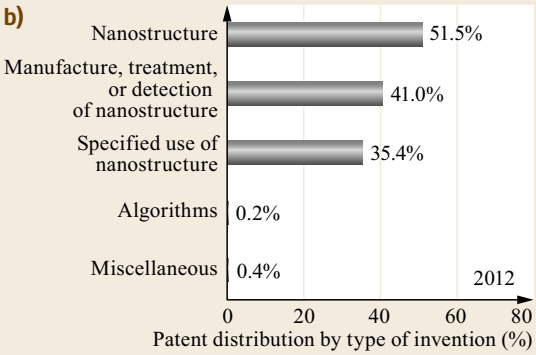
Table 1.6 Summary of six key indicators of nanotechnology R&D investment and output in the world and the US during 2000-2010 and projections for 2015 (after [1.3, 18]) (Venture Capital estimations were made by Lux Research)

World/US	R&D funding (public & private)	Venture capital	SCI papers	Patent applications	People – primary workforce	Final products market
2000	≈ $1.2 B/$0.37 B	≈ $0.21 B/$0.17 B	18 085/5342	1197/405	≈ 60 000/25 000	≈ $30 B/$13 B
2010	≈ $18 B/$4.1 B	≈ $1.3 B/$1.0 B	78 842/17 978	≈ 20 000/5000	≈ 600 000/ 220 000	≈ $300 B/$110 B
2000–2010 (aver. growth)	≈ 31%/ ≈ 27%	≈ 30%/ ≈ 35%	≈ 16%/ ≈ 13%	≈ 33%/ ≈ 28%	≈ 25%/ ≈ 23%	≈ 25%/ ≈ 24%
2015 (projected)					≈ 2 000 000/ 800 000	≈ $1000 B/ $400 B

a)

Area of technology	No. of Patents
Biotechnology and organic chemistry	1231
Chemical and materials engineering	2768
Computer architecture software and information security	28
Computer networking, network security, cryptography	5
Communications	259
Semiconductor, electrical, optical systems	3594
Transportation, construction, electronic commerce	66
Mechanical engineering, manufacturing and products	433
Total	8384

b)



c) Patent distribution by residence country (%)

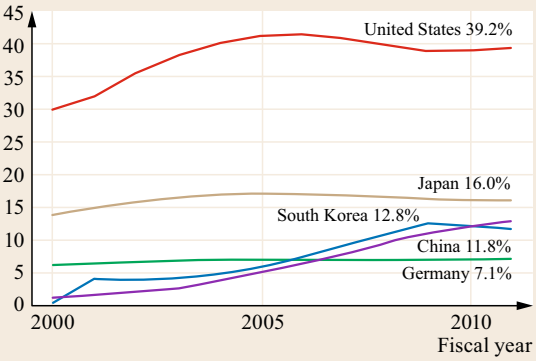


Fig. 1.9a–c The breakdown of the number of patents issued by the USPTO based on **(a)** technology area, as of December 31, 2012, **(b)** type of invention in 2012, and **(c)** residence – country of first-named inventor (after [1.3, 23])

Figure 1.10 shows the primary applications markets for nanoenabled products based on reviews gathered

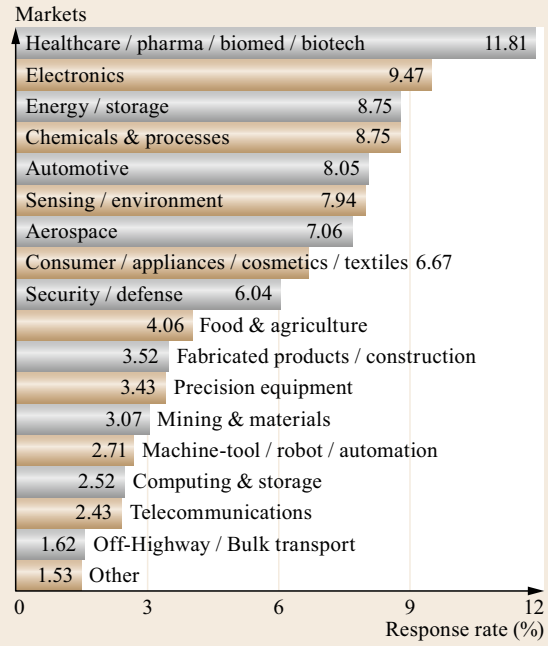


Fig. 1.10 Primary application markets in 2014 for nanoenabled products based on reviews as ranked in a survey (after [1.3, 24])

from 300 respondents in a 2014 survey project conducted by the National Center for Manufacturing Sciences (NCMS), Ann Arbor, Michigan. The data shows that some industries pursue simultaneous pathways for penetrating multiple markets and end users for their products. The largest application markets are listed below in decreasing order of response rate (percentage):

- Healthcare, pharmaceuticals, biomedical applications, and biotechnological applications
- Electronics
- Energy and energy storage
- Chemical and processes
- Automotive
- Sensing and environment, and
- Aerospace.

In summary, the US remains a global leader in both investments and output, as measured by the number of publications and patents, jobs, and nanoenabled product revenue. However, other countries are slowly catching up and competing with the US, including China which is rapidly increasing its share of investment.

1.6 Translation of Basic Research to Technology Commercialization

Research and some development occurs in academic labs, primarily funded by federal and state investments. Translational research is funded jointly by federal/state and private funding. Product launch is then carried out by startups and funded by venture capitalists or industry. Federal R&D agencies and public-private partnerships need to invest in technology transfer and developing research into a successful product. Technology transfer will drive the economy and facilitate in multiplying the investment in basic research. The move of US manufacturing to locations offshore and around the world has given rise to the idea that we *invent here and manufacture there*. This exists because the US economy is not able to provide the necessary investments and incentives for venture capitalists and industries to continue both development and manufacturing in the US [1.25].

Various concerns regarding the NNI have been raised in several reports on the mandated NNI Triennial Reviews; a recent one prepared by The National Academies on the Triennial Review in 2013 [1.26]. Major concerns include lack of information at the project level on who is performing research and its implications; poor planning, management, and coordination; lack of metrics for assessing progress; and lack of official data on return on investments. The Government Accountability Office (GAO) conducted a study and issued a report in 2014 entitled *Nanomanufacturing: Emergence and Implications for US Competitiveness, the Environment, and Human Health* [1.27]. It examined current issues related to nanotechnology and nanomanufacturing. This report also identified various concerns in regard to NNI implementation, including the valley of death – gaps in funding or support for technology development and manufacturing development (for a schematic, see Fig. 1.11). The report also stated that there is a lack of participation in setting standards for nanomanufacturing and nanotechnology, a lack of national vision for nanomanufacturing capability, and a need for an integrated framework to help assess and address the EHS implications.

The PCAST report on the Triennial Review issued in 2014 [1.15] reiterated many critiques advanced by the National Academies and GAO reviews. It stated that the NNI has reached a turning point. Although continued support of fundamental research in nanotechnology is critical, the transition toward commercialization is encouraged to translate the technologies and investments into commercial products and revenues. This may affect a range of technology areas, including drug delivery, energy technology, smart sensors, clean water, and quantum computing.

The report presented four general constraints on technology commercialization [1.15]. It stated that first-time academic entrepreneurs have little training or experience in moving a technical innovation out of the research lab and into a company. The second constraint has to do with a lack of marketing knowledge, i. e., how their innovation might address strategic goals of a company or could be translated into a successful product. Third, given that venture capitalists prefer to back experienced entrepreneurs, fund raising for newcomers becomes more difficult. Finally, an academic may not be allowed by the peer review process to pursue high-risk, high-return ideas. The report made recommendations to address all these constraints.

The report argued that the NSIs being used are shaped by coordination of three or more federal agencies around a topic. Instead, it recommends creation of *Grand Challenges*, which are large, outward-facing efforts with specific, measurable goals. A Grand Challenge has a well-defined technical goal and it addresses

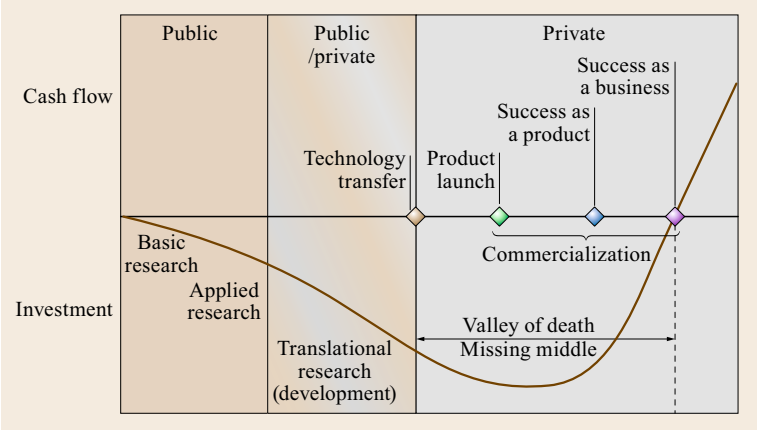


Fig. 1.11 Schematic showing the *valley of death* scenario (after [1.25])

an issue of significant societal impact. It has a measurable end-point with clear milestones en route.

There needs to be a focus on EHS as the use of nanotechnology products, particularly nanomaterials, is perceived to have potential EHS risks when exposed during manufacturing, use, or disposal. Established tools and techniques to study EHS risks do not exist for nanotechnology products. In addition, it is difficult to assess risks because nanomaterials may differ in size, shape, and surface chemistry. NIOSH is responsible for developing recommended exposure limits and necessary protective actions for nanomaterials [1.28–31].

1.7 Nanoscience and Nanotechnology Education

Education and training are essential to produce a new generation of scientists, engineers, and skilled workers with the flexible and interdisciplinary R&D approach necessary for rapid progress in nanosciences and nanotechnology. A trained nanotechnology workforce at both vocational and professional levels needs to be developed. This needs to be accomplished with both formal and informal education efforts, starting with the foundations in children and increasing in technical and specialized education through the university and continuing education levels.

Nanoscience and nanoetechnology education is concerned with five general topics related to this necessity: interdisciplinary education; K-12 education, excitement, and creativity; vocational, undergraduate, graduate, and continuing education programs; educator training; and informal science education [1.32].

1.7.1 Interdisciplinary Education

With decreasing numbers of people choosing to study science and engineering and with the rapid progress in nanoscience and nanotechnology demanding an increasing number of people, the problem of a trained workforce is expected to be acute. The question is being asked – is the traditional separation of academic disciplines into physics, chemistry, biology, and various engineering disciplines meaningful at the nanolevel? Arguably, nanoscience and nanoetechnology education should begin by covering the broad science topics in order to facilitate comprehensive understanding of foundational science. Generic skills and entrepreneurship are also needed to transfer scientific knowledge into products, increasing the importance and support for scientific education. Therefore, fundamental understanding of business, society, politics, and health and safety are required, with preparation to work in a global environment and solve global problems. Thus, nanoscience and nanoetechnology education should focus on a broad, interdisciplinary approach to science and engineering at the nanoscale.

1.7.2 K–12 Education, Excitement, and Creativity

To increase the pool of students interested in advanced education in science and technology, science needs to be conveyed as an exciting and creative endeavor from its first introduction. Innovative foundational education in traditional education environments, such as schools and supplementary instruction opportunities, can help to accomplish this goal. In order to encourage excitement and creativity in the fields of nanoscience and nanoetechnology, K-12 education needs to take a cross-disciplinary approach that includes combining arts and science.

To this end, since the mid-1990s, there has been a global push for science, technology, engineering, and mathematics (STEM) education, particularly in the K-12 arena. However, this curriculum, while rich in content, is lacking in critical and creative thinking and real-world application skills. However, since the early 2010s, there has been a drive, called STEAM, to include the arts in STEM education to encourage that creativity and critical thinking skills derived from arts-based education as well. To develop interdisciplinary curricula relevant for nanoscience and nanotechnology, a revamping of the STEAM education model for nanoscience, development of new courses and course material, and training of new instructors is required.

1.7.3 Vocational, Undergraduate, Graduate, and Continuing Education Programs

In the US, a strong example of vocational-level programs is the one-semester, 18 credit-hour *capstone* style program pioneered by Pennsylvania State University (Penn State), designed to complete a student's vocational degree at a *home* institution by integrating nanoscience into their foundational classes. In addition, Penn State developed a network to develop a standardized curriculum to assist vocational institutions in those foundations, as well as to train practitioners. Contin-

ually identified areas of improvement for vocational training include a best practices manual, creating nationwide programs that ensure appropriate education and training, but are sufficiently flexible to account for local and regional technical needs.

The current state of undergraduate nanoscience education is confined mostly to opportunities for undergraduate research. For example, the US NSF funds a Research Experience for Undergraduates, providing students with nanoscience-specific training and research education, as well as networking and professional development opportunities. However, there is a need to promote interdisciplinary cooperation and societal impacts in existing university programs to increase diversity in student populations.

There are also deficiencies in graduate training for nanoscience, including interdisciplinary education and diversity, but these programs will improve as the pipeline of students funneling into the programs have better foundational nanoscience knowledge due to strengthening of nanoscience education from kindergarten through undergraduate training.

Continuing education is necessary to educate and update practitioners on new technology and best practices. Presently, there is a wealth of information available at a variety of sources on nanoscience and nanotechnology, but there is a need to assemble the information into learning packages or short courses for targeted training, increasing accessibility and rapid dissemination.

1.7.4 Educator Training

In order to accomplish the above education goals and further nanoscience and nanotechnology education, including attraction and engagement, educator training is essential. Presently, educator training is focused broadly on STEM fields, without, perhaps, specific training in nanoscience and nanotechnology.

1.7.5 Informal Science Education

Education for the general public on the benefits of nanoscience and nanotechnology is an important part of furthering nanoscience and nanotechnology education goals. This means supporting and engaging community-based opportunities for science and education, and making them approachable, affordable, inclusive, and interesting. Currently, this is being done mostly by a mix of science museums [1.33] and university groups. In the US, these efforts are largely funded by the National Science Foundation. In Germany, efforts are supported by the Federal Ministry of Education and Research, who have a traveling exhibit that visits schools to supplement the formal science curriculum.

1.8 Summary and Outlook

Nanotechnology refers to any technology done on a nanoscale that has applications in the real world. The properties of matter at the nanoscale are different from those at a larger scale. The unique physical and chemical properties of nanomaterials can be exploited for commercial applications that benefit society. The discovery of novel materials, processes, and phenomena at the nanoscale and the development of new experimental and theoretical techniques for research at the end of the 20th century, provide fresh opportunities for the development of innovative nanosystems and nanomaterials. Nanotechnology represents a *megatrend*, and has become a *general purpose* technology.

The NNI was launched in the US in 2000 with a $497 million investment in the FY 2001 budget. In 2003, a 21st Century Nanotechnology R&D Act became law in the US, with $3.7 billion to be spent in the four year period beginning in October 2005 over six federal agencies. Through FY 2015, about $20 billion was spent by 20 federal agencies, with about $1.5 billion for FY 2015. Twenty top-level federal agencies (27 including all subsidiary and subagencies) fund NNI-related research with about 93% of the funding being spent by five agencies (NIH, NSF, DOE, DOD, and NIST). The US private sector R&D investment is estimated to be twice that of public funding, with a focus on translating fundamental research into commercial products.

NNI is governed by an intricate organizational structure overseen by OSTP. NNCO has the responsibility to coordinate R&D activities among various federal agencies. Congress mandates various Triennial Review reports, including an external triennial assessment by the National Academies and another by PCAST, which reports to OSTP. It should be noted that since the enactment of the Nanotechnology Act in 2003, it has not been reauthorized by US Congress.

The US and the rest of the world continue to make significant investments. As of FY 2015, the US and the EU have made comparable government investments, followed by Japan, China, Korea, and Taiwan. US private industry has spent far more than industry in other countries. However, the rest of the world,

particularly emerging economies such as China, are increasing funding at a rapid rate. On the basis of key indicators of nanotechnology development, which include R&D funding, venture capital, SCI papers, patent applications, primary workforce, and final product market, a significant growth rate in all of these indicators is observed worldwide. The revenue from nanoenabled products continues to grow, with over $200 billion in 2012 in the US alone. This represents an impressive return on R&D investment.

As an outlook, there appears to be a lack of investment in technology transfer. Since 2001, as expected, the major broad-based focus has been on foundational research. NNI is believed to have reached a turning point. One should assess the scientific progress made and its economic impact. On the basis of this detailed analysis, scientific focus may need to be on only a few areas of economic importance. In addition, the NNI needs to develop strategies for investment in translational research to facilitate commercialization, as well as research on EHS-related issues to protect public life and the environment. Finally, development of the nanotechnology workforce and curriculum development, and programs to educate the public of the benefits of nanotechnology and to change attitudes toward nanotechnology should be continued.

1.9 Organization of the Handbook

This handbook integrates knowledge of nanofabrication, materials, applications, and reliability. The organization of the handbook is straightforward. The handbook is divided into ten parts. The first two parts include an introduction to micro/nanofabrication techniques, nanomaterials, and nanostructures. The third and fourth parts introduce various MEMS/NEMS and BioMEMS/BioNEMS devices. The fifth part introduces various nanometrology techniques. The sixth part provides an overview of bio/nanotribology and bio/nanomechanics, which will prepare the reader to understand interface reliability in industrial applications. The seventh part provides an overview of molecularly thick films for lubrication. The eighth part focuses on the emerging field of biomimetics and bioinspired surfaces, in which one mimics nature to develop products and processes of interest. The ninth part focuses on industrial applications and micro/nanodevice reliability. The tenth part focuses on nanotechnology in the social context, environment, health and safety issues, and nanotechnology education.

References

1.1 National Science and Technology Council: *Nanotechnology: Shaping the World Atom by Atom* (National Science and Technology Council, Committee on Technology, The Interagency Working Group on Nanoscience, Engineering, and Technology, Washington, D.C. 1999)

1.2 National Science and Technology Council: *National Nanotechnology Initiative: Leading to the Next Industrial Revolution* (National Science and Technology Council, Committee on Technology, The Interagency Working Group on Nanoscience, Engineering, and Technology, Washington, D.C. 2000)

1.3 B. Bhushan: Governance, policy, and legislation of nanotechnology: A perspective, Microsyst. Technol. **21**, 1137–1155 (2015)

1.4 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer, Switzerland 2016)

1.5 B. Bhushan, D. Luo, S.R. Schricker, W. Sigmund, S. Zauscher: *Handbook of Nanomaterials Properties* (Springer, Heidelberg 2014)

1.6 M.J. Madou: *Fundamentals of Microfabrication and Nanotechnology*, 3rd edn. (CRC Press, Boca Raton 2011)

1.7 R.J. Chen, H.C. Choi, S. Bangsaruntip, E. Yenilmex, X. Tang, Q. Wang, Y.L. Chang, H. Dai: An investigation of the mechanisms of electrode sensing of protein adsorption on carbon nanotube devices, J. Am. Chem. Soc. **126**, 1563–1568 (2004)

1.8 D. Srivastava: Computational nanotechnology of carbon nanotubes. In: *Carbon Nanotubes: Science and Applications*, ed. by M. Meyyappan (CRC, Boca Raton 2004) pp. 25–36

1.9 W.G. van der Wiel, S. De Franceschi, J.M. Elzerman, T. Fugisawa, S. Tarucha, L.P. Kauwenhoven: Electron transport through double quantum dots, Rev. Mod. Phys. **75**, 1–22 (2003)

1.10 L.J. Hornbeck: Digital light processing update – Status and future applications, SPIE **3634**, 158–170 (1999)

1.11 M.C. Roco, W.S. Bainbridge: The new world of discovery, invention, and innovation: Convergence of knowledge, technology, and society, J. Nanopart. Res. **15**, 1–17 (2013)

1.12 D.J. Irvine: Drug delivery: One nanoparticle, one kill, Nat. Mater. **10**, 342–343 (2011)

1.13 R.P. Feynman: There's plenty of room at the bottom, Eng. Sci. **23**, 22–36 (1960)

1.14 National Science and Technology Council: *National Nanotechnology Initiative Strategic Plan* (National Science and Technology Council Committee on Technology, Subcommittee on Nanoscale Science, Engi-

neering, and Technology, Washington, D.C. 2014)
1.15 Executive Office of the President: *Report to the President and Congress on the Fifth Assessment of the National Nanotechnology Initiative* (President's Council of Advisors on Science and Technology, Executive Office of the President, Washington, D.C. 2014)
1.16 National Science and Technology Council: *The National Nanotechnology Initiative – Supplement to the President's 2015 Budget* (National Science and Technology Council and the Office of Science & Technology Policy, Washington, D.C. 2014)
1.17 M.C. Roco: The long view of nanotechnology development: The national nanotechnology initiative at 10 years, J. Nanopart. Res. **13**, 427–445 (2011)
1.18 M. Roco: NSF and long-view for nanotechnology development. In: *10th Int. Nanotechnol. Conf. Commun. Cooperation (INC10), NIST, Gaithersburg* (2014)
1.19 H. Chen, M.C. Roco, J. Son, S. Jiang, C.A. Larson, Q. Gao: Global nanotechnology development from 1991 to 2012: Patents, scientific publications, and the effect of NSF funding, J. Nanopart. Res. **15**, 1–21 (2013)
1.20 S. Arora, A. Porter, J. Youtie, P. Shapira: Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs, Scientometrics **95**, 351–370 (2013)
1.21 H. Flynn: *Nanotechnology Update: Corporations Up Their Spending as Revenues for Nano-enabled Products Increase* (Lux Research, Boston 2014)
1.22 Y. Li, S. Arora, J. Youtie, P. Shapira, S. Carley: *Program on Nanotechnology Research and Innovation System Assessment* (Georgia Institute of Technology, Atlanta 2014)
1.23 J. Lorengo: Nanotechnology-related patenting: USPTO & Global statistics. In: *NanoBusiness Commercialization Association Annu. Meet., Philadelphia* (2013)
1.24 M. Mehta: 2014 NCMS Study of Nanotechnology in the U.S. Manufacturing Industry (National Center for Manufacturing Sciences (NCMS), Ann Arbor 2015), Unpublished data
1.25 B. Bhushan: Perspective: Science and technology policy – What is at stake and why should scientists participate?, Sci. Public. Policy **42**, 887–900 (2015)
1.26 National Research Council: *Triennial Review of the National Nanotechnology Initiative* (The National Academies Press, Washington, D.C. 2013)
1.27 U.S. Government Accountability Office: *Nanomanufacturing: Emergence and Implications for U.S. Competitiveness, the Environment, and Human Health* (Government Accountability Office, Washington, D.C. 2014) GAO-14-181SP
1.28 National Institute for Occupational Safety and Health: *Approaches to Safe Nanotechnology – Managing the Health and Safety Concerns Associated with Engineering Nanomaterials* (US Department of Health and Human Services, Centers for Disease Control and Prevention, Cincinnati 2009) DHHS (NIOSH) Publication No. 2009-125.
1.29 National Institute for Occupational Safety and Health: *Current Strategies for Engineering Controls in Nanomaterial Production and Downstream Handling Processes* (US Department of Health and Human Services, Centers for Disease Control and Prevention, Cincinnati 2013) DHHS (NIOSH) Publication No. 2014–102.
1.30 National Institute for Occupational Safety and Health: *Protecting the Nanotechnology Workforce: NIOSH Nanotechnology Research and Guidance Strategic Plan, 2013–2016* (US Department of Health and Human Services, Centers for Disease Control and Prevention, Cincinnati 2013) DHHS (NIOSH) Publication 2014–106
1.31 National Research Council: *Research Progress on Environmental, Health, and Safety Aspects of Engineering Materials* (The National Academies Press, Washington, D.C. 2013)
1.32 K. Winkelmann, B. Bhushan: *Global Perspectives of Nanoscience and Engineering Education* (Springer, Switzerland 2016)
1.33 W.C. Crone: *Bringing Nano to the Public: A Collaboration Opportunity for Researchers and Museums – NISE Network* (Science Museums of Minnesota, Minneapolis 2006)

Micro/Nano Part A

Part A Micro/Nanofabrication Techniques

2 **Molecule-Based Devices**
Françisco M. Raymo, Coral Gables, USA

3 **Introduction to Micro-/Nanofabrication**
Gemma Rius, Bellaterra, Spain
Antoni Baldi, Bellaterra, Spain
Babak Ziaie, West Lafayette, USA
Massood Z. Atashbar, Kalamazoo, USA

4 **3-D Nanostructure Fabrication by Focused-Ion Beam, Electron- and Laser Beam**
Shinji Matsui, Hyogo, Japan
Hiroaki Misawa, Sapporo, Japan
Quan Sun, Sapporo, Japan

5 **Nanoimprint Lithography**
Helmut Schift, Villigen PSI, Switzerland
Anders Kristensen, Kongens Lyngby, Denmark

6 **Stamping Techniques for Micro- and Nanofabrication**
John A. Rogers, Evanston, USA
Etienne Menard, Le Bourget du Lac, France

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems**
Christian A. Zorman, Cleveland, USA

2. Molecule-Based Devices

Françisco M. Raymo

The constituent components of conventional devices are carved out of larger materials relying on physical methods. This top-down approach to engineered building blocks becomes increasingly challenging as the dimensions of the target structures approach the nanoscale. Nature, on the other hand, assembles nanoscaled biomolecules relying on chemical strategies. Small molecular building blocks are joined to produce nanostructures with defined geometries and specific functions. It is becoming apparent that Nature's bottom-up approach to functional nanostructures can be mimicked to produce artificial molecules with nanoscaled dimensions and engineered properties. Indeed, examples of artificial nanohelices, nanotubes and molecular motors are starting to be developed. Some of these fascinating chemical systems have intriguing electrochemical and photochemical properties, which can be exploited to manipulate chemical, electrical and optical signals at the molecular level. This tremendous opportunity has led to the development of the molecular equivalent of conventional logic gates. Indeed, simple logic operations can be reproduced with collections of molecules operating in solution. Most of these chemical systems, however, rely on bulk addressing to execute combinational and sequential logic operations. It is essential to devise methods to reproduce these useful functions in solid-state configurations and, eventually, with single molecules. These challenging objectives are stimulating the design of clever devices that interface small assemblies of organic molecules with macroscaled and nanoscaled electrodes. These strategies have already produced rudimentary examples of diodes, switches and transistors based on functional molecular components. The rapid and continuous progress of this exploratory research will, hopefully, lead to an entire generation of molecule-based devices that might ultimately find useful applications in a variety of fields ranging from biomedical research to information technology.

2.1 **Chemical Approaches to Nanostructured Materials** 24
2.1.1 From Molecular Building Blocks to Nanostructures 24
2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins 24
2.1.3 Chemical Synthesis of Artificial Nanostructures 26
2.1.4 From Structural Control to Designed Properties and Functions 27
2.2 **Molecular Switches and Logic Gates** 28
2.2.1 From Macroscopic to Molecular Switches 28
2.2.2 Digital Processing and Molecular Logic Gates 29
2.2.3 Molecular AND, NOT and OR Gates 29
2.2.4 Combinational Logic at the Molecular Level 31
2.2.5 Intermolecular Communication 32
2.3 **Solid-State Devices** 35
2.3.1 From Functional Solutions to Electroactive and Photoactive Solids 35
2.3.2 Langmuir–Blodgett Films 36
2.3.3 Self-Assembled Monolayers 40
2.3.4 Nanogaps and Nanowires 43
2.4 **Conclusions and Outlook** 46
References 47

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_2

2.1 Chemical Approaches to Nanostructured Materials

The fabrication of conventional devices relies on the assembly of macroscopic building blocks with specific configurations. The shapes of these components are carved out of larger materials by exploiting physical methods. This top-down approach to engineered building blocks is extremely powerful and can deliver effectively and reproducibly microscaled objects. However, this strategy becomes increasingly challenging as the dimensions of the target structures approach the nanoscale. Indeed, the physical fabrication of nanosized features with subnanometer precision is a formidable technological challenge.

2.1.1 From Molecular Building Blocks to Nanostructures

Nature efficiently builds nanostructures relying on chemical approaches. Tiny molecular building blocks are assembled with a remarkable degree of structural control in a variety of nanoscaled materials with defined shapes, properties and functions. In contrast to the top-down physical methods, small components are connected to produce larger objects in these bottom-up chemical strategies. It is becoming apparent that the limitations of the top-down approach to artificial nanostructures can be overcome by mimicking nature's bottom-up processes. Indeed, beautiful and ingenious examples of molecule-based strategies to chemically fabricate nanoscaled building blocks for functional materials and innovative devices are starting to emerge.

2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins

Nanoscaled macromolecules play a fundamental role in biological processes [2.1]. Nucleic acids, for example, ensure the transmission and expression of genetic information. These particular biomolecules are linear polymers incorporating nucleotide repeating units (Fig. 2.1a). Each nucleotide has a phosphate bridge and a sugar residue. Chemical bonds between the phosphate of one nucleotide and the sugar of the next ensures the propagation of a polynucleotide strand from the 5′ to the 3′ end. Along the sequence of alternating sugar and phosphate fragments, an extended chain of robust covalent bonds involving carbon, oxygen and phospho-

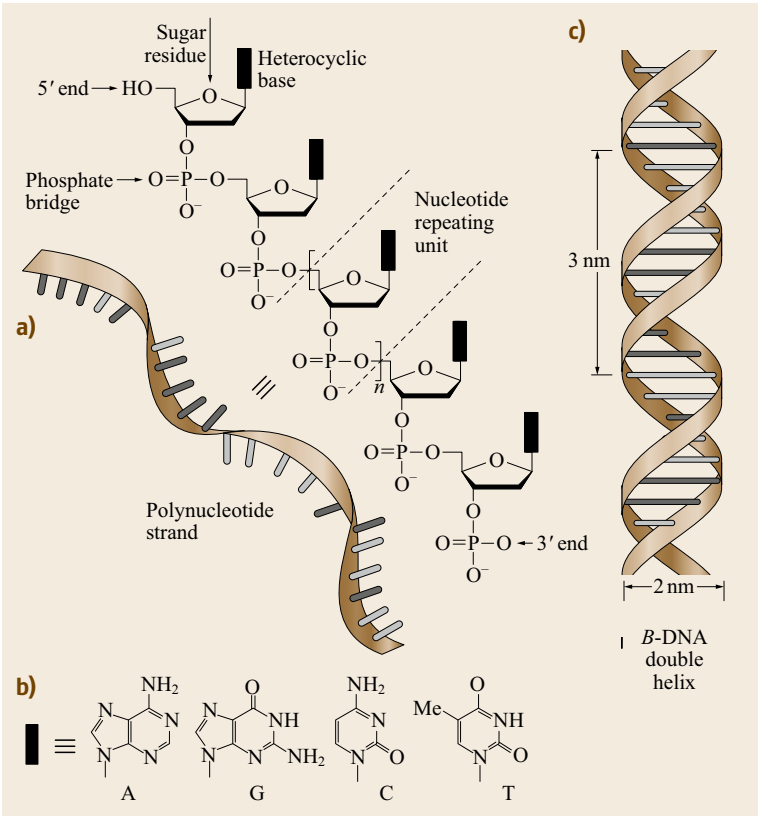


Fig. 2.1a–c A polynucleotide strand **(a)** incorporates alternating phosphate and sugar residues joined by covalent bonds. Each sugar carries one of four heterocyclic bases **(b)**. Noncovalent interactions between complementary bases in two independent polynucleotide strands encourage the formation of nanoscaled double helixes **(c)**

rous atoms forms the main backbone of the polymeric strand.

Every single nucleotide of a polynucleotide strand carries one of the four heterocyclic bases shown in Fig. 2.1b. For a strand incorporating 100 nucleotide repeating units, a total of 4^{100} unique polynucleotide sequences are possible. It follows that nature can fabricate a huge number of closely related nanostructures relying on four building blocks only. The heterocyclic bases appended to the main backbone of alternating phosphate and sugar units can sustain hydrogen bonding and $[\pi \cdots \pi]$ stacking interactions. Hydrogen bonds, formed between [N–H] donors and either N or O acceptors, encourage the pairing of adenine (A) with thymine (T) and of guanine (G) with cytosine (C). The stacking interactions involve attractive contacts between the extended π-surfaces of heterocyclic bases.

In the *B* conformation of deoxyribonucleic acid (DNA), the synergism of hydrogen bonds and $[\pi \cdots \pi]$ stacking glues pairs of complementary polynucleotide strands into fascinating double helical supermolecules (Fig. 2.1c) with precise structural control at the subnanometer level. The two polynucleotide strands wrap around a common axis to form a right-handed double helix with a diameter of ca. 2 nm. The hydrogen bonded and $[\pi \cdots \pi]$ stacked base pairs lie at the core of the helix with their π-planes perpendicular to the main axis of the helix. The alternating phosphate and sugar units define the outer surface of the double helix. In *B*-DNA, approximately ten base pairs define each helical turn corresponding to a rise per turn or helical pitch of ca. 3 nm. Considering that these molecules can incorporate up to approximately 10^{11} base pairs, extended end-to-end lengths spanning from only a few nanometers to hundreds of meters are possible.

Nature's operating principles to fabricate nanostructures are not limited to nucleic acids. Proteins are also built joining simple molecular building blocks, the amino acids, by strong covalent bonds [2.1]. More precisely, nature relies on twenty amino acids differing in their side chains to assemble linear polymers, called polypeptides, incorporating an extended backbone of robust [C–N] and [C–C] bonds (Fig. 2.2a). For a single polymer strand of 100 repeating amino acid units, a total of 20^{100} unique combinations of polypeptide sequences are possible. Considering that proteins can incorporate more than one polypeptide chain with over 4000 amino acid residues each, it is obvious that nature can assemble an enormous number of different biomolecules relying on the same fabrication strategy and a relatively small pool of building blocks.

The covalent backbones of the polypeptide strands form the main skeleton of a protein molecule. In addition, a myriad of secondary interactions, involving noncovalent contacts between portions of the amino acid residues, control the arrangement of the individual polypeptide chains. Intrastrand hydrogen bonds curl single polypeptide chains around a longitudinal axis in a helical fashion to form tubular nanostructures ca. 0.5 nm wide and ca. 2 nm long (Fig. 2.2b). Similarly, interstrand hydrogen bonds can align from 2 up to 15 parallel or antiparallel polypeptide chains to form nanoscaled sheets with average dimensions of 2×3 nm (Fig. 2.2c). Multiple nanohelices and/or nanosheets

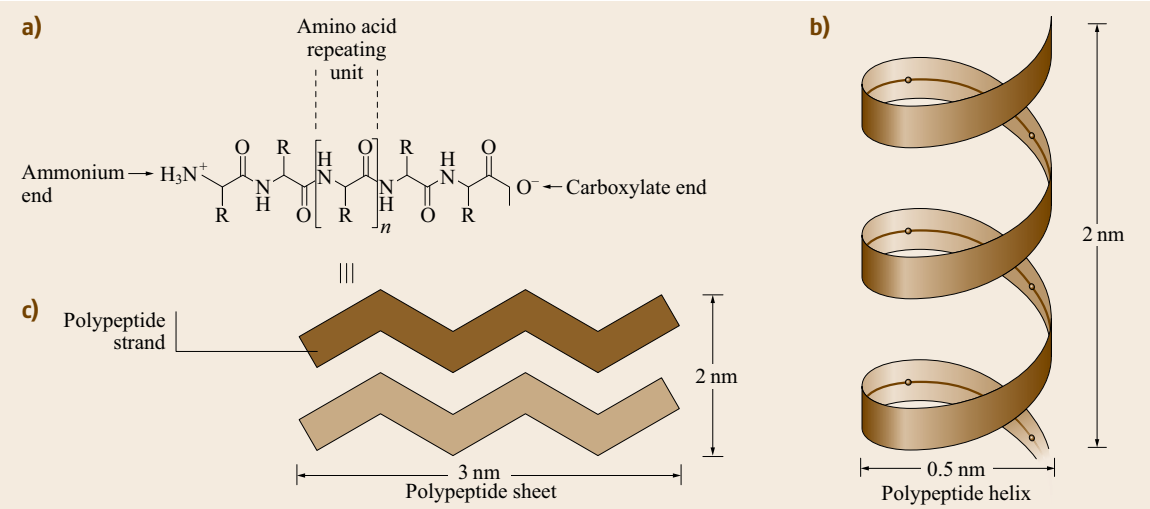


Fig. 2.2a–c A polypeptide strand **(a)** incorporates amino acid residues differing in their side chains and joined by covalent bonds. Hydrogen bonding interactions curl a single polypeptide strand into a helical arrangement **(b)** or lock pairs of strands into nanoscaled sheets **(c)**

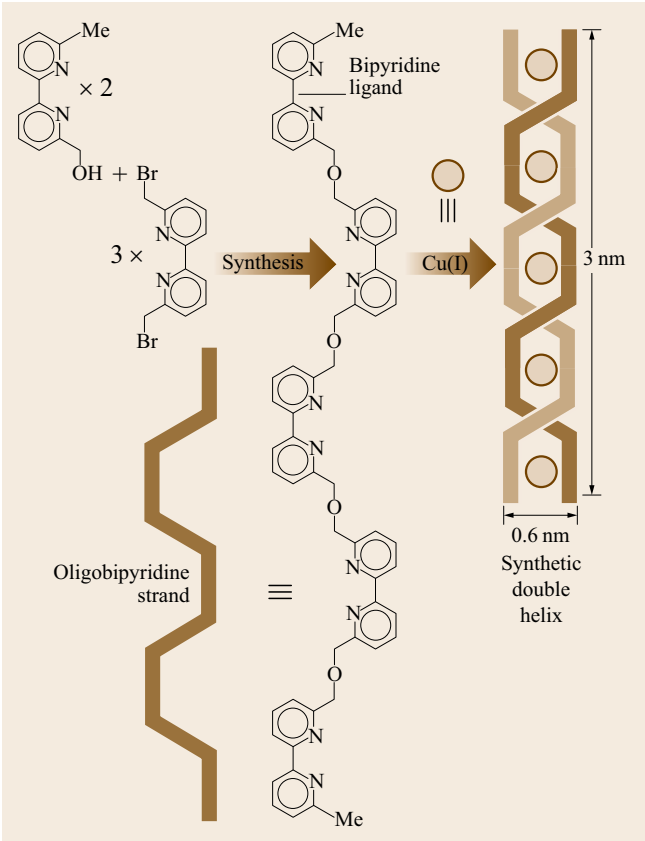


Fig. 2.3 An oligobipyridine strand can be synthesized joining five bipyridine subunits by covalent bonds. The tetrahedral coordination of pairs of bipyridine ligands by Cu(I) ions encourages the assembly of two oligobipyridine strands into a double helical arrangement

combine into a unique three-dimensional arrangement dictating the overall shape and dimensions of a protein.

2.1.3 Chemical Synthesis of Artificial Nanostructures

Nature fabricates complex nanostructures relying on simple criteria and a relatively small pool of molecular building blocks. Robust chemical bonds join the basic components into covalent scaffolds. Noncovalent interactions determine the three-dimensional arrangement and overall shape of the resulting assemblies. The multitude of unique combinations possible for long sequences of chemically connected building blocks provides access to huge libraries of nanoscaled biomolecules.

Modern chemical synthesis has evolved considerably over the past few decades [2.2]. Experimental procedures to join molecular components with structural control at the picometer level are available. A multitude of synthetic schemes to encourage the formation of chemical bonds between selected atoms in reacting molecules have been developed. Furthermore, the tremendous progress of crystallographic and spectroscopic techniques has provided efficient and reliable tools to probe directly the structural features of artificial inorganic and organic compounds. It follows that designed molecules with engineered shapes and dimensions can be now prepared in a laboratory relying on the many tricks of chemical synthesis and the power of crystallographic and spectroscopic analyses.

The high degree of sophistication reached in this research area translates into the possibility of mimicking the strategies successfully employed by nature to chemically fabricate nanostructures [2.3]. Small molecular building blocks can be synthesized and covalently joined following routine laboratory procedures. It is even possible to design the stereoelectronic properties of the assembling components in order to shape the geometry of the final product with the assistance of noncovalent interactions. For example, five bipyridine building blocks (Fig. 2.3) can be connected in five synthetic steps to produce an oligobipyridine strand [2.4]. The five repeating units are bridged by [C–O] bonds and can chelate metal cations in the bay regions defined by their two nitrogen atoms. The spontaneous assembly of two organic strands in a double helical arrangement occurs in the presence of inorganic cations. In the resulting helicate, the two oligobipyridine strands wrap around an axis defined by five Cu(I) centers. Each inorganic cation coordinates two bipyridine units with a tetrahedral geometry imposing a diameter of ca. 0.6 nm on the nanoscaled helicate [2.5]. The overall length from one end of the helicate to the other is ca. 3 nm [2.6]. The analogy between this artificial double helix and the *B*-DNA double helix shown in Fig. 2.1c is obvious. In both instances, a supramolecular glue combines two independent molecular strands into nanostructures with defined shapes and dimensions.

The chemical synthesis of nanostructures can borrow Nature's design criteria as well as its molecular building blocks. Amino acids, which are the basic components of proteins, can be assembled into artificial macrocycles. In the example of Fig. 2.4, eight amino acid residues are joined through the formation of [C–N] bonds in multiple synthetic steps [2.7]. The resulting covalent backbone defines a circular cavity with a diameter of ca. 0.8 nm [2.8]. In analogy to the polypeptide chains of proteins, the amino acid residues of this artificial oligopeptide can sustain hydrogen bonding interactions. It follows that multiple macrocycles can pile on top of each other to form tubular nanostructures. The walls of the resulting nanotubes are maintained in

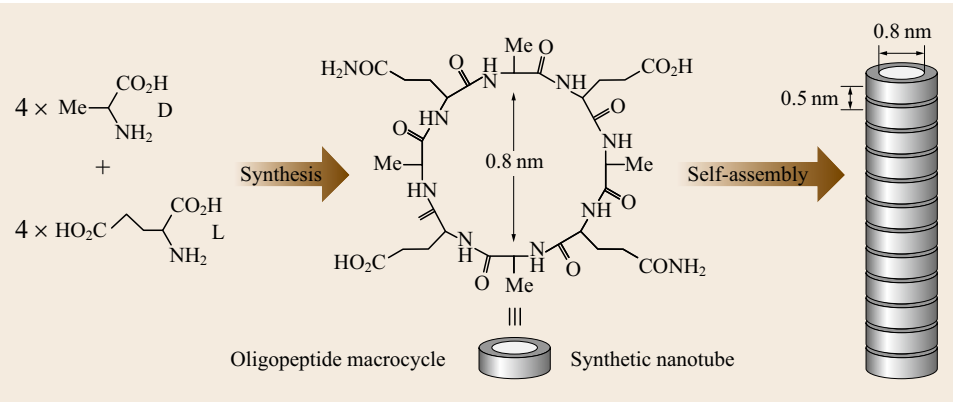


Fig. 2.4 Cyclic oligopeptides can be synthesized joining eight amino acid residues by covalent bonds. The resulting macrocycles self-assemble into nanoscaled tube-like arrays

position by the cooperative action of at least eight primary hydrogen bonding contacts per macrocycle. These noncovalent interactions maintain the mean planes of independent macrocycles in an approximately parallel arrangement with a plane-to-plane separation of ca. 0.5 nm.

2.1.4 From Structural Control to Designed Properties and Functions

The examples in Figs. 2.3 and 2.4 demonstrate that modular building blocks can be assembled into target compounds with precise structural control at the picometer level through programmed sequences of synthetic steps. Indeed, modern chemical synthesis offers access to complex molecules with nanoscaled dimensions and, thus, provides cost-effective strategies for the production and characterization of billions of engineered nanostructures in parallel. Furthermore, the high degree of structural control is accompanied by the possibility of designing specific properties into the target nanostructures. Electroactive and photoactive components can be integrated chemically into functional molecular machines [2.9]. Extensive electrochemical investigations have demonstrated that inorganic and organic compounds can exchange electrons with macroscopic electrodes [2.10]. These studies have unraveled the processes responsible for the oxidation and reduction of numerous functional groups and indicated viable design criteria to adjust the ability of molecules to accept or donate electrons [2.11]. Similarly, detailed photochemical and photophysical investigations have elucidated the mechanisms responsible for the absorption and emission of photons at the molecular level [2.12]. The vast knowledge established on the interactions between light and molecules offers the opportunity to engineer chromophoric and fluorophoric functional groups with defined absorption and emission properties [2.13].

The power of chemical synthesis to deliver functional molecules is, perhaps, better illustrated by the molecular motor shown in Fig. 2.5. The preparation of this [2]rotaxane requires twelve synthetic steps starting from known precursors [2.14]. This complex molecule incorporates a Ru(II)-trisbipyridine stopper bridged to a linear tetracationic fragment by a rigid triaryl spacer. The other end of the tetracationic portion is terminated by a bulky tetraarylmethane stopper. The bipyridinium unit of this dumbbell-shaped compound is encircled by a macrocyclic polyether. No covalent bonds join the macrocyclic and linear components. Rather, hydrogen bonding and $[\pi \cdots \pi]$ stacking interactions maintain the macrocyclic polyether around the bipyridinium unit. In addition, mechanical constraints associated with the bulk of the two terminal stoppers prevent the macrocycle to slip off the thread. The approximate end-to-end distance for this [2]rotaxane is ca. 5 nm.

The bipyridinium and the 3,3′-dimethyl bipyridinium units within the dumbbell-shaped component undergo two consecutive and reversible monoelectronic reductions [2.14]. The two methyl substituents on the 3,3′-dimethyl bipyridinium dication make this electroactive unit more difficult to reduce. In acetonitrile, its redox potential is ca. 0.29 V more negative than that of the unsubstituted bipyridinium dication. Under irradiation at 436 nm in degassed acetonitrile, the excitation of the Ru(II)-trisbipyridine stopper is followed by electron transfer to the unsubstituted bipyridinium unit. In the presence of a sacrificial electron donor (triethanolamine) in solution, the photogenerated hole in the photoactive stopper is filled and the undesired back electron transfer is suppressed. The permanent and light-induced reduction of the dicationic bipyridinium unit to a radical cation depresses significantly the magnitude of the noncovalent interactions holding the macrocyclic polyether in position. As a result, the macrocycle shuttles from the reduced unit to the adjacent dicationic 3,3′-dimethyl bipyridinium. After

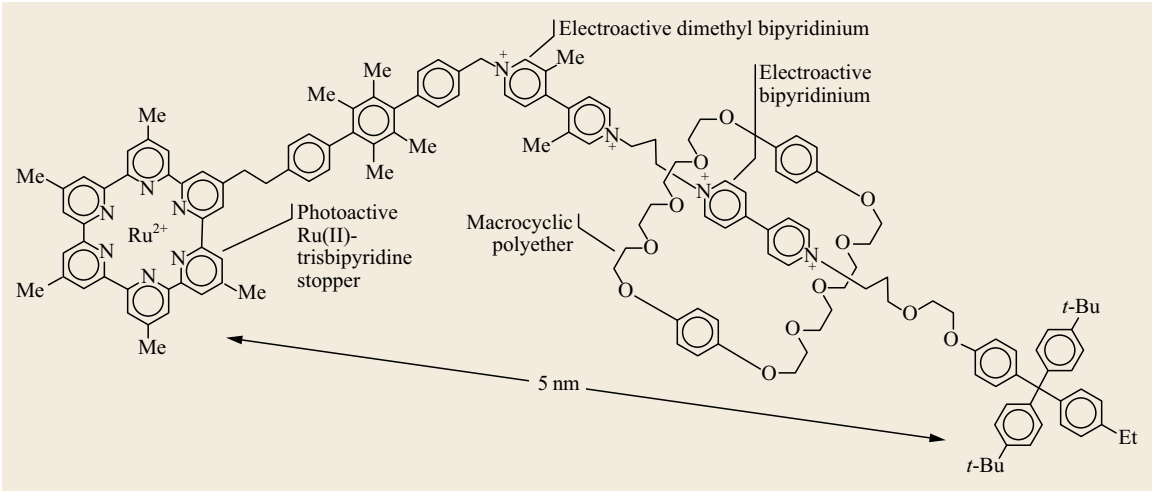


Fig. 2.5 This nanoscaled [2]rotaxane incorporates a photoactive Ru(II)-trisbipyridine stopper and two electroactive bipyridinium units. Photoinduced electron transfer from the photoactive stopper to the encircled electroactive unit forces the macrocyclic polyether to shuttle to the adjacent bipyridinium dication

the diffusion of molecular oxygen into the acetonitrile solution, oxidation occurs restoring the dicationic form of the bipyridinium unit and its ability to sustain strong noncovalent bonds. As a result, the macrocyclic polyether shuttles back to its original position. This amazing example of a molecular shuttle reveals that dynamic processes can be controlled reversibly at the molecular level relying on the clever integration of electroactive and photoactive fragments into functional and nanoscaled molecules.

2.2 Molecular Switches and Logic Gates

2.2.1 From Macroscopic to Molecular Switches

Everyday, we routinely perform dozens of switching operations. We turn our personal computers, cellular phones, CD players, radios or simple light bulbs off and on at a click of a button. Every single time, our finger exerts a mechanical stimulation on a control device, namely a switch. The external stimulus changes the physical state of the switch, closing or opening an electric circuit and enabling or preventing the passage of electrons. Overall, the switch transduces a mechanical input into an electrical output.

The use of switching devices is certainly not limited to electric circuits. For example, a switch at the junction of a railroad can divert trains from one track to another. Similarly, a faucet in a lavatory pipe can block or release the flow of water. Of course, the nature of the control stimulations and the character of the final outcome vary significantly from case to case, but the operating principle behind each switching device is the same. In all cases, input stimulations reach the switch changing its physical state and producing a specific output.

The development of nanoscaled counterparts to conventional switches is expected to have fundamental scientific and technological implications. Indeed, practical applications for ultraminiaturized switches can be easily envisaged in a diversity of areas ranging from biomedical research to information technology. The major challenge in the quest for nanoswitches, however, is the identification of reliable design criteria and operating principles for these devices. Chemical approaches to implement molecule-sized switches appear to be extremely promising. The intrinsically small dimensions of organic molecules coupled to the power of chemical synthesis are the main driving forces behind these exploratory investigations.

Certain organic molecules adjust their structural and electronic properties when stimulated with chemical, electrical or optical inputs. Generally, the change is accompanied by an electrochemical or spectroscopic response. Overall, these nanostructures transduce input stimulations into detectable outputs and, appropriately, are called molecular switches [2.15, 16]. The chemical transformations associated with these switching processes are often reversible. The chemical system returns

to the original state when the input signal is turned off. The interconverting states of a molecular switch can be isomers, an acid and its conjugated base, the oxidized and reduced forms of a redox active molecule or even the complexed and uncomplexed forms of a receptor [2.9, 15, 16]. The output of a molecular switch can be a chemical, electrical and/or optical signal that varies in intensity with the interconversion process. For example, changes in absorbance, fluorescence, pH or redox potential can accompany the reversible transformation of a molecular switch.

2.2.2 Digital Processing and Molecular Logic Gates

In present computer networks, data are electronically elaborated by microprocessor systems [2.17] and are exchanged optically between remote locations [2.18]. Data processing and communication require the encoding of information in electrical and optical signals in the form of binary digits. Using arbitrary assumptions, logic thresholds can be established for each signal and, then, 0 and 1 digits can be encoded following simple conventions. Sequences of electronic devices manipulate the encoded bits executing logic functions as a result of basic switching operations.

The three basic AND, NOT and OR operators combine binary inputs into binary outputs following precise logic protocols [2.17]. The NOT operator converts an input signal into an output signal. When the input is 0, the output is 1. When the input is 1, the output is 0. Because of the inverse relationship between the input and output values, the NOT gate is often called inverter [2.19]. The OR operator combines two input signals into a single output signal. When one or both inputs are 1, the output is 1. When both inputs are 0, the output is 0. The AND gate also combines two input signals into one output signal. In this instance, however, the output is 1 only when both inputs are 1. When at least one input is 0, the output is 0.

The output of one gate can be connected to one of the inputs of another operator. A NAND gate, for example, is assembled connecting the output of an AND operator to the input of a NOT gate. Now, the two input signals are converted into the final output after two consecutive logic operations. In a similar fashion, a NOR gate can be assembled connecting the output of an OR operator to the input of a NOT gate. Once again, two consecutive logic operations determine the relation between two input signals and a single output. The NAND and NOR operations are termed universal functions because any conceivable logic operation can be implemented relying on one of these two gates only [2.17]. In fact, digital circuits are fabricated routinely, interconnecting exclusively NAND or exclusively NOR operators [2.19].

The logic gates of conventional microprocessors are assembled interconnecting transistors and their input and output signals are electrical [2.19]. However, the concepts of binary logic can be extended to chemical, mechanical, optical, pneumatic or any other type of signal. First, it is necessary to design devices that can respond to these stimulations in the same way transistors respond to electrical signals. Molecular switches respond to a variety of input stimulations producing specific outputs and can, therefore, be exploited to implement logic functions [2.20, 21].

2.2.3 Molecular AND, NOT and OR Gates

More than a decade ago, a potential strategy to execute logic operations at the molecular level was proposed [2.22]. Later, the analogy between molecular switches and logic gates was recognized in a seminal article [2.23]. There, it is demonstrated that AND, NOT and OR operations can be reproduced with fluorescent molecules. The pyrazole derivative **1** (Fig. 2.6) is a molecular NOT gate. It imposes an inverse relation between a chemical input (concentration of H^+) and an optical output (emission intensity). In a mixture of methanol and water, the fluorescence quantum yield of **1** is 0.13 in the presence of only 0.1 equivalents of H^+ [2.23]. The quantum yield drops to 0.003 when the equivalents of H^+ are 1000. Photoinduced electron transfer from the central pyrazoline unit to the pendant benzoic acid quenches the fluorescence of the protonated form. Thus, a change in H^+ concentration (**I**) from a low to a high value switches the emission intensity (**O**) from a high to a low value. The inverse relationship between the chemical input **I** and the optical output **O** translates into the truth table of a NOT operation if a positive logic convention (low = 0, high = 1) is applied to both signals. The emission intensity is high ($\mathbf{O} = 1$) when the concentration of H^+ is low ($\mathbf{I} = 0$). The emission intensity is low ($\mathbf{O} = 0$) when the concentration of H^+ is high ($\mathbf{I} = 1$).

The anthracene derivative **2** (Fig. 2.6) is a molecular OR gate. It transduces two chemical inputs (concentrations of Na^+ and K^+) into an optical output (emission intensity). In methanol, the fluorescence quantum yield is only 0.003 in the absence of metal cations [2.23]. Photoinduced electron transfer from the nitrogen atom of the azacrown fragment to the anthracene fluorophore quenches the emission. After the addition of 1000 equivalents of either Na^+ or K^+, the quantum yield raises to 0.053 and 0.14, respectively. Similarly, the quantum yield is 0.14 when both metal cations are present in solution. The complexation of one of the

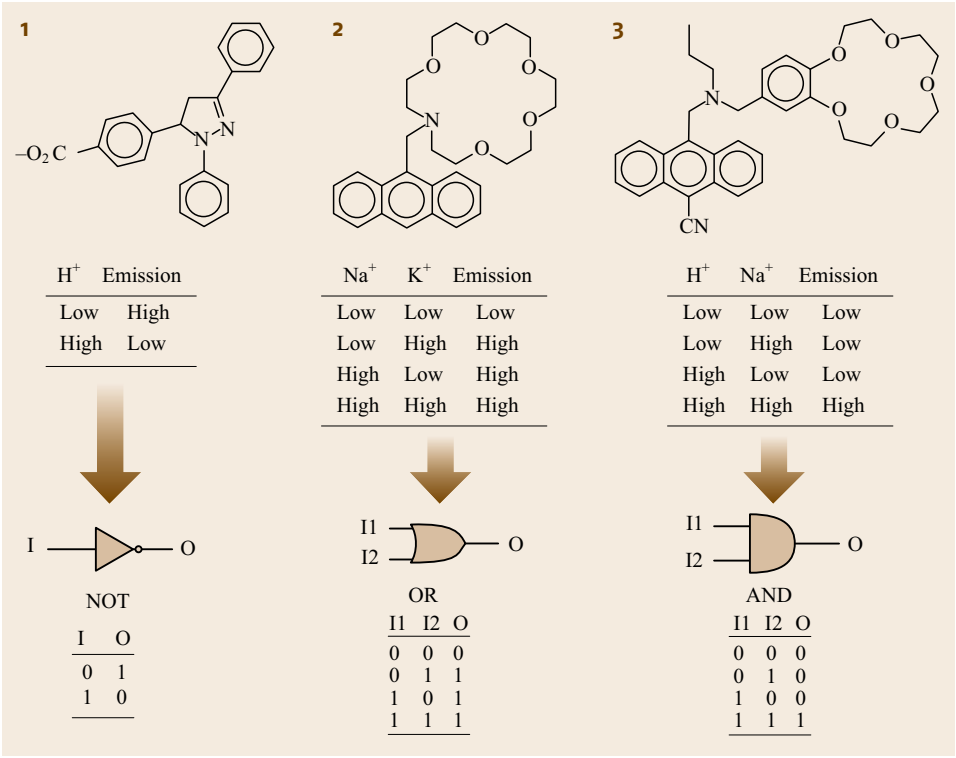


Fig. 2.6 The fluorescence intensity of the pyrazoline derivative **1** is high when the concentration of H^+ is low and vice versa. The fluorescence intensity of the anthracene derivative **2** is high when the concentration of Na^+ and/or K^+ is high. The emission is low when both concentrations are low. The fluorescence intensity of the anthracene **3** is high only when the concentrations of H^+ and Na^+ are high. The emission is low in the other three cases. The signal transductions of the molecular switches **1**, **2** and **3** translate into the truth tables of NOT, OR and AND gates, respectively, if a positive logic convention is applied to all inputs and outputs (low = 0, high = 1)

two metal cations inside the azacrown receptor depresses the efficiency of the photoinduced electron transfer enhancing the fluorescence. Thus, changes in the concentrations of Na^+ (**I1**) and/or K^+ (**I2**) from low to high values switch the emission intensity (**O**) from a low to a high value. The relationship between the chemical inputs **I1** and **I2** and the optical output **O** translates into the truth table of an OR operation if a positive logic convention (low = 0, high = 1) is applied to all signals. The emission intensity is low (**O** = 0) only when the concentration of Na^+ and K^+ are low (**I1** = 0, **I2** = 0). The emission intensity is high (**O** = 1) for the other three input combinations.

The anthracene derivative **3** (Fig. 2.6) is a molecular AND gate. It transduces two chemical inputs (concentrations of H^+ and Na^+) into an optical output (emission intensity). In a mixture of methanol and *iso*-propanol, the fluorescence quantum yield is only 0.011 in the absence of H^+ or Na^+ [2.23]. Photoinduced electron transfer from either the tertiary amino group or the catechol fragment to the anthracene fluorophore quenches the emission. After the addition of either 100 equivalents of H^+ or 1000 equivalents of Na^+, a modest change of the quantum yield to 0.020 and 0.011, respectively, is observed. Instead, the quantum yield increases to 0.068 when both species are present in solution. The protonation of the amino group and the insertion of the metal cation in the benzocrown ether receptor depress the efficiency of the photoinduced electron transfer processes enhancing the fluorescence. Thus, changes in the concentrations of H^+ (**I1**) and Na^+ (**I2**) from low to high values switch the emission intensity (**O**) from a low to a high value. The relationship between the chemical inputs **I1** and **I2** and the optical output **O** translates into the truth table of an AND operation if a positive logic convention (low = 0, high = 1) is applied to all signals. The emission intensity is high (**O** = 1) only when the concentration of H^+ and Na^+ are high (**I1** = 1, **I2** = 1). The emission intensity is low (**O** = 0) for the other three input combinations.

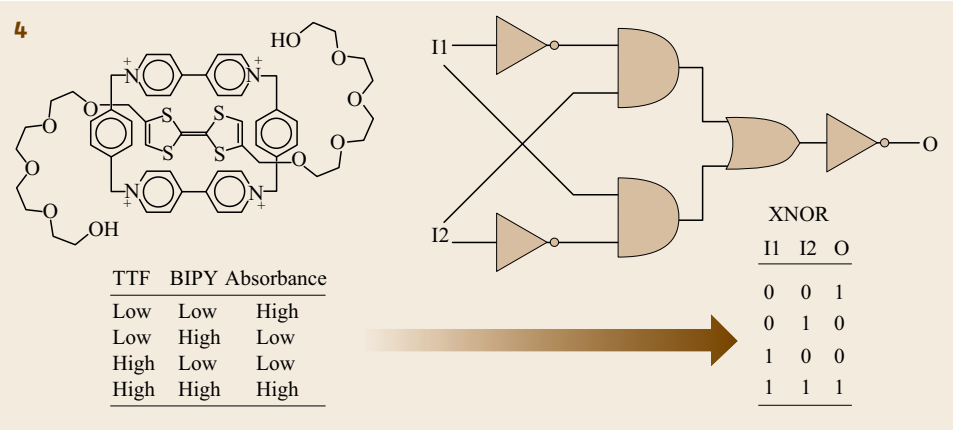
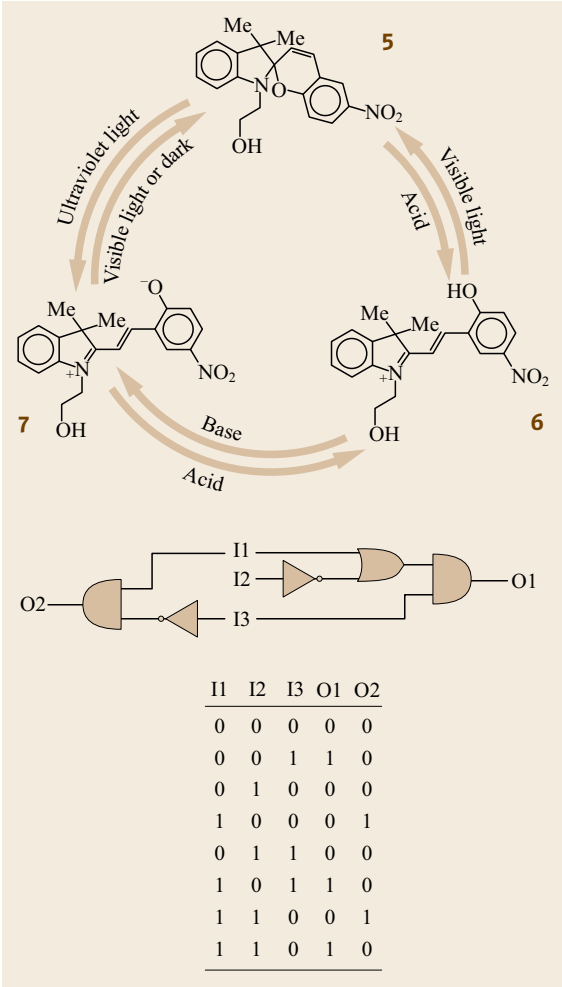


Fig. 2.7 The charge-transfer absorbance of complex **4** is high when the voltage input addressing the tetrathiafulvalene (TTF) unit is low and stimulating the bipyridinium (BIPY) units is high, and vice versa. If a positive logic convention is applied to the TTF input and to the absorbance output (low = 0, high = 1) while a negative logic convention is applied to the BIPY input (low = 0, high = 1), the signal transduction of **4** translates into the truth table of a XNOR circuit

2.2.4 Combinational Logic at the Molecular Level

The fascinating molecular AND, NOT and OR gates illustrated in Fig. 2.6 have stimulated the design of related chemical systems able to execute the three basic logic operations and simple combinations of them [2.20, 21]. Most of these molecular switches convert chemical inputs into optical outputs. However, the implementation of logic operations at the molecular level is not limited to the use of chemical inputs. For example, electrical signals and reversible redox processes can be exploited to modulate the output of a molecular switch [2.24]. The supramolecular assembly **4** (Fig. 2.7) executes a XNOR function relying on these operating principles. The π-electron-rich tetrathiafulvalene (TTF) guest threads the cavity of a π-electron deficient bipyridinium (BIPY) host. In acetonitrile, an absorption band associated with the charge-transfer interactions between the complementary π-surfaces is observed at 830 nm. Electrical stimulations alter the redox state of either the TTF or the BIPY units encouraging the separation of the two components of the complex and the disappearance of the charge-transfer band. Electrolysis at a potential of +0.5 V oxidizes the neutral TTF unit to a monocationic state. The now cationic guest is expelled from the cavity of the tetracationic host as a result of electrostatic repulsion. Consistently, the absorption band at 830 nm disappears. The charge-transfer band, however, is restored after the exhaustive back reduction of the TTF unit at a potential of 0 V. Similar changes in the absorption properties can be induced addressing the BIPY units. Electrolysis at −0.3 V reduces the dicationic BIPY units to their monocationic forms encouraging the separation of the two components of the complex and the disappearance of the absorption band. The original absorption spectrum is restored after the exhaustive back oxidation of the BIPY units at a potential of 0 V. Thus, this supramolecular system responds to electrical stimulations producing an optical output. One of the electrical inputs (**I1**) controls the redox state of the TTF unit switching between 0 and +0.5 V. The other (**I2**) determines the redox state of the bipyridinium units switching between −0.3 and 0 V. The optical output (**O**) is the absorbance of the charge-transfer band. A positive logic convention (low = 0, high = 1) can be applied to the input **I1** and output **O**. A negative logic convention (low = 1, high = 0) can be applied to the input **I2**. The resulting truth table corresponds to that of a XNOR circuit (Fig. 2.7). The charge-transfer absorbance is high (**O** = 1) only when one voltage input is low and the other is high (**I1** = 0, **I2** = 0) or vice versa (**I1** = 1, **I2** = 1). It is important to note that the input string with both **I1** and **I2** equal to 1 implies that input potentials of +0.5 and −0.3 V are applied simultaneously to a solution containing the supramolecular assembly **4** and not to an individual complex. Of course, the concomitant oxidation of the TTF guest and reduction of the BIPY units in the very same complex would be unrealistic. In bulk solution, instead, some complexes are oxidized while other are reduced, leaving the average solution composition unaffected. Thus, the XNOR operation executed by this supramolecular system is a consequence of bulk properties and not a result of unimolecular signal transduction.

Optical inputs can be employed to operate the three-state molecular switch of Fig. 2.8 in acetonitrile



I1	I2	I3	O1	O2
0	0	0	0	0
0	0	1	1	0
0	1	0	0	0
1	0	0	0	1
0	1	1	0	0
1	0	1	1	0
1	1	0	0	1
1	1	0	1	0

Fig. 2.8 Ultraviolet light (**I1**), visible light (**I2**) and H^+ (**I3**) inputs induce the interconversion between the three states **5**, **6** and **7**. The colorless state **5** does not absorb in the visible region. The yellow-green state **6** absorbs at 401 nm (**O1**). The purple state **7** absorbs at 563 nm (**O2**). The truth table illustrates the conversion of input strings of three binary digits (**I1**, **I2** and **I3**) into output strings of two binary digits (**O1** and **O2**) operated by this three-state molecular switch. A combinational logic circuit incorporating nine AND, NOT and OR operators correspond to this particular truth table

solution [2.25]. This chemical system responds to three inputs producing two outputs. The three input stimulations are ultraviolet light (**I1**), visible light (**I2**) and the concentration of H^+ (**I3**). One of the two optical outputs is the absorbance at 401 nm (**O1**), which is high when the molecular switch is in the yellow-green state **6** and low in the other two cases. The other optical output is the absorbance at 563 nm (**O2**), which is high when the molecular switch is in the purple state **7** and low in the other two cases. The colorless spiropyran state **5** switches to the merocyanine form **7** upon irradiation with ultraviolet light. It switches to the protonated merocyanine from **6** when treated with H^+. The colored state **7** isomerizes back to **5** in the dark or upon irradiation with visible light. Alternatively, **7** switches to **6** when treated with H^+. The colored state **6** switches to **5**, when irradiated with visible light, and to **7**, after the removal of H^+. In summary, this three-state molecular switch responds to two optical inputs (**I1** and **I2**) and one chemical input (**I3**) producing two optical outputs (**O1** and **O2**). Binary digits can be encoded on each signal applying positive logic conventions (low = 0, high = 1). It follows that the three-state molecular switch converts input strings of three binary digits into output strings of two binary digits. The corresponding truth table (Fig. 2.8) reveals that the optical output **O1** is high (**O1** = 1) when only the input **I3** is applied (**I1** = 0, **I2** = 0, **I3** = 1), when only the input **I2** is not applied (**I1** = 1, **I2** = 0, **I3** = 1) or when all three inputs are applied (**I1** = 1, **I2** = 1, **I3** = 1). The optical output **O2** is high (**O2** = 1) when only the input **I1** is applied (**I1** = 1, **I2** = 0, **I3** = 0) or when only the input **I3** is not applied (**I1** = 1, **I2** = 1, **I3** = 0). The combinational logic circuit (Fig. 2.8) equivalent to this truth table shows that all three inputs determine the output **O1**, while only **I1** and **I3** control the value of **O2**.

2.2.5 Intermolecular Communication

The combinational logic circuits in Figs. 2.7 and 2.8 are arrays of interconnected AND, NOT and OR operators. The digital communication between these basic logic elements ensures the execution of a sequence of simple logic operations that results in the complex logic function processed by the entire circuit. It follows that the logic function of a given circuit can be adjusted altering the number and type of basic gates and their interconnection protocol [2.17]. This modular approach to combinational logic circuits is extremely powerful. Any logic function can be implemented connecting the appropriate combination of simple AND, NOT and OR gates.

The strategies followed so far to implement complex logic functions with molecular switches is based on the careful design of the chemical system and on the judicious choice of the inputs and outputs [2.20, 21]. A specific sequence of AND, NOT and OR operations is programmed in a single molecular switch. No digital communication between distinct gates is needed since they are built in the same molecular entity. Though extremely elegant, this strategy does not have the same versatility of a modular approach. A different molecule

has to be designed, synthesized and analyzed every single time a different logic function has to be realized. In addition, the degree of complexity that can be achieved with only one molecular switch is fairly limited. The connection of the input and output terminals of independent molecular AND, NOT and OR operators, instead, would offer the possibility of assembling any combinational logic circuit from three basic building blocks.

In digital electronics, the communication between two logic gates can be realized connecting their terminals with a wire [2.19]. Methods to transmit binary data between distinct molecular switches are not so obvious and must be identified. Recently, we developed two strategies to communicate signals between compatible molecular components. In one instance, a chemical signal is communicated between two distinct molecular switches [2.26]. They are the three-state switch illustrated in Fig. 2.8 and the two-state switch of Fig. 2.9. The merocyanine form **7** is a photogenerated base. Its *p*-nitrophenolate fragment, produced upon irradiation of the colorless state **5** with ultraviolet light, can abstract a proton from an acid present in the same solution. The resulting protonated form **6** is a photoacid. It releases a proton upon irradiation with visible light and can protonate a base codissolved in the same medium. The orange azopyridine **8** switches to the red-purple azopyridinium **9** upon protonation. This process is reversible and the addition of a base restores the orange state **8**. It follows that photoinduced proton transfer can be exploited to communicate a chemical signal from **6** to **8** and from **9** to **7**. The two colored states **8** and **9** have different absorption properties in the visible region. In acetonitrile, the orange state **8** absorbs at 422 nm and the red-purple state **9** absorbs at 556 nm. The changes in absorbance of these two bands can be exploited to monitor the photoinduced exchange of protons between the two communicating molecular switches.

The three-state molecular switch and the two-state molecular switch can be operated sequentially when dissolved in the same acetonitrile solution. In the presence of one equivalent of H^+, the two-state molecular switch is in state **9** and the absorbance at 556 nm is high (**O** = 1). Upon irradiation with ultraviolet light (**I1** = 1), **5** switches to **7**. The photogenerated base deprotonates **9** producing **8** and **6**. As a result, the absorbance at 556 nm decreases (**O** = 0). Upon irradiation with visible light (**I2** = 1), **6** switches to **5** releasing H^+. The result is the protonation of **8** to form **9** and restore the high absorbance at 556 nm (**O** = 1). In summary, the three-state molecular switch transduces two optical inputs (**I1** = ultraviolet light, **I2** = visible light) into a chemical signal (proton transfer) that is communicated to the two-state molecular switch and converted into a final optical output (**O** = absorbance at 556 nm).

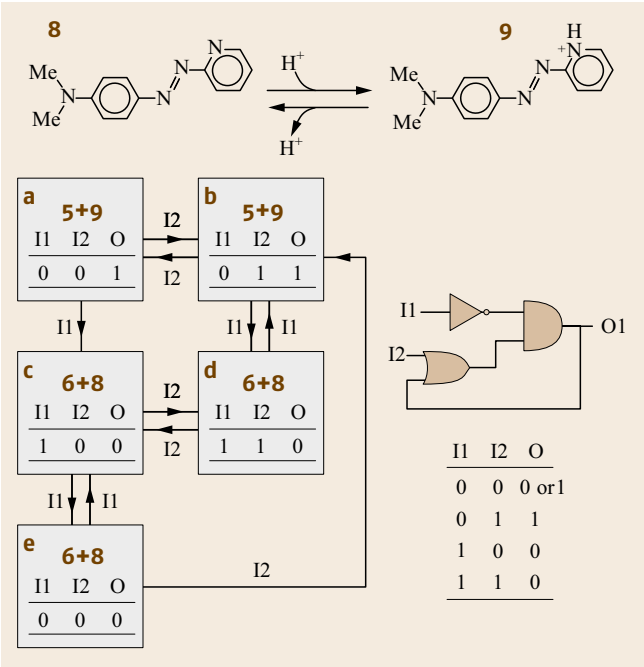
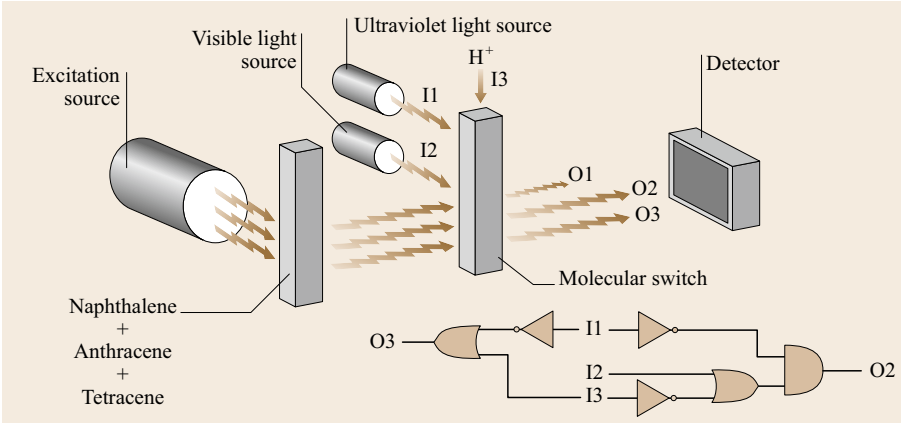


Fig. 2.9 The concentration of H^+ controls the reversible interconversion between the two states **8** and **9**. In response to ultraviolet (**I1**) and visible (**I2**) inputs, the three-state molecular switch in Fig. 2.7 modulates the ratio between these two forms and the absorbance (**O**) of **9** through photoinduced proton transfer. The truth table and sequential logic circuit illustrate the signal transduction behavior of the two communicating molecular switches. The interconversion between the five three-digit strings of input (**I1** and **I2**) and output (**O**) data is achieved varying the input values in steps

The logic behavior of the two communicating molecular switches is significantly different from those of the chemical systems illustrated in Figs. 2.6–2.8 [2.26]. The truth table in Fig. 2.9 lists the four possible combinations of two-digit input strings and the corresponding one-digit output. The output digit **O** for the input strings 01, 10 and 11 can take only one value. In fact, the input string 01 is transduced into a 1 and the input strings 10 and 11 are converted into 0. Instead, the output digit **O** for the input string 00 can be either 0 or 1. The sequence of events leading to the input string 00 determines the value of the output. The boxes a–e in Fig. 2.9 illustrates this effect. They correspond to the five three-digit input/output strings. The transformation of one box into any of the other four is achieved in one or two steps changing the values of **I1** and/or **I2**. In two instances (a and b), the two-state molecular switch is in state **9** and the output signal is high (**O** = 1). In the other three cases (c, d and e), the two-state molecular switch is in state **8** and the output signal is low (**O** = 0). The strings 000 (e) and 001 (a) correspond to the first

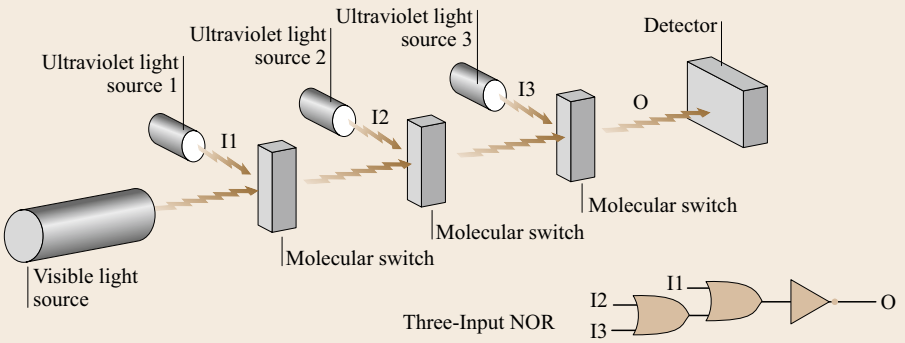
entry of the truth table. They share the same input digits, but differ in the output value. The string 000 (e) can be obtained only from the string 100 (c) varying the value of **I1**. Similarly, the string 001 (a) can be accessed only from the string 011 (b) varying the value of **I2**. In both transformations, the output digit remains unchanged. Thus, the value of **O1** in the parent string is memorized and maintained in the daughter string when both inputs become 0. This memory effect is the fundamental operating principle of sequential logic circuits [2.17], which are used extensively to assemble the memory elements of modern microprocessors. The sequential logic circuit equivalent to the truth table of the two communicating molecular switches is also shown in Fig. 2.9. In this circuit, the input data **I1** and **I2** are combined through NOT, OR and AND operators. The output of the AND gate **O** is also an input of the OR gate and controls, together with **I1** and **I2**, the signal transduction behavior.

The other strategy for digital transmission between molecules is based on the communication of optical signals between the three-state molecular switch (Fig. 2.8) and fluorescent compounds [2.25, 27, 28]. In the optical network of Fig. 2.10, three optical signals travel from an excitation source to a detector after passing through two quartz cells. The first cell contains an equimolar acetonitrile solution of naphthalene, anthracene and tetracene. The second cell contains an acetonitrile solution of the three-state molecular switch. The excitation source sends three consecutive monochromatic light beams to the first cell stimulating the emission of the three fluorophores. The light emitted in the direction perpendicular to the exciting beam reaches the second cell. When the molecular switch is in state **5**, the naphthalene emission at 335 nm is absorbed and a low intensity output (**O1**) reaches the detector. Instead, the anthracene and tetracene emissions at 401 and 544 nm, respectively, pass unaffected and high intensity outputs (**O2** and **O3**) reach the detector. When the molecular switch is in state **6**, the naphthalene and anthracene emissions are absorbed and only the tetracene emission reaches the detector (**O1** = 0, **O2** = 0, **O3** = 1). When the molecular switch is state **7**, the emission of all three fluorophores is absorbed (**O1** = 0, **O2** = 0, **O3** = 0). The interconversion of the molecular switch between the three states is induced addressing the second cell with ultraviolet (**I1**), visible (**I2**) and H^+ (**I3**) inputs. Thus, three independent optical outputs (**O1**, **O2** and **O3**) can be modulated stimulating the molecular switch with two optical and one chemical input. The truth table in Fig. 2.10 illustrates the relation between the three inputs (**I1**, **I2** and **I3**) and the three outputs (**O1**, **O2** and **O3**), when positive logic conventions are applied to all signals. The equivalent logic circuit shows that all three inputs control the anthracene channel **O2**, but only **I1** and **I3** influence the tetracene channel **O3**. Instead, the intensity of the naphthalene channel **O1** is always low and it is not affected by the three inputs.



I1	I2	I3	O1	O2	O3
0	0	0	0	1	1
0	0	1	0	0	1
0	1	0	0	1	1
1	0	0	0	0	0
0	1	1	0	1	1
1	0	1	0	0	1
1	1	0	0	0	0
1	1	1	0	0	1

Fig. 2.10 The excitation source sends three monochromatic light beams (275, 357 and 441 nm) to a quartz cell containing an equimolar acetonitrile solution of naphthalene, anthracene and tetracene. The three fluorophores absorb the exciting beams and reemit at 305, 401 and 544 nm, respectively. The light emitted in the direction perpendicular to the exciting beams passes through another quartz cell containing an acetonitrile solution of the three-state molecular switch shown in Fig. 2.7. Ultraviolet (**I1**), visible (**I2**) and H^+ (**I3**) inputs control the interconversion between the three states of the molecular switch. They determine the intensity of the optical outputs reaching the detector and correspond to the naphthalene (**O1**), anthracene (**O2**) and tetracene (**O3**) emissions. The truth table and equivalent combinational logic illustrate the relation between the three inputs and the three outputs. The output **O1** is always 0 and it is not influenced by the three inputs. Only two inputs determine the value of **O3**, while all of them control the output **O2**



I1	I2	I3	O
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	0
0	1	1	0
1	0	1	0
1	1	0	0
1	1	1	0

Fig. 2.11 The visible source sends a monochromatic beam (563 nm) to the detector. The traveling light is forced to pass through three quartz cells containing the molecular switch illustrated in Fig. 2.7. The three switching elements are operated by independent ultraviolet inputs. When at least one of them is on, the associated molecular switch is in the purple form **7**, which can absorb and block the traveling light. The truth table and equivalent logic circuit illustrate the relation between the three inputs **I1**, **I2** and **I3** and the optical output **O**

The operating principles of the optical network in Fig. 2.10 can be simplified to implement all-optical logic gates. The chemical input inducing the formation of the protonated form **6** of the molecular switch can be eliminated. The interconversion between the remaining two states **5** and **7** can be controlled relying exclusively on ultraviolet inputs. Indeed, ultraviolet irradiation induces the isomerization of the colorless form **5** to the colored species **7**, which reisomerizes to the original state in the dark. Thus, a single ultraviolet source is sufficient to control the switching from **5** to **7** and vice versa. On the basis of these considerations, all-optical NAND, NOR and NOT gates can be implemented operating sequentially or in parallel from one to three independent switching elements [2.29]. For example, the all-optical network illustrated in Fig. 2.11 is a three-input NOR gate. A monochromatic optical signal travels from a visible source to a detector. Three switching elements are aligned along the path of the traveling light. They are quartz cells containing an acetonitrile solution of the molecular switch shown in Fig. 2.8. The interconversion of the colorless form **5** into the purple isomer **7** is induced stimulating the cell with an ultraviolet input. The reisomerization from **7** to **5** occurs spontaneously, as the ultraviolet source is turned off. Using three distinct ultraviolet sources, the three switching elements can be controlled independently.

The colorless form **5** does not absorb in the visible region, while the purple isomer **7** has a strong absorption band at 563 nm. Thus, a 563 nm optical signal leaving the visible source can reach the detector unaffected only if all three switching elements are in the nonabsorbing state **5**. If one of the three ultraviolet inputs **I1**, **I2** or **I3** is turned on, the intensity of the optical output **O** drops to 3–4% of its original value. If two or three ultraviolet inputs are turned on simultaneously, the optical output drops to 0%. Indeed, the photogenerated state **7** absorbs and blocks the traveling light. Applying positive logic conventions to all signals, binary digits can be encoded in the three optical inputs and in the optical output. The resulting truth table is illustrated in Fig. 2.11. The output **O** is 1 only if all three inputs **I1**, **I2** or **I3** are 0. The output **O** is 0 if at least one of the three inputs **I1**, **I2** or **I3** is 1. This signal transduction corresponds to that executed by a three-input NOR gate, which is a combination of one NOT and two OR operators.

2.3 Solid-State Devices

2.3.1 From Functional Solutions to Electroactive and Photoactive Solids

The fascinating chemical systems illustrated in Figs. 2.6–2.11 demonstrate that logic functions can be implemented relying on the interplay between designed molecules and chemical, electrical and/or optical signals [2.20, 21]. These molecular switches, however, are operated exclusively in solution and remain far from potential applications in information technology at this stage. The integration of liquid components and volatile organic solvents in practical digital devices is hard to

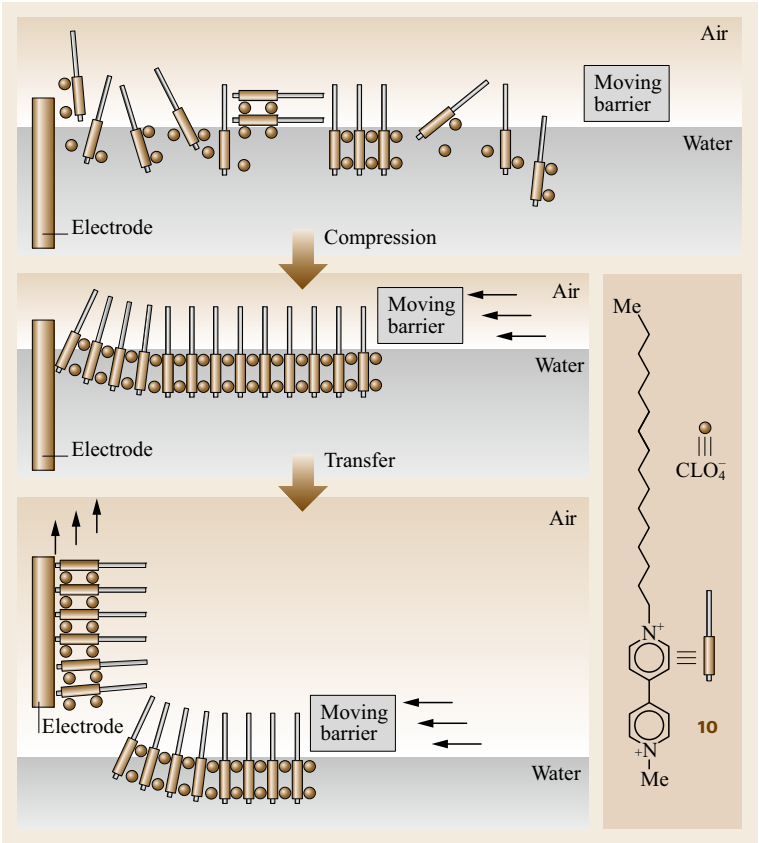


Fig. 2.12 The compression of the amphiphilic dication **10** with a moving barrier results in the formation of a packed monolayer at the air–water interface. The lifting of an electrode preimmersed in the aqueous subphase encourages the transfer of part of the monolayer on the solid support

envisage. Furthermore, the logic operations executed by these chemical systems rely on bulk addressing. Although the individual molecular components have nanoscaled dimensions, macroscopic collections of them are employed for digital processing. In some instances, the operating principles cannot even be scaled down to the unimolecular level. Often, bulk properties are responsible for signal transduction. For example, a single fluorescent compound **2** cannot execute an OR operation. Its azacrown appendage can accommodate only one metal cation. As a result, an individual molecular switch can respond to only one of the two chemical inputs. It is a collection of numerous molecular switches dissolved in an organic solvent that responds to both inputs enabling an OR operation.

The development of miniaturized molecule-based devices requires the identification of methods to transfer the switching mechanisms developed in solution to the solid state [2.30]. Borrowing designs and fabrication strategies from conventional electronics, the integration of molecular components into functional circuits and devices is starting to be explored [2.31–34]. Generally, these strategies combine lithography and surface chemistry to assemble nanometer-thick organic films on the surfaces of microscaled or nanoscaled electrodes. Two main approaches for the deposition of organized molecular arrays on inorganic supports have emerged so far. In one instance, amphiphilic molecular building blocks are compressed into organized monolayers at air–water interfaces. The resulting films can be transferred on supporting solids employing the Langmuir–Blodgett technique [2.35]. Alternatively, certain molecules can be designed to adsorb spontaneously on the surfaces of compatible solids from liquid or vapor phases. The result is the self-assembly of organic layers on inorganic supports [2.36].

2.3.2 Langmuir–Blodgett Films

Films of amphiphilic molecules can be deposited on a variety of solid supports employing the Langmuir–Blodgett technique [2.35]. This method can be extended to electroactive compounds incorporating hydrophilic and hydrophobic groups. For example, the amphiphile **10** (Fig. 2.12) has a hydrophobic hexadecyl tail attached to a hydrophilic bipyridinium dication [2.37, 38]. This compound dissolves in mixtures of chloroform and

methanol, but it is not soluble in moderately concentrated aqueous solutions of sodium perchlorate. Thus, the spreading of an organic solution of **10** on an aqueous sodium perchlorate subphase affords a collection of disorganized amphiphiles floating on the water surface (Fig. 2.12), after the evaporation of the organic solvent. The molecular building blocks can be compressed into a monolayer with the aid of a moving barrier. The hydrophobic tails align away from the aqueous phase. The hydrophilic dicationic heads and the accompanying perchlorate counterions pack to form an organized monolayer at the air–water interface. The compression process can be monitored by recording the surface pressure (π)–area per molecule (A) isotherm, which indicates a limiting molecular area of ca. 50 Å^2. This value is larger than the projected area of an oligomethylene chain. However, it correlates reasonably with the overall area of a bipyridinium dication plus two perchlorate anions.

The monolayer prepared at the air–water interface (Fig. 2.12) can be transferred on the surface of an indium-tin oxide electrode preimmersed in the aqueous phase. The slow lifting of the solid support drags the monolayer away from the aqueous subphase. The final result is the coating of the electrode with an organic film containing electroactive bipyridinium building blocks. The modified electrode can be integrated in a conventional electrochemical cell to probe the redox response of the electroactive layer. The resulting cyclic voltammograms reveal the characteristic waves for the first reduction process of the bipyridinium dications, confirming the successful transfer of the electroactive amphiphiles from the air–water interface to the electrode surface. The integration of the redox waves indicates a surface coverage of ca. 4×10^{-10} mol cm^{-2}. This value corresponds to a molecular area of ca. 40 Å^2 and is in excellent agreement with the limiting molecular area of the π–A isotherm.

These seminal experiments demonstrate that electroactive amphiphiles can be organized into uniform monolayers at the air–water interface and then transferred efficiently on the surface of appropriate substrates to produce electrode–monolayer junctions. The resulting electroactive materials can become the functional components of molecule-based devices. For example, bipyridinium-based photodiodes can be fabricated following this approach [2.39, 40]. Their operating principles rely on photoinduced electron transfer from chromophoric units to bipyridinium acceptors. The electroactive and photoactive amphiphile **11** (Fig. 2.13) incorporates hydrophobic ferrocene and pyrene tails and a hydrophilic bipyridinium head. Chloroform solutions of **11** containing ten equivalents of arachidic acid can be spread on an aqueous calcium chloride subphase in a Langmuir trough. The amphiphiles can be compressed into a mixed monolayer, after the evaporation of the organic solvent. Pronounced steps in the corresponding π–A isotherm suggest that the bulky ferrocene and pyrene groups are squeezed away from the water surface. In the final arrangement, both photoactive groups align above the hydrophobic dication.

A mixed monolayer of **11** and arachidic acid can be transferred from the air–water interface to the surface of a transparent gold electrode following the methodology illustrated for the system in Fig. 2.12. The coated electrode can be integrated in a conventional electrochemical cell. Upon irradiation at 330 nm under an inert atmosphere, an anodic photocurrent of ca. 2 nA develops at a potential of 0 V relative to a saturated calomel electrode. Indeed, the illumination of the electroactive monolayer induces the electron transfer from the pyrene appendage to the bipyridinium acceptor and then from the reduced acceptor to the electrode. A second intramolecular electron transfer from the ferrocene donor to the oxidized pyrene fills its photogenerated hole. Overall, a unidirectional flow of electrons across the monolayer–electrode junction is established under the influence of light.

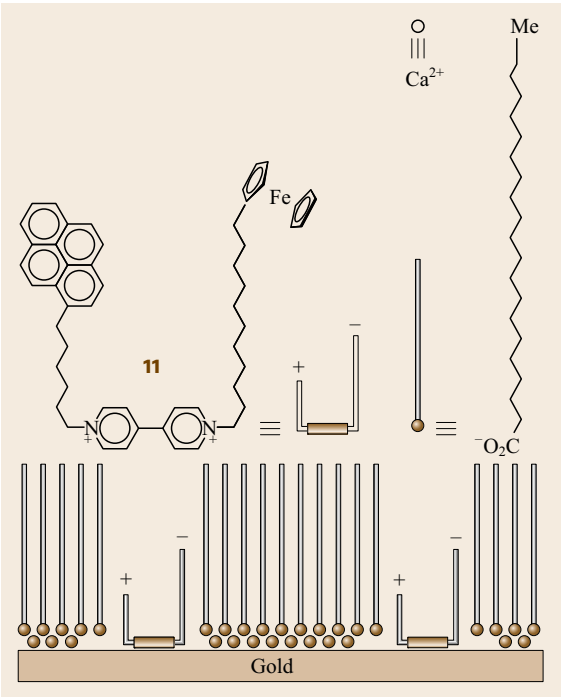


Fig. 2.13 Mixed monolayers of the amphiphile **11** and arachidic acid can be transferred from the air–water interface to the surface of an electrode to generate a molecule-based photodiode

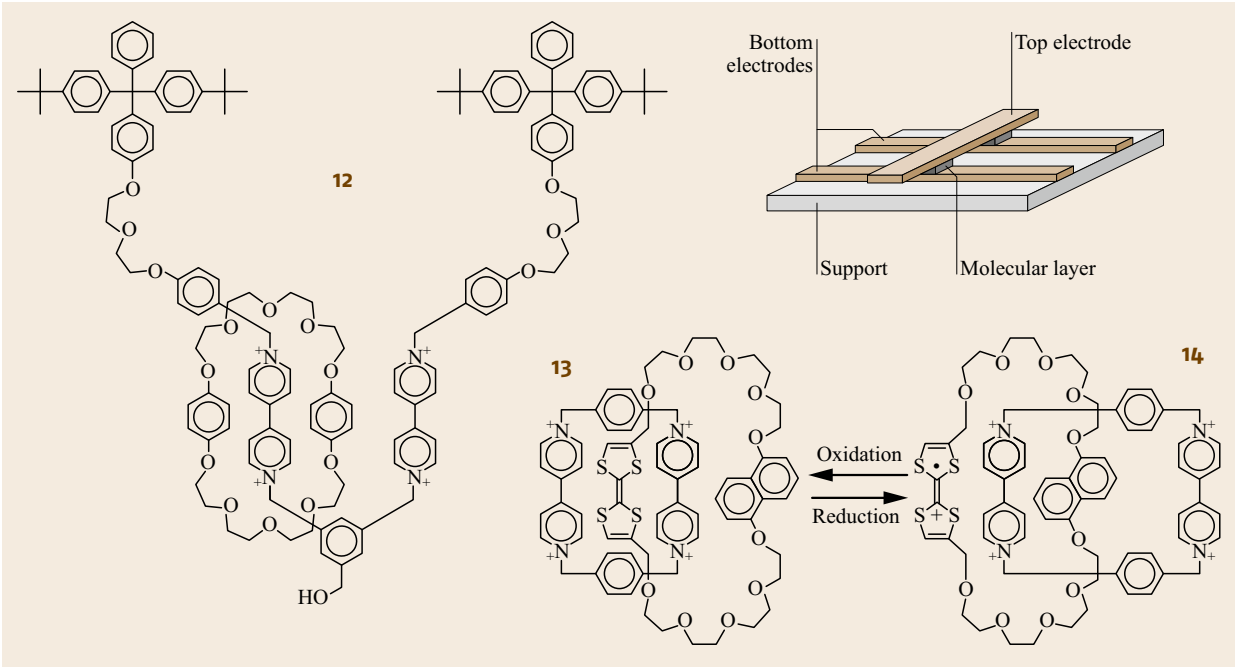


Fig. 2.14 The [2]rotaxane **12** and the [2]catenane **13** can be compressed into organized monolayers at air–water interfaces. The resulting monolayers can be transferred on the bottom electrodes of a patterned silicon support. After the deposition of a top electrode, electrode–monolayer–electrode junctions can be assembled. Note that only the portion of the monolayer sandwiched between the top and bottom electrodes is shown in the diagram. The oxidation of the tetrathiafulvalene unit of the [2]catenane **13** is followed by the circumrotation of the macrocyclic polyether to afford the [2]catenane **14**. The process is reversible and the reduction of the cationic tetrathiafulvalene unit restores the original state

The ability to transfer electroactive monolayers from air–water interfaces to electrode surfaces can be exploited to fabricate molecule-based electronic devices. In particular, arrays of interconnected electrode–monolayer–electrode tunneling junctions can be assembled combining the Langmuir–Blodgett technique with electron beam evaporation [2.33]. A schematic representation of the resulting devices is illustrated in Fig. 2.14. Initially, parallel fingers are patterned on a silicon wafer with a silicon dioxide overlayer by electron beam evaporation. The bottom electrodes deposited on the support can be either aluminum wires covered by an aluminum oxide or n-doped silicon lines with silicon dioxide overlayers. Their widths are ca. 6 or 7 μm, respectively. The patterned silicon chip is immersed in the aqueous subphase of a Langmuir trough prior to monolayer formation. After the compression of electroactive amphiphiles at the air–water interface, the substrate is pulled out of the aqueous phase to encourage the transfer of the molecular layer on the parallel bottom electrodes as well as on the gaps between them. Then, a second set of electrodes orthogonal to the first is deposited through a mask by electron beam evaporation. They consist of a titanium underlayer plus an aluminum overlayer. Their thicknesses are ca. 0.05 and 1 μm, respectively, and their width is ca. 10 μm. In the final assembly, portions of the molecular layer become sandwiched between the bottom and top electrodes. The active areas of these electrode–monolayer–electrode junctions are ca. 60–70 μm^2 and correspond to ca. 10^6 molecules.

The [2]rotaxane **12** (Fig. 2.14) incorporates a macrocyclic polyether threaded onto a bipyridinium-based backbone [2.41, 42]. The two bipyridinium dications are bridged by a *m*-phenylene spacer and terminated by tetraarylmethane appendages. These two bulky groups mechanically trap the macrocycle preventing its dissociation from the tetracationic backbone. In addition, their hydrophobicity complements the hydrophilicity of the two bipyridinium dications imposing amphiphilic character on the overall molecular assembly. This compound does not dissolve in aqueous solutions and can be compressed into organized monolayers at air-water interfaces. The corresponding π–A isotherm reveals a limiting molecular area of ca. 130 $Å^2$. This large value is a consequence of the bulk associated with the hydrophobic tetraarylmethane tails and the macrocycle encircling the tetracationic backbone.

Monolayers of the [2]rotaxane **12** can be transferred from the air–water interface to the surfaces of the bottom aluminum/aluminum oxide electrodes of a patterned silicon chip with the hydrophobic tetraarylmethane groups pointing away from the supporting substrate. The subsequent assembly of a top titanium/aluminum electrode affords electrode–monolayer–electrode junctions. Their current–voltage signature can be recorded by grounding the top electrode and scanning the potential of the bottom electrode. A pronounced increase in current is observed when the potential is lowered below −0.7 V. Under these conditions, the bipyridinium-centered lowest unoccupied molecular orbitals (LUMOs) mediate the tunneling of electrons from the bottom to the top electrode leading to a current enhancement. A similar current profile is observed if the potential is returned to 0 and then back to −2 V. Instead, a modest increase in current in the opposite direction is observed when the potential is raised above +0.7 V. Presumably, this trend is a result of the participation of the phenoxy-centered highest occupied molecular orbitals (HOMOs) in the tunneling process. After a single positive voltage pulse, however, no current can be detected if the potential is returned to negative values. In summary, the positive potential scan irreversibly suppresses the conducting ability of the electrode–molecule–electrode junction. The behavior of this device correlates with the redox response of the [2]rotaxane **12** in solution. Cyclic voltammograms reveal reversible monoelectronic reductions of the bipyridinium dications. However, they also show two irreversible oxidations associated, presumably, to the phenoxy rings of the macrocycle and tetraarylmethane groups. These observations suggest that a positive voltage pulse applied to the electrode–monolayer–electrode junction oxidizes irreversibly the sandwiched molecules suppressing their ability to mediate the transfer of electrons from the bottom to the top electrode under a negative bias.

The device incorporating the [2]rotaxane **13** can be exploited to implement simple logic operations [2.41]. The two bottom electrodes can be stimulated with voltage inputs (**I1** and **I2**) while measuring a current output (**O**) at the common top electrode. When at least one of the two inputs is high (0 V), the output is low (< 0.7 nA). When both inputs are low (−2 V), the output is high (ca. 4 nA). If a negative logic convention is applied to the voltage inputs (low = 1, high = 0) and a positive logic convention is applied to the current output (low = 0, high = 1), the signal transduction behavior translates into the truth table of an AND gate. The output **O** is 1 only when both inputs are 1. Instead, an OR operation can be executed if the logarithm of the current is considered as the output. The logarithm of the current is −12 when both voltage inputs are 0 V. It raises to ca. −9 when one or both voltage inputs are lowered to −2 V. This signal transduction behavior translates into the truth table of an OR gate if a negative logic convention is applied to the voltage inputs (low = 1, high = 0) and a positive logic convention is applied to the current output (low = 0, high = 1). The output **O** is 1 when at least one of the two inputs is 1.

The [2]catenane **13** (Fig. 2.14) incorporates a macrocyclic polyether interlocked with a tetracationic cyclophane [2.43, 44]. Organic solutions of the hexafluorophosphate salt of this [2]catenane and six equivalents of the sodium salt of dimyristoylphosphatidic acid can be cospread on the water surface of a Langmuir trough [2.45]. The sodium hexafluorophosphate formed dissolves in the supporting aqueous phase, while the hydrophilic bipyridinium cations and the amphiphilic anions remain at the interface. Upon compression, the anions align their hydrophobic tails away from the water surface forming a compact monolayer above the cationic bipyridinium derivatives. The corresponding π–A isotherm indicates limiting molecular areas of ca. 125 Å^2. This large value is a consequence of the bulk associated with the two interlocking macrocycles.

Monolayers of the [2]catenane **13** can be transferred from the air–water interface to the surfaces of the bottom n-doped silicon/silicon dioxide electrodes of a patterned silicon chip with the hydrophobic tails of the amphiphilic anions pointing away from the supporting substrate [2.46, 47]. The subsequent assembly of a top titanium/aluminum electrode affords electrode–monolayer–electrode arrays. Their junction resistance can be probed grounding the top electrode and maintaining the potential of the bottom electrode at +0.1 V. If a voltage pulse of +2 V is applied to the bottom electrode before the measurement, the junction resistance probed is ca. 0.7 GΩ. After a pulse of −2 V applied to the bottom electrode, the junction resistance probed at +0.1 V drops ca. 0.3 GΩ. Thus, alternating positive and negative voltage pulses can switch reversibly the junction resistance between high and low values. This intriguing behavior is a result of the redox and dynamic properties of the [2]catenane **13**.

Extensive spectroscopic and crystallographic studies [2.43, 44] demonstrated that the tetrathiafulvalene unit resides preferentially inside the cavity of the tetracationic cyclophane of the [2]catenane **13** (Fig. 2.14). Attractive [$\pi \cdots \pi$] stacking interactions between the neutral tetrathiafulvalene and the bipyridinium dications are responsible for this coconformation. Oxidation of the tetrathiafulvalene generates a cationic form that is expelled from the cavity of the tetracationic cyclophane. After the circumrotation of the macrocyclic polyether, the oxidized tetrathiafulvalene is exchanged

with the neutral 1,5-dioxynaphthalene producing the [2]catenane **14** (Fig. 2.14). The reduction of the tetrathiafulvalene back to its neutral state is followed by the circumrotation of the macrocyclic polyether, which restores the original state **14**. The voltage pulses applied to the bottom electrode of the electrode–monolayer–junction oxidize and reduce the tetrathiafulvalene unit inducing the interconversion between the forms **13** and **14**. The difference in the stereoelectronic properties of these two states translates into distinct current–voltage signatures. Indeed, their ability to mediate the tunneling of electrons across the junction differs significantly. As a result, the junction resistance probed at a low voltage after an oxidizing pulse is significantly different from that determined under the same conditions after a reducing pulse.

2.3.3 Self-Assembled Monolayers

In the examples illustrated in Figs. 2.12–2.15, monolayers of amphiphilic and electroactive derivatives are assembled at air–water interfaces and then transferred on the surfaces of appropriate substrates. An alternative strategy to coat electrodes with molecular layers relies on the ability of certain compounds to adsorb spontaneously on solid supports from liquid or vapor phases [2.36]. In particular, the affinity of certain sulfurated functional groups for gold can be exploited to encourage the self-assembly of organic molecules on microscaled and nanoscaled electrodes.

The electrode–monolayer–electrode junction in Fig. 2.15 incorporates a molecular layer between two gold electrodes mounted on a silicon nitride support. This device can be fabricated combining chemical vapor deposition, lithography, anisotropic etching and self-assembly [2.48]. Initially, a silicon wafer is coated with a 50 nm thick layer of silicon nitride by low pressure chemical vapor deposition. Then, a square of $400 \times 400\,\mu m$ is patterned on one side of the coated wafer by optical lithography and reactive ion etching. Anisotropic etching of the exposed silicon up to the other side of the wafer leaves a suspended silicon nitride membrane of $40 \times 40\,\mu m$. Electron beam lithography and reactive ion etching can be used to carve a bowl-shaped hole (diameter = 30–50 nm) in the membrane. Evaporation of gold on the membrane fills the pore producing a bowl-shaped electrode. Immersion of the substrate in a solution of the thiol **15** results in the self-assembly of a molecular layer on the narrow part of the bowl-shaped electrode. The subsequent evaporation of a gold film on the organic monolayer produces an electrode–monolayer–electrode junction (Fig. 2.15) with a contact area of less than $2000\,nm^2$ and approximately 1000 molecules.

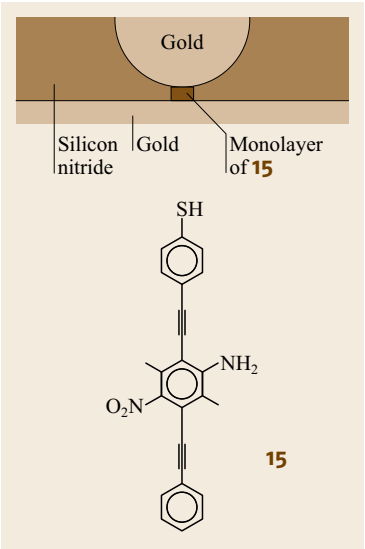


Fig. 2.15 A monolayer of the thiol **15** is embedded between two gold electrodes maintained in position by a silicon nitride support

Under the influence of voltage pulses applied to one of the two gold electrodes in Fig. 2.15, the conductivity of the sandwiched monolayer switches reversibly between low and high values [2.49]. In the initial state, the monolayer is in a low conducting mode. A current output of only 30 pA is detected, when a probing voltage of +0.25 V is applied to the bowl-shaped electrode. If the same electrode is stimulated with a short voltage pulse of +5 V, the monolayer switches to a high conducting mode. Now a current output of 150 pA is measured at the same probing voltage of +0.25 V. Repeated probing of the current output at various intervals of time indicates that the high conducting state is memorized by the molecule-based device and it is retained for more than 15 min. The low conducting mode is restored after either a relatively long period of time or the stimulation of the bowl-shaped electrode with a reverse voltage pulse of −5 V. Thus the current output switches from a low to a high value, if a high voltage input is applied. It switches from a high to a low value, under the influence of a low voltage pulse. This behavior offers the opportunity to store and erase binary data in analogy to a conventional random access memory [2.17]. Binary digits can be encoded on the current output of the molecule-based device applying a positive logic convention (low = 0, high = 1). It follows that a binary 1 can be stored in the molecule-based device applying a high voltage input and it can be erased applying a low voltage input [2.49].

The ability of thiols to self-assemble on the surface of gold can be exploited to fabricate nanocomposite materials integrating organic and inorganic components. For example, the bisthiol **16** forms monolayers (Fig. 2.16a) on gold electrodes with surface coverages

of ca. 4.1×10^{-10} mol cm^2 [2.50, 51]. The formation of a thiolate–gold bond at one of the two thiol ends of **16** is responsible for adsorption. The remaining thiol group points away from the supporting surface and can be exploited for further functionalization. Gold nanoparticles adsorb on the molecular layer (Fig. 2.16b), once again, as a result of thiolate–gold bond formation. The immersion of the resulting material in a methanol solution of **16** encourages the adsorption of an additional organic layer (Fig. 2.16c) on the composite material. Following these procedures, up to ten alternating organic and inorganic layers can be deposited on the electrode surface. The resulting assembly can mediate the unidirectional electron transfer from the supporting electrode to the redox active species in solution. For example, the cyclic voltammogram of the $[Ru(NH_3)_6]^{3+/2+}$ couple recorded with a bare gold electrode reveals a reversible reduction process. In the presence of ten alternating molecular and nanoparticle layers on the electrode surface, the reduction potential shifts by ca. −0.2 V and the back oxidation wave disappears. The pronounced potential shift indicates that $[Ru(NH_3)_6]^{3+}$ accepts electrons only after the surface-confined bipyridinium dications have been reduced. The lack of reversibility indicates that the back oxidation to the bipyridinium dications inhibits the transfer of electrons from the $[Ru(NH_3)_6]^{2+}$ to the electrode. Thus, the electroactive multilayer allows the flow of electrons in one direction only in analogy to conventional diodes.

The current–voltage behavior of individual nanoparticles in Fig. 2.16b can be probed by scanning tunneling spectroscopy in an aqueous electrolyte under an inert atmosphere [2.52]. The platinum-iridium tip of a scanning tunneling microscope is positioned above one of the gold particles. The voltage of the gold substrate relative to the tip is maintained at −0.2 V, while that relative to a reference electrode immersed in the same electrolyte is varied to control the redox state of the electroactive units. Indeed, the bipyridinium dications in the molecular layer can be reduced to a monocationic state reversibly. The resulting monocations can be reduced further and, once again, reversibly to a neutral form. Finally, the current flowing from the gold support to the tip of the scanning tunneling microscope is monitored as the tip–particle distance increases. From the distance dependence of the current, inverse length decays of ca. 16 and 7 nm^{-1} for the dicationic and monocationic states, respectively, of the molecular spacer can be determined. The dramatic decrease indicates that the reduction of the electroactive unit facilitates the tunneling of electrons through the gold–molecule–nanoparticle–tip junction. In summary, a change in the redox state of the bipyridinium components can be exploited to reversibly gate the current flowing through this nanoscaled device.

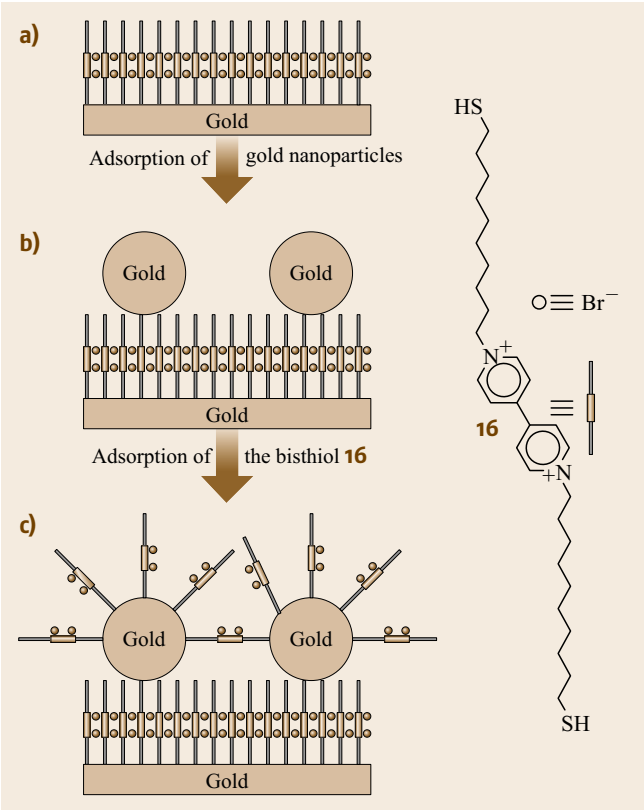


Fig. 2.16 **(a)** The bisthiol **16** self-assembles on gold electrodes as a result of thiolate–gold bond formation. **(b)** Gold nanoparticles adsorb spontaneously on the molecular layer. **(c)** Exposure of the composite assembly to a solution of **16** results in the formation of an additional molecular layer on the surface of the gold nanoparticles

Similar nanostructured materials, combining molecular and nanoparticle layers, can be prepared on layers of indium-tin oxide electrodes following multistep procedures [2.53]. The hydroxylated surfaces of indium-tin oxide supports can be functionalized with 3-ammoniumpropylsilyl groups and then exposed to gold nanoparticles having a diameter of ca. 13 nm [2.54, 55]. Electrostatic interactions promote the adsorption of the nanoparticles on the organic layer (Fig. 2.17a). The treatment of the composite film with the bipyridinium cyclophane **17** produces an organic layer on the gold nanoparticles (Fig. 2.18b). Following this approach, alternating layers of inorganic nanoparticles and organic building blocks can be assembled on the indium-tin oxide support. Cyclic voltammograms of the resulting materials show the oxidation of the gold nanoparticles and the reduction

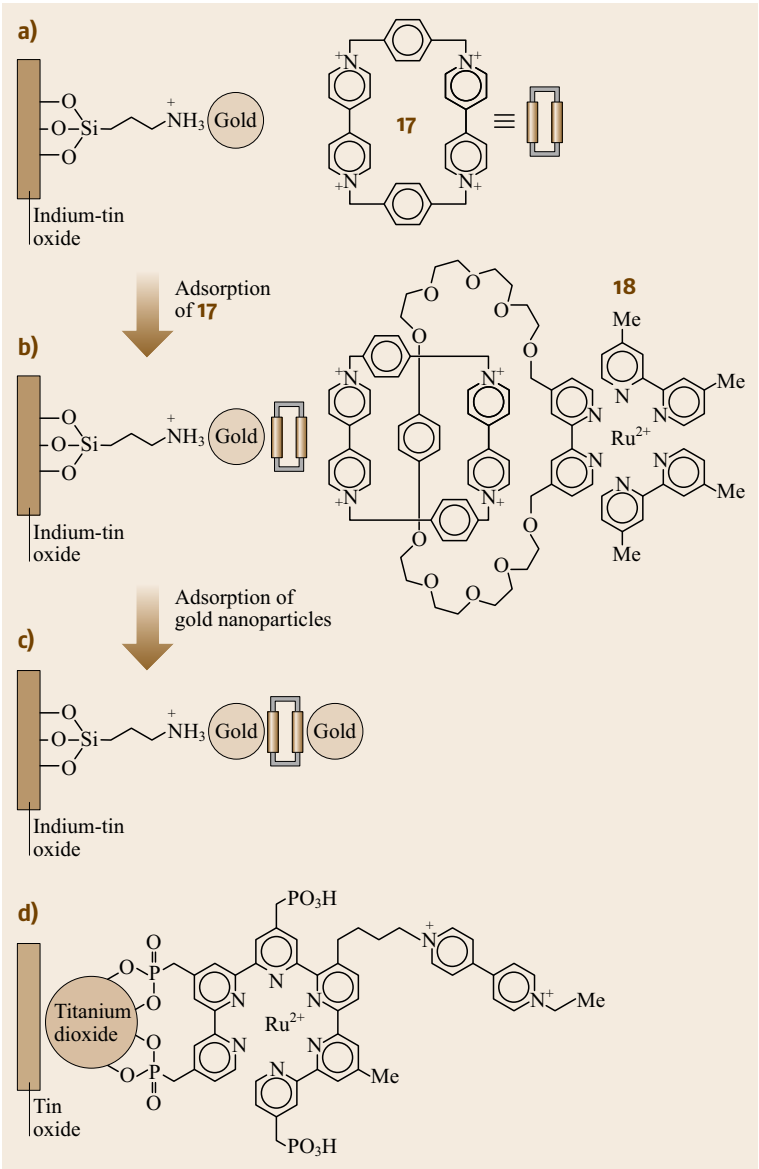


Fig. 2.17 **(a)** Gold nanoparticles assemble spontaneously on prefunctionalized indium-tin oxide electrodes. **(b)** Electrostatic interactions encourage the adsorption of the tetracationic cyclophane **17** on the surface-confined nanoparticles. **(c)** An additional layer of nanoparticles assembles on the cationic organic coating. Similar composite films can be prepared using the tetracationic [2]catenane **18** instead of the cyclophane **17**. **(d)** Phosphonate groups can be used to anchor molecular building blocks to titanium dioxide nanoparticles

of the bipyridinium units. The peak current for both processes increases with the number of alternating layers. Comparison of these values indicates that the ratio between the number of tetracationic cyclophanes and that of the nanoparticles is ca. 100:1.

The tetracationic cyclophane **17** binds dioxyarenes in solution [2.56, 57]. Attractive supramolecular forces between the electron deficient bipyridinium units and the electron-rich guests are responsible for complexation. This recognition motif can be exploited to probe the ability of the composite films in Fig. 2.17b,c to sense electron-rich analytes. In particular, hydroquinone is expected to enter the electron deficient cavities of the surface-confined cyclophanes. Consistently, cyclic voltammograms reveal the redox waves associated with the reversible oxidation of hydroquinone even when very small amounts of the guest (ca. 1×10^{-5} M) are added to the electrolyte solution [2.54, 55]. No redox response can be detected with a bare indium-tin oxide electrode under otherwise identical conditions. The supramolecular association of the guest and the surface-confined cyclophanes increases the local concentration of hydroquinone at the electrode–solution interface enabling its electrochemical detection.

Following a related strategy, the [2]catenane **18** (Fig. 2.17) can be incorporated into similar composite

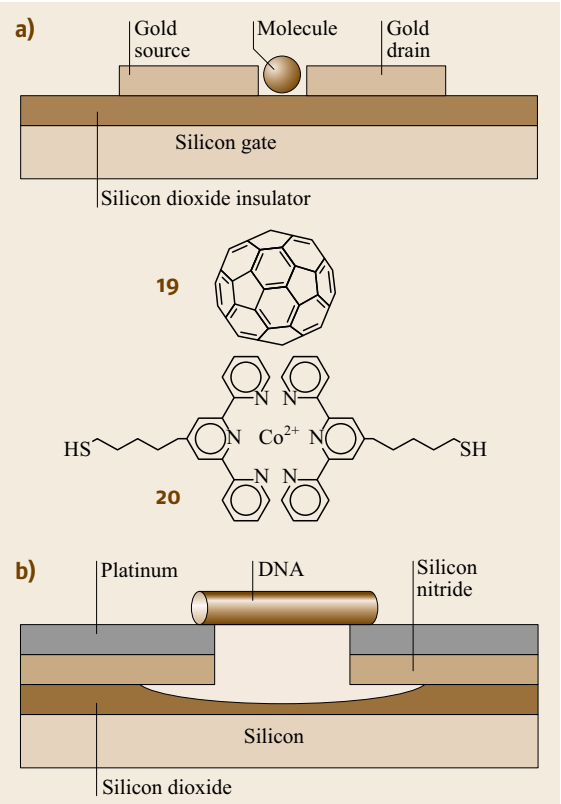


Fig. 2.18 **(a)** Nanoscaled transistors can be fabricated by inserting a single molecule (**19** or **20**) between source and drain electrodes mounted on a silicon/silicon dioxide support. **(b)** A DNA nanowire can bridge nanoelectrodes suspended above a silicon dioxide support

arrays [2.58, 59]. This interlocked molecule incorporates a Ru(II)/trisbipyridine sensitizer and two bipyridinium acceptors. Upon irradiation of the composite material at 440 nm, photoinduced electron transfer from the sensitizer to the appended acceptors occurs. The photogenerated hole in the sensitizer is filled after the transfer of an electron from a sacrificial electron donor present in the electrolyte solution. Under a positive voltage bias applied to the supporting electrode, an electron flow from the bipyridinium acceptors to the indium-tin oxide support is established. The resulting current switches between high and low values as the light source is turned off and on.

Another photoresponsive device, which is assembled combining inorganic nanoparticles with molecular building blocks, is illustrated in Fig. 2.17d. Phosphonate groups can be used to anchor a Ru(II)/trisbipyridine complex with an appended bipyridinium dication to titanium dioxide nanoparticles deposited on a doped tin oxide electrode [2.60, 61]. The resulting composite array can be integrated in a conventional electrochemical cell filled with an aqueous electrolyte containing triethanolamine. Under a bias voltage of −0.45 V and irradiation at 532 nm, 95% of the excited ruthenium centers transfer electrons to the titanium dioxide nanoparticles. The other 5% donate electrons to the bipyridinium dications. All the electrons transferred to the bipyridinium acceptors return to the ruthenium centers, while only 80% of those accepted by the nanoparticles return to the transition metal complexes. The remaining 15% reach the bipyridinium acceptors, while electron transfer from sacrificial triethanolamine donors fills the photogenerated holes left in the ruthenium sensitizers. The photoinduced reduction of the bipyridinium dication is accompanied by the appearance of the characteristic band of the radical cation in the absorption spectrum. This band persists for hours under open circuit conditions. However, it fades in ca. 15 s under a voltage bias of +1 V, as the radical cation is oxidized back to the dicationic form. In summary, an optical stimulation accompanied by a negative voltage bias reduces the bipyridinium building block. The state of the photogenerated form can be read optically, recording the absorption spectrum in the visible region, and erased electrically, applying a positive voltage pulse.

2.3.4 Nanogaps and Nanowires

The operating principles of the electroactive and photoactive devices illustrated in Figs. 2.12–2.17 exploit the ability of small collections of molecular components to manipulate electrons and photons. Designed molecules are deposited on relatively large electrodes and can be addressed electrically and/or optically by controlling the voltage of the support and/or illuminating its surface. The transition from devices relying on collections of molecules to unimolecular devices requires the identification of practical methods to contact single molecules. This fascinating objective demands the rather challenging miniaturization of contacting electrodes to the nanoscale.

A promising approach to unimolecular devices relies on the fabrication of nanometer-sized gaps in metallic features followed by the insertion of individual molecules between the terminals of the gap. This strategy permits the assembly of nanoscaled three-terminal devices equivalent to conventional transistors [2.62–64]. A remarkable example is illustrated in Fig. 2.18a [2.62]. It incorporates a single molecule in the nanogap generated between two gold electrodes. Initially, electron beam lithography is used to pattern a gold wire on a doped silicon wafer covered by an insulating silicon dioxide layer. Then, the gold feature is broken by electromigration to generate the nanogap.

The lateral size of the separated electrodes is ca. 100 nm and their thickness is ca. 15 nm. Scanning electron microcopy indicates that the facing surfaces of the separated electrodes are not uniform and that tiny gaps between their protrusions are formed. Current–voltage measurements suggest that the size of the smallest nanogap is ca. 1 nm. When the breakage of the gold feature is preceded by the deposition of a dilute toluene solution of C_{60} (**19**), junctions with enhanced conduction are obtained. This particular molecule has a diameter of ca. 0.7 nm and can insert in the nanogap, facilitating the flow of electrons across the junction.

The unique configuration of the molecule-based device in Fig. 2.18a can reproduce the functions of a conventional transistor [2.19] at the nanoscale. The two gold terminals of the junction are the drain and source of this nanotransistor and the underlying silicon wafer is the gate. At a temperature of 1.5 K, the junction conductance is very small, when the gate bias is low, and increases in steps at higher voltages [2.62]. The conductance gap is a consequence of the finite energy that is required to oxidize/reduce the single C_{60} positioned in the junction. Interestingly, the zero-conductance window also changes with the gate voltage and can be opened and closed reversibly adjusting the gate bias.

A similar strategy can be employed to fabricate a nanoscaled transistor incorporating the Co(II) complex **20** shown in Fig. 2.18 [2.64]. In this instance, a silicon dioxide layer with a thickness of ca. 30 nm is grown thermally on a doped silicon substrate. Then, a gold wire with a width of ca. 200 nm and a thickness of ca. 10−15 nm is patterned on the silicon dioxide overlayer by electron beam lithography. After extensive washing of the substrate with acetone and methylene chloride and cleaning with oxygen plasma, the gold wire is exposed to a solution of the bisthiol **20**. The formation of thiolate–gold bonds promotes the self-assembly of the molecular building block on the gold surface. At this point, electromigration-induced breakage produces a gap of 1−2 nm in the gold wire. The surface-confined bisthiol **20** is only 0.24 nm long and, therefore, it can insert in the nanogap producing an electrode–molecule–electrode junction.

The cobalt center in **20** can be oxidized/reduced reversibly between Co(II) and Co(III) states [2.64]. When this electroactive molecule is inserted in a nanogap (Fig. 2.18a), its ability to accept and donate electrons dictates the current–voltage profile of the resulting electrode–molecule–electrode junction. More precisely, no current flows across the junction below a certain voltage threshold. As the source voltage is raised above this particular value, the drain current increases in steps. Interestingly, the threshold associated with the source voltage varies in magnitude with the gate voltage. This intriguing behavior is a consequence of the finite energy necessary to oxidize/reduce the cobalt center and of a change in the relative stabilities of the oxidized and reduced forms Co(II) and Co(III) with the gate voltage. In summary, the conduction of the electrode–molecule–electrode junction can be tuned adjusting the voltage of the silicon support. The behavior of this molecule-based nanoelectronic device is equivalent to that of a conventional transistor [2.19]. In both instances, the gate voltage regulates the current flowing from the source to the drain.

The electromigration-induced breakage of preformed metallic features successfully produces nanogaps by moving apart two fragments of the same wire. Alternatively, nanogaps can be fabricated reducing the separation of the two terminals of much larger gaps. For example, gold electrodes separated by a distance of 20−80 nm can be patterned on a silicon/silicon dioxide substrate by electron beam lithography [2.65]. The relatively large gap between them can be reduced significantly by the electrochemical deposition of gold on the surfaces of both electrodes. The final result is the fabrication of two nanoelectrodes separated by ca. 1 nm and with a radius of curvature of 5−15 nm. The two terminals of this nanogap can be *contacted* by organic nanowires grown between them [2.66]. In particular, the electropolymerization of aniline produces polyaniline bridges between the gold nanoelectrodes. The conductance of the resulting junction can be probed immersing the overall assembly in an electrolyte solution. Employing a bipotentiostat, the bias voltage of the two terminals of the junction can be maintained at 20 mV, while their potentials are scanned relative to that of a silver/silver chloride reference electrode. Below ca. 0.15 V, the polymer wire is in an insulating state and the current flowing across the junction is less than 0.05 nA. At this voltage threshold, however, the current rises abruptly to ca. 30 nA. This value corresponds to a conductivity for the polymer nanojunction of 10−100 S cm^{-1}. When the potential is lowered again below the threshold, the current returns back to very low values. Interestingly, the abrupt decrease in current in the backward scan is observed at a potential that is slightly more negative than that causing the abrupt current increase in the forward scan. In summary, the conductance of this nanoscaled junction switches off and on as a potential input is switched above and below a voltage threshold.

It is interesting to note that the influence of organic bridges on the junction conductance can be exploited for chemical sensing. Nanogaps fabricated following a similar strategy, but lacking the polyaniline bridge, alter their conduction after exposure to dilute solutions of small organic molecules [2.67]. Indeed, the organic

analytes dock into the nanogaps producing a marked decrease in the junction conductance. The magnitude of the conductance drop happens to be proportional to the analyte–nanoelectrode binding strength. Thus, the presence of the analyte in solution can be detected probing the current–voltage characteristics of the nanogaps.

Nanogaps between electrodes patterned on silicon/silicon dioxide supports can be bridged also by DNA double strands [2.68, 69]. The device in Fig. 2.18b has a 10.4 nm long poly(G)–poly(C) DNA oligomer suspended between two nanoelectrodes. It can be fabricated patterning a 30 nm wide slit in a silicon nitride overlayer covering a silicon/silicon dioxide support by electron beam evaporation. Underetching the silicon dioxide layer leaves a silicon nitride finger, which can be sputtered with a platinum layer and chopped to leave a nanogap of 8 nm. At this point, a microdroplet of a dilute solution of DNA is deposited on the device and a bias of 5 V is applied between the two electrodes. Electrostatic forces encourage the deposition of a single DNA wire on top of the nanogap. As soon as the nanowire is in position, current starts to flow across the junction. The current–voltage signature of the electrode–DNA–electrode junction shows currents below 1 pA at low voltage biases. Under these conditions, the DNA nanowire is an insulator. Above a certain voltage threshold, however, the nanowire becomes conducting and currents up to 100 nA can flow across the junction through a single nanowire. Assuming that direct tunneling from electrode to electrode is extremely unlikely for a relatively large gap of 8 nm, the intriguing current–voltage behavior must be a consequence of the participation of the molecular states in the electron transport process. Two possible mechanisms can be envisaged. Sequential hopping of the electrons between states localized in the DNA base pairs can allow the current flow above a certain voltage threshold. However, this mechanism would presumably result in a Coulomb blockade voltage gap that is not observed experimentally. More likely, electronic states delocalized across the entire length of the DNA nanowire are producing a molecular conduction band. The off-set between the molecular conduction band and the Fermi levels of the electrodes is responsible for the insulating behavior at low biases. Above a certain voltage threshold, the molecular band and one of the Fermi levels align facilitating the passage of electrons across the junction.

Carbon nanotubes are extremely versatile building blocks for the assembly of nanoscaled electronic devices. They can be used to bridge nanogaps [2.70–73] and assemble nanoscaled cross junctions [2.74–76]. In Fig. 2.19a, a single-wall carbon nanotube crosses over another one in an orthogonal arrangement [2.74]. Both nanotubes have electrical contacts at their ends.

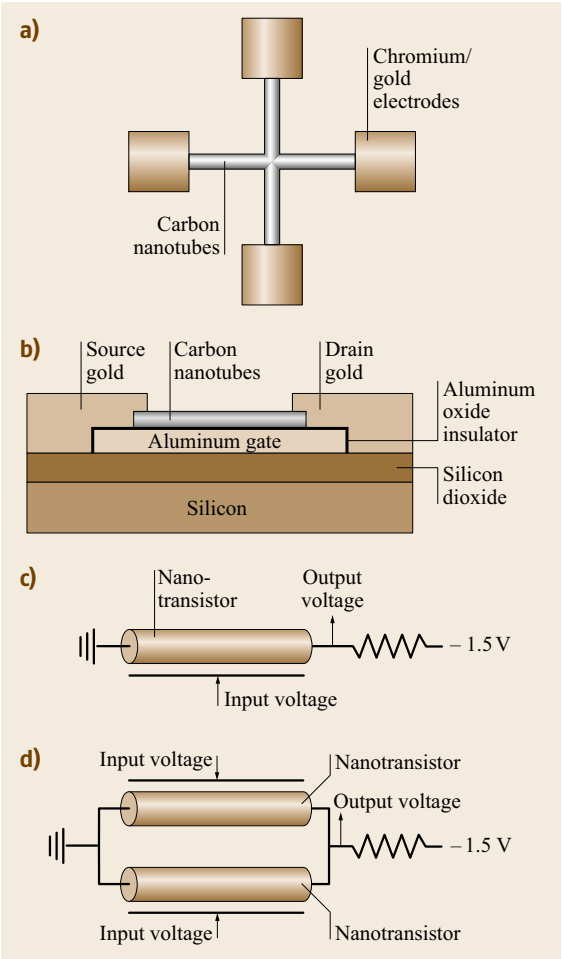


Fig. 2.19 **(a)** Nanoscaled junctions can be assembled on silicon/silicon dioxide supports crossing pairs of orthogonally arranged single-wall carbon nanotubes with chromium/gold electrical contacts at their ends. **(b)** Nanotransistors can be fabricated contacting the two ends of a single-wall carbon nanotube deposited on an aluminum/aluminum oxide gate with gold sources and drain. One or two nanotube transistors can be integrated into nanoscaled NOT **(c)** and NOR **(d)** logic gates

The fabrication of this device involves three main steps. First, alignment marks for the electrodes are patterned on a silicon/silicon dioxide support by electron beam lithography. Then, the substrate is exposed to a dichloromethane suspension of single-wall carbon nanotubes. After washing with isopropanol, crosses of carbon nanotubes in an appropriate alignment relative to the electrode marks are identified by tapping mode atomic force microscopy. Finally, chromium/gold electrodes are fabricated on top of the nanotube ends, again, by electron beam lithography. The conductance of in-

dividual nanotubes can be probed exploiting the two electric contacts at their ends. These two-terminal measurements reveal that certain nanotubes have metallic behavior, while others are semiconducting. It follows that three distinct types of cross junctions differing in the nature of their constituent nanotubes can be identified on the silicon/silicon dioxide support. Four terminal current–voltage measurements indicate that junctions formed by two metallic nanotubes have high conductance and ohmic behavior. Similarly, high junction conductance and ohmic behavior is observed when two semiconducting nanotubes cross. The current–voltage signature of junctions formed when a metallic nanotube crosses a semiconducting one are, instead, completely different. The metallic nanotube depletes the semiconducting one at the junction region producing a nanoscaled Schottky barrier with a pronounced rectifying behavior.

Similar fabrication strategies can be exploited to assemble nanoscaled counterparts of conventional transistors. The device in Fig. 2.19b is assembled patterning an aluminum finger on a silicon/silicon dioxide substrate by electron beam lithography [2.76]. After exposure to air, an insulating aluminum oxide layer forms on the aluminum finger. Then, a dichloromethane suspension of single-wall carbon nanotubes is deposited on the resulting substrate. Atomic force microscopy can be used to select carbon nanotubes with a diameter of ca. 1 nm that are positioned on the aluminum finger. After registering their coordinates relative to alignment markers, gold contacts can be evaporated on their ends by electron beam lithography. The final assembly is a nanoscaled three-terminal device equivalent to a conventional field effect transistor [2.19]. The two gold contacts are the source and drain terminals, while the underlying aluminum finger reproduces the function of the gate. At a source to drain bias of ca. −1.3 V, the drain current jumps from ca. 0 to ca. 50 nA when the gate voltage is lowered from −1.0 to −1.3 V. Thus, moderate changes in the gate voltage vary significantly the current flowing through the nanotube-based device in analogy to a conventional enhancement mode p-type field effect transistor [2.19].

The nanoscaled transistor in Fig.2.18a has a microscaled silicon gate that extends under the entire chip [2.62, 64]. The configuration in Fig. 2.19b, instead, has nanoscaled aluminum gates for every single carbon nanotube transistor fabricated on the same support [2.76]. It follows that multiple nanoscaled transistors can be fabricated on the same chip and operated independently following this strategy. This unique feature offers the possibility of fabricating nanoscaled digital circuits by interconnecting the terminals of independent nanotube transistors. The examples in Fig. 2.19c,d illustrate the configurations of nanoscaled NOT and NOR gates implemented using one or two nanotube transistors. In Fig. 2.19c, an off-chip bias resistor is connected to the drain terminal of a single transistor while the source is grounded. A voltage input applied to the gate modulates the nanotube conductance altering the voltage output probed at the drain terminal. In particular, a voltage input of −1.5 V lowers the nanotube resistance (26 MΩ) below that of the bias resistor (100 MΩ). As a result, the voltage output drops to 0 V. When the voltage input is raised to 0 V, the nanotube resistance increases above that of the bias resistor and the voltage output becomes −1.5 V. Thus, the output of this nanoelectronic device switches from a high (0 V) and to a low (−1.5 V) level as the input shifts from a low (−1.5 V) to a high (0 V) value. The inverse relation between input and output translates into a NOT operation if a negative logic convention (low = 1, high = 0) is applied to both signals.

In Fig. 2.15d, the source terminals of two independent nanotube transistors fabricated on the same chip are connected by a gold wire and grounded [2.76]. Similarly, the two drain terminals are connected by another gold wire and contacted to an off-chip bias resistor. The gate of each nanotube can be stimulated with a voltage input and the voltage output of the device can be probed at their interconnected drain terminals. When the resistance of at least one of the two nanotubes is below that of the resistor, the output is 0 V. When both nanotubes are in a nonconducting mode, the output voltage is −1.5 V. Thus, if a low voltage input −1.5 V is applied to one or both transistors, the output is high (0 V). When both voltage inputs are high (0 V), the output is low (−1.5 V). If a negative logic convention (low = 1, high = 0) is applied to all signals, the signal transduction behavior translates into a NOR operation.

2.4 Conclusions and Outlook

Nature builds nanostructured biomolecules relying on a highly modular approach [2.1]. Small building blocks are connected by robust chemical bonds to generate long strands of repeating units. The synergism of a multitude of attractive supramolecular forces determines the three-dimensional arrangement of the resulting polymeric chains and controls the association of independent strands into single and well-defined entities. Nucleic acids and proteins are two representative classes of biomolecules assembled with sub-

nanometer precision through the subtle interplay of covalent and noncovalent bonds starting from a relatively small pool of nucleotide and amino acid building blocks.

The power of chemical synthesis [2.2] offers the opportunity of mimicking nature's modular approach to nanostructured materials. Following established experimental protocols, small molecular building blocks can be joined together relying on the controlled formation of covalent bonds between designed functional groups. Thus, artificial molecules with nanoscaled dimensions can be assembled piece by piece with high structural control. Indeed, helical, tubular, interlocked and highly branched nanostructures have been prepared, and already exploit this general strategy and the synergism of covalent and noncovalent bonds [2.3].

The chemical construction of nanoscaled molecules from modular building blocks also offers the opportunity of engineering specific properties in the resulting assemblies. In particular, electroactive and photoactive fragments can be integrated into single molecules. The ability of these functional subunits to accept/donate electrons and photons can be exploited to design nanoscaled electronic and photonic devices. Indeed, molecules that respond to electrical and optical stimulations producing detectable outputs have been designed already [2.15]. These chemical systems can be employed to control the interplay of input and output signals at the molecular level. Their conceptual analogy with the signal transduction operated by conventional logic gates in digital circuits is evident. In fact, electroactive and photoactive molecules able to reproduce AND, NOT and OR operations as well as simple combinations of these basic logic functions, are already a reality [2.20, 21].

Most of the molecular switches for digital processing developed so far rely on bulk addressing. In general, relatively large collections of functional molecules are addressed simultaneously in solution. The realization of molecule-based devices with reduced dimensions as well as practical limitations associated with liquid phases in potential applications are encouraging a transition from the solution to the solid state. The general strategy followed so far relies on the deposition of functional molecules on the surfaces of appropriate electrodes following either the Langmuir–Blodgett methodology [2.77] or self-assembly processes [2.36]. The combination of these techniques with the nanofabrication of insulating, metallic and semiconducting features on appropriate supports has already allowed for the realization of fascinating molecule-based devices [2.31–34, 53]. The resulting assemblies integrate inorganic and organic components and, in some instances, even biomolecules to execute specific functions. They can convert optical stimulations into electrical signals. They can execute irreversible and reversible switching operations. They can sense qualitatively and quantitatively specific analytes. They can reproduce the functions of conventional rectifiers and transistors. They can be integrated within functioning nanoelectronic devices capable of simple logic operations.

The remarkable examples of molecule-based materials and devices already available are an obvious demonstration of the great potential and promise for this research area. At this stage, the only limit factor of designing functional molecules is the imagination of the synthetic chemist. All sorts of molecular building blocks with tailored dimensions, shapes and properties are more or less accessible with the assistance of modern chemical synthesis. Now the major challenges are (1) to master the operating principles of the molecule-based devices that have been and continue to be assembled, and (2) to expand and improve the fabrication strategies available to incorporate molecules into reliable device architectures. As we continue to gather further insights in these directions, design criteria for a wide diversity of molecule-based devices will emerge. It is not unrealistic to foresee the evolution of an entire generation of nanoscaled devices, based on engineered molecular components, which will find applications in a variety of fields ranging from biomedical research to information technology. Perhaps, Nature can once again illuminate our path teaching us not only how to synthesize nanostructured molecules but also how to use them. After all, Nature is full of examples of extremely sophisticated molecule-based devices. From tiny bacteria to higher animals, we are all a collection of molecule-based devices!

References

2.1 D. Voet, J.G. Voet: *Biochemistry* (Wiley, New York 2010)

2.2 K.C. Nicolau, J.S. Chen: *Classics in Total Synthesis* (Wiley-VCH, Weinheim 2011)

2.3 J.W. Steed, J.L. Atwood: *Supramolecular Chemistry* (Wiley, New York 2009)

2.4 M.M. Harding, U. Koert, J.-M. Lehn, A. Marquis-Rigault, C. Piguet, J. Siegel: Synthesis of unsubstituted and 4,4′-substituted oligobipyridines as ligand strands for helicate self-assembly, Helv. Chim. Acta **74**, 594–610 (1991)

2.5 J.-M. Lehn, A. Rigault, J. Siegel, B. Harrowfield, B. Chevrier, D. Moras: Spontaneous assembly of double-stranded helicates from oligobipyridine ligands and copper(I) cations: Structure of an inorganic double helix, Proc. Natl. Acad. Sci. USA **84**, 2565–2569 (1987)

2.6 J.-M. Lehn, A. Rigault: Helicates: Tetra- and pentanuclear double helix complexes of Cu(I) and poly(bipyridine) strands, Angew. Chem. Int. Ed. Engl. **27**, 1095–1097 (1988)

2.7 J.D. Hartgerink, J.R. Granja, R.A. Milligan, M.R. Ghadiri: Self-assembling peptide nanotubes, J. Am. Chem. Soc. **118**, 43–50 (1996)

2.8 M.R. Ghadiri, J.R. Granja, R.A. Milligan, D.E. McRee, N. Khazanovich: Self-assembling organic nanotubes based on a cyclic peptide architecture, Nature **366**, 324–327 (1993)

2.9 V. Balzani, A. Credi, M. Venturi: *Molecular Devices and Machines: Concepts and Perspectives for the Nanoworld* (Wiley-VCH, Weinheim 2008)

2.10 C.H. Hamann, A. Hamnett, W. Vielstich: *Electrochemistry* (Wiley-VCH, Weinheim 2007)

2.11 S. Fukuzumi: *Electron Transfer: Mechanisms and Applications* (Wiley-VCH, Weinheim 2016)

2.12 V. Balzani, P. Ceroni, A. Juris: *Photochemistry and Photophysics: Concepts, Research, Applications* (Wiley-VCH, Weinheim 2014)

2.13 N.J. Turro, J.C. Scaiano, V. Ramamurthy: *Modern Molecular Photochemistry of Organic Molecules* (University Science Books, Herndon 2010)

2.14 P.R. Ashton, R. Ballardini, V. Balzani, A. Credi, K.R. Dress, E. Ishow, C.J. Kleverlaan, O. Kocian, J.A. Preece, N. Spencer, J.F. Stoddart, M. Venturi, S. Wenger: A photochemically driven molecular-level abacus, Chem. Eur. J. **6**, 3558–3574 (2000)

2.15 B.L. Feringa, W.R. Browne (Eds.): *Molecular Switches* (Wiley-VCH, Weinheim 2011)

2.16 M. Irie, Y. Yokoyama, T. Seki (Eds.): *New Frontiers in Photochromism* (Springer, Tokyo 2013)

2.17 J.-P. Deschamps, E. Valderrama, L. Terés: *Digital Systems: From Logic Gates to Processors* (Springer, New York 2016)

2.18 D.R. Smith: *Digital Transmission Systems* (Springer, New York 2003)

2.19 O. Bishop: *Electronics: Circuits and Systems* (Routledge, Burlington 2011)

2.20 F.M. Raymo: Digital processing and communication with molecular switches, Adv. Mater. **14**, 401–414 (2002)

2.21 A.P. de Silva: Molecular computation – Molecular logic gets loaded, Nature Mater. **4**, 15–16 (2005)

2.22 A. Aviram: Molecules for memory, Logic Amplif, J. Am. Chem. Soc. **110**, 5687–5692 (1988)

2.23 A.P. de Silva, H.Q.N. Gunaratne, C.P. McCoy: A molecular photoionic AND gate based on fluorescent signaling, Nature **364**, 42–44 (1993)

2.24 M. Asakawa, P.R. Ashton, V. Balzani, A. Credi, G. Mattersteig, O.A. Matthews, M. Montalti, N. Spencer, J.F. Stoddart, M. Venturi: Electrochemically induced molecular motions in pseudorotaxanes: A case of dual-mode (oxidative and reductive) dethreading, Chem. Eur. J. **3**, 1992–1996 (1997)

2.25 F.M. Raymo, S. Giordani, A.J.P. White, D.J. Williams: Digital processing with a three-state molecular switch, J. Org. Chem. **68**, 4158–4169 (2003)

2.26 F.M. Raymo, S. Giordani: Signal communication between molecular switches, Org. Lett. **3**, 3475–3478 (2001)

2.27 F.M. Raymo, S. Giordani: Digital communication through intermolecular fluorescence modulation, Org. Lett. **3**, 1833–1836 (2001)

2.28 F.M. Raymo, S. Giordani: Multichannel digital transmission in an optical network of communicating molecules, J. Am. Chem. Soc. **124**, 2004–2007 (2002)

2.29 F.M. Raymo, S. Giordani: All-optical processing with molecular switches, Proc. Natl. Acad. Sci. USA **99**, 4941–4944 (2002)

2.30 J.C. Cuevas, E. Scheer: *Molecular Electronics: An Introduction to Theory and Experiment* (World Scientific Publishing, Singapore 2013)

2.31 C. Joachim, J.K. Gimzewski, A. Aviram: Electronics using hybrid-molecular and mono-molecular devices, Nature **408**, 541–548 (2000)

2.32 J.M. Tour: Molecular electronics. Synthesis and testing of components, Acc. Chem. Res. **33**, 791–804 (2000)

2.33 A.R. Pease, J.O. Jeppesen, J.F. Stoddart, Y. Luo, C.P. Collier, J.R. Heath: Switching devices based on interlocked molecules, Acc. Chem. Res. **34**, 433–444 (2001)

2.34 R.M. Metzger: Unimoleular electrical rectifiers, Chem. Rev. **103**, 3803–3834 (2003)

2.35 W.A. Barlow (Ed.): *Langmuir-Blodgett Films* (North Holland, Amsterdam 2013)

2.36 Y.S. Lee: *Self-Assembly and Nanotechnology Systems: Design, Characterization and Applications* (Wiley, New York 2011)

2.37 C. Lee, A.J. Bard: Comparative electrochemical studies of *N*-methyl- *N'*-hexadecyl viologen monomolecular films formed by irreversible adsorption and the Langmuir-Blodgett method, J. Electroanal. Chem. **239**, 441–446 (1988)

2.38 C. Lee, A.J. Bard: Cyclic voltammetry and Langmuir film isotherms of mixed monolayers of *N*-docosoyl-*N'*-methyl viologen with arachidic acid, Chem. Phys. Lett. **170**, 57–60 (1990)

2.39 M. Fujihira, K. Nishiyama, H. Yamada: Photoelectrochemical responses of optically transparent electrodes modified with Langmuir-Blodgett films consisting of surfactant derivatives of electron donor, acceptor and sensitizer molecules, Thin Solid Films **132**, 77–82 (1985)

2.40 M. Fujihira: Photoelectric conversion with Langmuir-Blodgett films. In: *Nanostructures Based on Molecular Materials*, ed. by W. Göpel, C. Ziegler (VCH, Weinheim 1992) pp. 27–46

2.41 C.P. Collier, E.W. Wong, M. Belohradsky, F.M. Raymo, J.F. Stoddart, P.J. Kuekes, R.S. Williams, J.R. Heath: Electronically configurable molecular-based logic gates, Science **285**, 391–394 (1999)

2.42 E.W. Wong, C.P. Collier, M. Belohradsky, F.M. Raymo, J.F. Stoddart, J.R. Heath: Fabrication and transport properties of single-molecule-thick electrochemical junctions, J. Am. Chem. Soc. **122**, 5831–5840 (2000)

2.43 M. Asakawa, P.R. Ashton, V. Balzani, A. Credi, C. Hamers, G. Mattersteig, M. Montalti, A.N. Shipway, N. Spencer, J.F. Stoddart, M.S. Tolley, M. Venturi, A.J.P. White, D.J. Williams: A chemically and electrochemically switchable [2]catenane incorporating a tetrathiafulvalene unit, Angew. Chem. Int. Ed. **37**, 333–337 (1998)

2.44 V. Balzani, A. Credi, G. Mattersteig, O.A. Matthews, F.M. Raymo, J.F. Stoddart, M. Venturi, A.J.P. White, D.J. Williams: Switching of pseudorotaxanes and catenanes incorporating a tetrathiafulvalene unit by redox and chemical inputs, J. Org. Chem. **65**, 1924–1936 (2000)

2.45 M. Asakawa, M. Higuchi, G. Mattersteig, T. Nakamura, A.R. Pease, F.M. Raymo, T. Shimizu, J.F. Stoddart: Current/voltage characteristics of monolayers of redox-switchable [2]catenanes on gold, Adv. Mater. **12**, 1099–1102 (2000)

2.46 C.P. Collier, G. Mattersteig, E.W. Wong, Y. Luo, K. Beverly, J. Sampaio, F.M. Raymo, J.F. Stoddart, J.R. Heath: A [2]catenane based solid-state electronically reconfigurable switch, Science **289**, 1172–1175 (2000)

2.47 C.P. Collier, J.O. Jeppesen, Y. Luo, J. Perkins, E.W. Wong, J.R. Heath, J.F. Stoddart: Molecular-based electronically switchable tunnel junction devices, J. Am. Chem. Soc. **123**, 12632–12641 (2001)

2.48 J. Chen, M.A. Reed, A.M. Rawlett, J.M. Tour: Large on-off ratios and negative differential resistance in a molecular electronic device, Science **286**, 1550–1552 (1999)

2.49 M.A. Reed, J. Chen, A.M. Rawlett, D.W. Price, J.M. Tour: Molecular random access memory cell, Appl. Phys. Lett. **78**, 3735–3737 (2001)

2.50 D.I. Gittins, D. Bethell, R.J. Nichols, D.J. Schiffrin: Redox-controlled multilayers of discrete gold particles: A novel electroactive nanomaterial, Adv. Mater. **9**, 737–740 (1999)

2.51 D.I. Gittins, D. Bethell, R.J. Nichols, D.J. Schiffrin: Diode-like electron transfer across nanostructured films containing a redox ligand, J. Mater. Chem. **10**, 79–83 (2000)

2.52 D.I. Gittins, D. Bethell, D.J. Schiffrin, R.J. Nichols: A nanometer-scale electronic switch consisting of a metal cluster and redox-addressable groups, Nature **408**, 67–69 (2000)

2.53 A.N. Shipway, M. Lahav, I. Willner: Nanostructured gold colloid electrodes, Adv. Mater. **12**, 993–998 (2000)

2.54 A.N. Shipway, M. Lahav, R. Blonder, I. Willner: Bis-bipyridinium cyclophane receptor–Au nanoparticle superstructure for electrochemical sensing applications, Chem. Mater. **11**, 13–15 (1999)

2.55 M. Lahav, A.N. Shipway, I. Willner, M.B. Nielsen, J.F. Stoddart: An enlarged bis-bipyridinum cyclophane–Au nanoparticle superstructure for selective electrochemical sensing applications, J. Electroanal. Chem. **482**, 217–221 (2000)

2.56 R.E. Gillard, F.M. Raymo, J.F. Stoddart: Controlling self-assembly, Chem. Eur. J. **3**, 1933–1940 (1997)

2.57 F.M. Raymo, J.F. Stoddart: From supramolecular complexes to interlocked molecular compounds, Chemtracts – Org. Chem. **11**, 491–511 (1998)

2.58 M. Lahav, T. Gabriel, A.N. Shipway, I. Willner: Assembly of a Zn(II)-porphyrin-bipyridinium dyad and Au-nanoparticle superstructures on conductive surfaces, J. Am. Chem. Soc. **121**, 258–259 (1999)

2.59 M. Lahav, V. Heleg-Shabtai, J. Wasserman, E. Katz, I. Willner, H. Durr, Y. Hu, S.H. Bossmann: Photoelectrochemistry with integrated photosensitizer-electron acceptor Au-nanoparticle arrays, J. Am. Chem. Soc. **122**, 11480–11487 (2000)

2.60 G. Will, S.N. Rao, D. Fitzmaurice: Heterosupramolecular optical write-read-erase device, J. Mater. Chem. **9**, 2297–2299 (1999)

2.61 A. Merrins, C. Kleverlann, G. Will, S.N. Rao, F. Scandola, D. Fitzmaurice: Time-resolved optical spectroscopy of heterosupramolecular assemblies based on nanostructured TiO_2 films modified by chemisorption of covalently linked ruthenium and viologen complex components, J. Phys. Chem. B **105**, 2998–3004 (2001)

2.62 H. Park, J. Park, A.K.L. Lim, E.H. Anderson, A.P. Alivisatos, P.L. McEuen: Nanomechanical oscillations in a single C_{60} transistor, Nature **407**, 57–60 (2000)

2.63 W. Liang, M.P. Shores, M. Bockrath, J.R. Long, H. Park: Kondo resonance in a single-molecule transistor, Nature **417**, 725–729 (2002)

2.64 J. Park, A.N. Pasupathy, J.I. Goldsmith, C. Chang, Y. Yaish, J.R. Petta, M. Rinkoski, J.P. Sethna, H.D. Abruna, P.L. McEuen, D.C. Ralph: Coulomb blockade and the Kondo effect in single-atom transistors, Nature **417**, 722–725 (2002)

2.65 C.Z. Li, H.X. He, N.J. Tao: Quantized tunneling current in the metallic nanogaps formed by electrodeposition and etching, Appl. Phys. Lett. **77**, 3995–3997 (2000)

2.66 H. He, J. Zhu, N.J. Tao, L.A. Nagahara, I. Amlani, R. Tsui: A conducting polymer nanojunction switch, J. Am. Chem. Soc. **123**, 7730–7731 (2001)

2.67 A. Bogozi, O. Lam, H. He, C. Li, N.J. Tao, L.A. Nagahara, I. Amlani, R. Tsui: Molecular adsorption onto metallic quantum wires, J. Am. Chem. Soc. **123**, 4585–4590 (2001)

2.68 A. Bezryadin, C.N. Lau, M. Tinkham: Quantum suppression of superconductivity in ultrathin nanowires, Nature **404**, 971–974 (2000)

2.69 D. Porath, A. Bezryadin, S. de Vries, C. Dekker: Direct measurement of electrical transport through DNA molecules, Nature **403**, 635–638 (2000)

2.70 S.J. Tans, M.H. Devoret, H. Dai, A. Thess, E.E. Smalley, L.J. Geerligs, C. Dekker: Individual single-wall carbon nanotubes as quantum wires, Nature **386**, 474–477 (1997)

2.71 A.F. Morpurgo, J. Kong, C.M. Marcus, H. Dai: Gate-controlled superconducting proximity effect in carbon nanotubes, Nature **286**, 263–265 (1999)

2.72 J. Nygård, D.H. Cobden, P.E. Lindelof: Kondo physics in carbon nanotubes, Nature **408**, 342–346 (2000)

2.73 W. Liang, M. Bockrath, D. Bozovic, J.H. Hafner, M. Tinkham, H. Park: Fabry-Perot interference in a nanotube electron waveguide, Nature **411**, 665–669 (2001)

2.74 M.S. Fuhrer, J. Nygård, L. Shih, M. Forero, Y.-G. Yoon, M.S.C. Mazzoni, H.J. Choi, J. Ihm, S.G. Louie, A. Zettl, P.L. McEuen: Crossed nanotube junctions, Science **288**, 494–497 (2000)
2.75 T. Rueckes, K. Kim, E. Joselevich, G.Y. Tseng, C.-L. Cheung, C.M. Lieber: Carbon nanotube-based nonvolatile random access memory for molecular computing, Science **289**, 94–97 (2000)
2.76 A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker: Logic circuits with carbon nanotube transistors, Science **294**, 1317–1320 (2001)
2.77 M.C. Petty: *Langmuir-Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)

3. Introduction to Micro-/Nanofabrication

Gemma Rius, Antoni Baldi, Babak Ziaie, Massood Z. Atashbar

This chapter outlines and discusses important micro- and nanofabrication techniques. We focus on the most basic methods borrowed from the integrated circuit (IC) industry, such as thin-film deposition, lithography and etching, and then move on to look at microelectromechanical systems (MEMS) and nanofabrication technologies. We cover a broad range of dimensions, from the submillimeter to the nanometer scale. Although most of the current research is being geared towards the nanodomain, a good understanding of basic top-down methods for fabricating micron-sized objects can aid our understanding of this research. Due to space constraints, we focus here on the most important technologies; in the microdomain these include surface, bulk, and high-aspect-ratio micromachining; in the nanodomain we concentrate on e-beam lithography, epitaxial growth, scanning probe lithography, template manufacturing, and self-assembly. MEMS technology has matured rapidly, with some new technologies displacing older ones that have proven to be unsuited to manufacture on a commercial scale. However, due to limitations encountered on these methods used in the nanodomain, it appears that bottom-up methods or introduction of novel nanoforms and nanomaterials are the most feasible and promising solutions. Disruptive approaches are expected to have a major impact in a variety of application areas such as biology, medicine, environmental monitoring, and nanoelectronics.

3.1 **Basic Microfabrication Techniques** 52
3.1.1 Lithography 52
3.1.2 Thin-Film Deposition and Doping 53
3.1.3 Etching and Substrate Removal 58
3.1.4 Substrate Bonding 62
3.2 **MEMS Fabrication Techniques** 62
3.2.1 Bulk Micromachining 63
3.2.2 Surface Micromachining 66
3.2.3 High-Aspect-Ratio Micromachining 70
3.3 **Nanofabrication Techniques** 74
3.3.1 E-Beam Nanofabrication 75
3.3.2 Epitaxy and Strain Engineering 75
3.3.3 Scanning Probe Microscope Techniques 75
3.3.4 Self-Assembly and Template Manufacturing 78
3.4 **Summary and Conclusions** 82
References 82

Continuous efforts in the area of micro and nanofabrication have created a unique opportunity for manufacturing structures in the nm–mm range, particularly, utilizing the robust methods of semiconductor technology. The available six orders of magnitude dimensional span can be used to fabricate novel electronic, optical, magnetic, mechanical, and chemical/biological devices with applications ranging from sensors to computation and control. In this chapter, we will introduce major micro/nanofabrication techniques currently used to produce structures from the nm to several hundred μm range. We will mainly focus on the most important and widely used techniques and will not discuss specialized methods. After a brief introduction to basic microfabrication, we will discuss MEMS fabrication techniques used to build microstructures down to about 1 μm in dimensions. Following this, we will briefly discuss several top-down and bottom-up nanofabrication methods. As a comprehensive review of the diversity of nanofabrication techniques used nowadays is out of the scope of this chapter, we will refer to other dedicated chapters in this Handbook or provide relevant literature for the main ones.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_3

3.1 Basic Microfabrication Techniques

Most micro/nanofabrication techniques have their roots in the standard fabrication methods developed for the semiconductor industry [3.1–3]. Therefore, a clear understanding of these techniques is necessary for anyone starting to embark on a research and development path in the micro/nano area. In this section, we will discuss the major microfabrication methods used most frequently in the manufacturing of micro/nanostructures. Some of these techniques such as thin-film deposition and etching are common between the micro/nano and very large-scale integration (VLSI) microchip fabrication disciplines. However, several other techniques which are more specific to the micro/nanofabrication area will also be discussed in this section. In essence, silicon technology is based on a combination or sequence of additive and substractive methods, such as ion implantation or sputtering and physical etching or resist development, respectively; typically applied together with a number of lithography or direct patterning steps.

3.1.1 Lithography

Lithography is the technique used to transfer a computer-generated pattern onto a substrate (silicon, glass, GaAs, etc.). This pattern is subsequently used to etch an underlying thin film (oxide, nitride, etc.) for various purposes (doping, etching, etc.). Although photolithography, i. e., lithography using an ultraviolet (UV) light source, is by far the most widely used lithography technique in the microelectronic fabrication, electron-beam (e-beam) and x-ray lithography are two other alternatives which have attracted considerable attention in the MEMS and nanofabrication areas. We will discuss photolithography in this section and postpone discussion of e-beam and x-ray techniques to subsequent sections dealing with MEMS and nanofabrication.

The starting point subsequent to the creation of the computer layout for a specific fabrication sequence is the generation of a photomask. This involves a sequence of photographic processes (using optical or e-beam pattern generators), which results in a glass plate having the desired pattern in the form of a thin ($\approx 100\,\mathrm{nm}$) chromium layer. Following the generation of the photomask, the lithography process can proceed as shown in Fig. 3.1. This sequence demonstrates the pattern transfer onto a Si substrate coated with silicon dioxide; however, the same technique is applicable to other materials. After depositing the desired material on the substrate, the photolithography process starts with spin-coating the substrate with a photoresist. This is a polymeric photosensitive material which can be spun onto the wafer in liquid from (usually an adhesion promoter such as hexamethyldisilazane (HMDS) is used prior to the application of the resist). The spinning speed and photoresist viscosity will determine the final resist thickness, which is typically in the range $0.5-2.5\,\mu\mathrm{m}$. Two different kinds of photoresist are available: positive and negative. With positive resist, UV-exposed areas will be dissolved in the subsequent development stage, whereas with negative photoresist, the exposed areas will remain intact after UV development. Due to its better performance with regard to process control in small geometries, positive resist is the most extensively used in the VLSI processes. After spinning the photoresist onto the wafer, the substrate is soft-baked ($5-30\,\mathrm{min}$ at $60-100\,^{\circ}\mathrm{C}$) in order to remove the solvents from the resist and improve adhesion. Subsequently, the mask is aligned to the wafer and the photoresist is, therefore, selectively exposed to a UV source.

Depending on the separation between the mask and the wafer three different exposure systems are available:

1. Contact
2. Proximity
3. Projection.



Fig. 3.1 Lithography process flow

Although contact printing gives better resolution compared with the proximity technique, the constant contact of the mask with the photoresist reduces the process yield and can damage the mask. Projection printing uses a dual-lens optical system to project the mask image onto the wafer. Since only one die at a time can be exposed, this requires a step-and-repeat system to cover the whole wafer area. Projection printing is by far the most widely used system in microfabrication and can yield superior resolutions compared with the contact and proximity methods. The exposure source for photolithography depends on the resolution (minimum feature size and pattern density). Above 0.25 μm minimum line width, a high-pressure mercury lamp is adequate (436 nm g-line and 365 nm i-line). However, between 0.25 and 0.13 μm, deep-UV sources such as excimer lasers (248 nm KrF and 193 nm ArF) are required. Although there has been extensive competition for the below-0.13 μm regime (including e-beam and x-ray), extreme UV (EUV) with wavelength of 10–14 nm seems to be the preferred technique, although major technical challenges still remain [3.4]. Immersion lithography (i. e., using a liquid in the space between the lens and substrate in order to increase the numerical aperture), a recent innovation, has allowed the minimum feature size to be reduced to 32 nm without the requirement for EUV sources [3.5].

After exposure, the photoresist is developed in a chemical process similar to the development of photographic films. The resist is subsequently hard-baked (20–30 min at 120–180 °C) in order to further improve adhesion. The hard-bake step concludes the photolithography sequence by creating the desired pattern on the wafer. Next, the underlying thin film is etched and the photoresist is stripped using acetone or other organic removal solvents. Figure 3.2 shows a schematic of the photolithography steps with a positive photoresist.

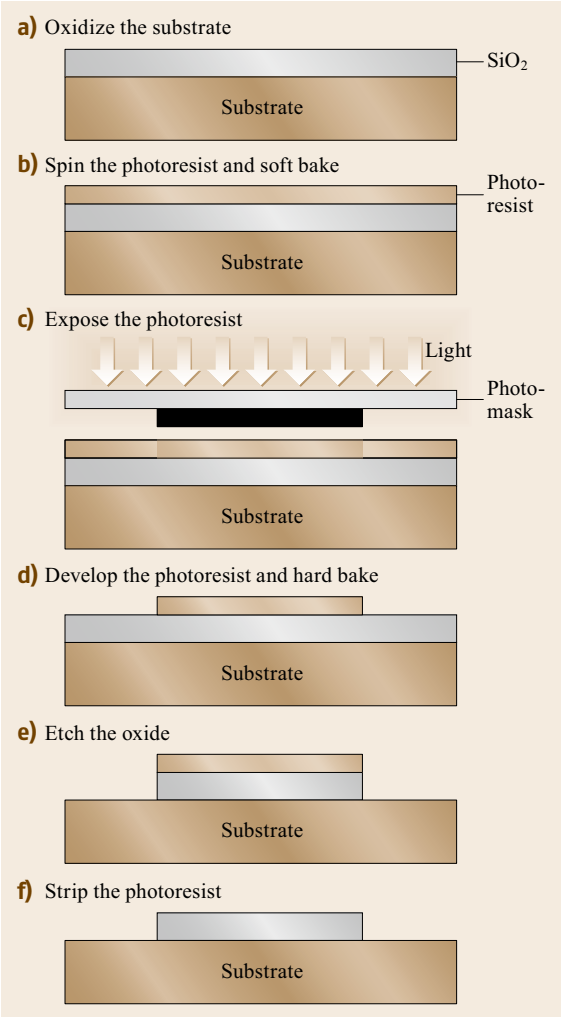


Fig. 3.2a–f Example of patterning. Schematic drawing of the photolithographic steps with a positive PR and pattern transfer. **(a)** Oxidation of silicon substrate. **(b)** Photoresist deposition and soft bake. **(c)** Photoresist exposure through mask. **(d)** Photoresist development and hard bake. **(e)** Silicon oxide etching. **(f)** Photoresist stripping

3.1.2 Thin-Film Deposition and Doping

Thin-film deposition and doping are extensively used in micro/nanofabrication technologies. Most fabricated micro/nanostructures contain materials other than that of the substrate, which are obtained by various deposition techniques or by modification of the substrate. Following is a list of a few typical applications for the deposited and/or doped materials used in micro/nanofabrication, which gives an idea of the required properties:

- Mechanical structure
- Electrical isolation
- Electrical connection
- Sensing or actuating
- Mask for etching and doping
- Support or mold during deposition of other materials (sacrificial materials)
- Passivation.

Most of the deposited thin films have properties different from those of their corresponding *bulk* forms (for example, metals shows higher resistivities in thin-film form). In addition, the techniques utilized to deposit these materials have a great impact on their final properties. For instance, internal stress (compressive or tensile) in a film is strongly process dependent. Ex-

cessive stress may crack or detach the film from the substrate and therefore must be minimized, although it may also be useful for certain applications. Adhesion is another important issue that needs to be taken into account when depositing thin films. In some cases such as the deposition of noble metals (e.g., gold) an intermediate layer (chromium or titanium) may be needed to improve adhesion. Finally, step coverage and conformality are two properties that can also influence the choice of one or another deposition technique. Figure 3.3 illustrates these concepts.

Oxidation

Oxidation of silicon is a process used to obtain a thin film of SiO_2 with excellent quality (very low density of defects) and thickness homogeneity. Although it is not properly a deposition (often referred indistinctively as growth), the result is the same; i. e., a thin layer of a new material covering the surface is produced. The oxidation process is typically carried out at temperatures in the range of 900–1200 °C in the presence of O_2 (dry oxidation) or H_2O (wet oxidation). The reactions for oxide formation are

$$Si_{(solid)} + O_{2\,(gas)} \rightarrow SiO_{2\,(solid)}$$

and

$$Si_{(solid)} + 2H_2O_{(steam)} \rightarrow SiO_{2\,(solid)} + 2H_{2\,(gas)} \,.$$

Although the rate of oxide growth is higher for wet oxidation, this is achieved at the expense of lower oxide quality (density). Since silicon atoms from the substrate participate in the reaction, the substrate is consumed as the oxide grows ($\approx 44\%$ of the total thickness lies above the line of the original silicon surface). The oxidation of silicon also occurs at room temperature, however a layer of about 20 Å (namely, native oxide) is enough to passivate the surface and prevent further oxidation. To grow thicker oxides, wafers are introduced into an electric resistance furnace such as that represented in Fig. 3.4. Tens of wafers can be processed in a single batch in such equipment. By strictly controlling the timing, temperature, and gas flow entering the quartz tube the desired thickness can be achieved with high accuracy. Thicknesses ranging from a few tens of Angstroms to 2 µm can be obtained in reasonable times. Despite the good quality of the SiO_2 obtained by silicon oxidation (also called *thermal oxide*), the use of this process is often limited to the early stages of the fabrication, since some of the materials added during the formation of structures may not withstand the high temperatures. The contamination of the furnace, when the substrates have been previously in contact with certain etchants such as KOH or when materials such as metals have been deposited, also poses limitations in most cases.

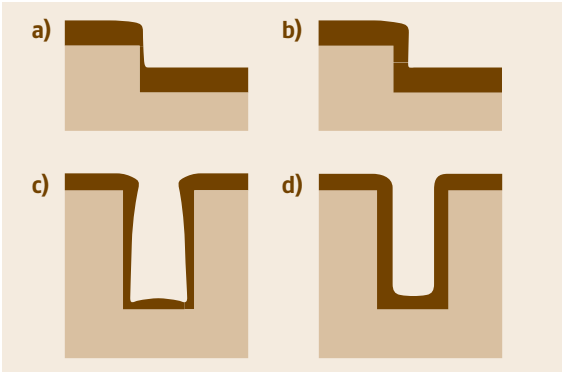


Fig. 3.3a–d Step coverage and conformality: **(a)** poor step coverage, **(b)** good step coverage, **(c)** nonconformal layer, and **(d)** conformal layer

Doping

The introduction of certain impurities in a semiconductor can change its electrical, chemical, and even mechanical properties. Typical impurities or *dopants* used in silicon include boron (to form p-type regions) and phosphorous or arsenic (to form n-type regions). Doping is the key process used in the microelectronic industry to fabricate major components such as diodes and transistors. In micro/nanofabrication technologies doping has additional applications such as the formation of piezoresistors for mechanical transducers or the creation of etch stop-layers. Two different techniques are used to introduce impurities into a semiconductor substrate: diffusion and ion implantation.

Diffusion is the process which became dominant in the initial years following the invention of the integrated circuit to form n- and p-type regions in the silicon. The diffusion of impurities into silicon occurs only at high temperature (above 800 °C). Furnaces used to carry out this process are similar to those used for oxidation. Dopants are introduced in the furnace gaseous atmosphere from liquid or solid sources. Figure 3.5 il-

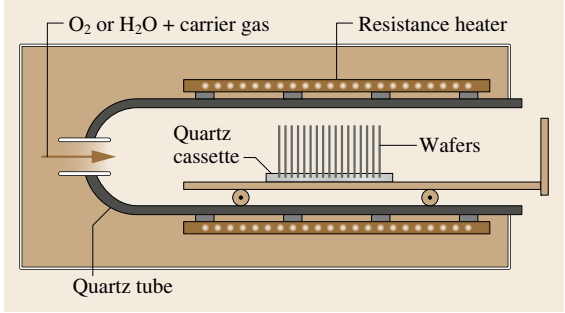


Fig. 3.4 Schematic representation of a typical oxidation furnace

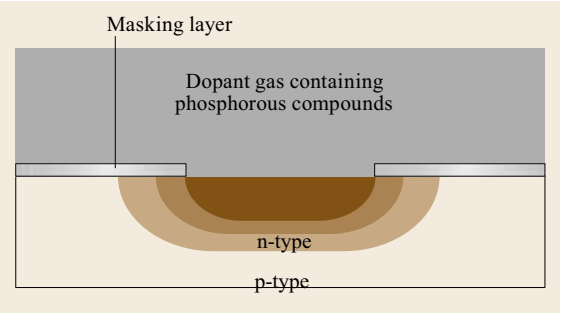


Fig. 3.5 Formation of an n-type region on a p-type silicon substrate by diffusion of phosphorous

lustrates the process of creating an n-type region by diffusion of phosphor from the surface into a p-type substrate. A masking material is previously deposited and patterned on the surface to define the areas to be doped. However, because diffusion is an isotropic process, the doped area will also extend underneath the mask. In microfabrication, diffusion is mainly used for the formation of very highly boron-doped regions (p^{++}), which are usually used as an etch stop in bulk micromachining.

Ion implantation allows more precise control of the dose (the total amount of impurities introduced per area unit) and the impurity profile (the concentration versus depth). In ion implantation the impurities are ionized and accelerated towards the semiconductor surface, at high energies. The penetration of impurities into the material follows a Gaussian distribution. After implantation, an annealing process is needed to activate the impurities and repair the damage in the crystal structure produced by ion collisions. A *drive-in* process to redistribute the impurities, done in a standard furnace such as those used for oxidation or diffusion, may be required as well.

Chemical Vapor Deposition and Epitaxy

As its name suggests, chemical vapor deposition (CVD) includes all deposition techniques using the reaction of chemicals in gas phase to form the deposited thin film. The energy needed for the chemical reaction to occur is usually supplied by maintaining the substrate at elevated temperature. Alternative energy sources such as plasma or optical excitation are also used, with the advantage of requiring a lower temperature at the substrate. The most common CVD processes in microfabrication are low-pressure CVD (LPCVD) and plasma-enhanced (PECVD).

The LPCVD process is typically carried out in electrically heated tubes, similar to oxidation tubes, equipped with pumping capabilities to achieve the low pressures required (0.1–1.0 Torr). Large numbers of wafers can be processed simultaneously where the material is deposited on both sides of the wafers. The process temperature depends on the material to be deposited, but generally is in the range 550–900 °C. As in oxidation, high temperatures and contamination issues can restrict the type of processes used previous to the LPCVD. Typical materials deposited by LPCVD include silicon oxide (e.g., $SiCl_2H_2 + 2N_2O \rightarrow SiO_2 + 2N_2 + 2HCl$ at 900 °C), silicon nitride (e.g., $3SiH_4 + 4NH_3 \rightarrow Si_3N_4 + 12H_2$ at 700–900 °C), and polysilicon (e.g., $SiH_4 \rightarrow Si + 2H_2$ at 600 °C). Due to its faster etch rate in HF, in situ phosphorous-doped LPCVD oxide (phosphosilicate glass, PSG) is extensively used in surface micromachining as the sacrificial layer. Conformality in this process is excellent, even for very high-aspect-ratio structures. Mechanical properties of LPCVD materials are good compared with others such as PECVD, and are often used as structural materials in microfabricated devices. Stress in the deposited layers depends on the material, deposition conditions, and subsequent thermal history (e.g., postdeposition annealing). Typical values are: 100–300 MPa (compressive) for oxide, ≈ 1 GPa (tensile) for stoichiometric nitride, and ≈ 200–300 MPa (tensile) for polysilicon. The stress in nitride layers can be reduced to nearly zero by using a silicon-rich composition. Since the stress values can vary over a wide range, one has to measure and characterize the internal stress of deposited thin films for any specific equipment and deposition conditions.

The PECVD process is performed in plasma systems such as that represented in Fig. 3.6. The use of radiofrequency (RF) energy to create highly reactive species in the plasma allows for the use of lower temperature at the substrate (150–350 °C). Parallel-plate plasma reactors normally used in microfabrication can

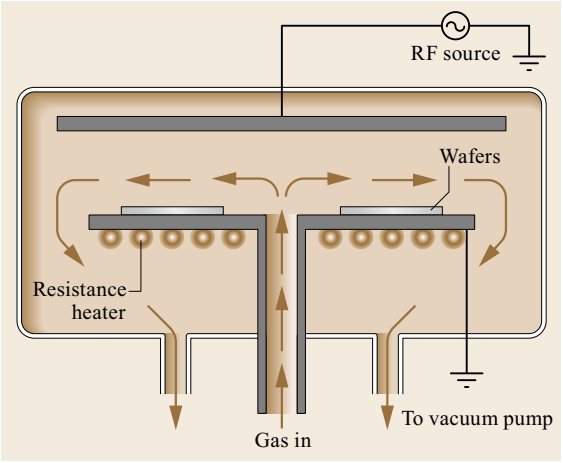


Fig. 3.6 Schematic representation of a typical PECVD system

only process a limited number of wafers per batch. The wafers are positioned horizontally on top of the lower electrode so only one side gets deposited. Typical materials deposited with PECVD include silicon oxide, nitride, and amorphous silicon. Conformality is good for low-aspect-ratio structures, but becomes very poor for deep trenches (20% of the surface thickness inside through-wafer holes with an aspect ratio of 10). Stress depends on deposition parameters and can be either compressive or tensile. PECVD nitrides are typically nonstoichiometric (Si_xN_y) and are much less resistant to etchants in masking applications.

Another interesting type of CVD is epitaxial growth. In this process, a single-crystalline material is grown as an extension of the crystal structure of the substrate. It is possible to grow dissimilar materials if the crystal structures are somehow similar (lattice matched; preferably mismatch below 8%). Silicon-on-sapphire (SoS) substrates and some heterostructures are fabricated in this way. However, most common in microfabrication is the growth of silicon on another silicon substrate. Of particular interest for the formation of microstructures is selective epitaxial growth. In this process the silicon crystal is allowed to grow only in windows patterned on a masking material. Many CVD techniques have been used to produce epitaxial growth. The most common for silicon is thermal chemical vapor deposition or vapor-phase epitaxy (VPE). Metalorganic chemical vapor deposition (MOCVD) and molecular-beam epitaxy (MBE) are the most common for growing high-quality III–V compound layers with nearly atomic abrupt interfaces. The former uses vapors of organic compounds with group III atoms such as trimethylgallium ($Ga(CH_3)_3$) and group V hydrides such as AsH_3 in a CVD chamber with fast gas switching capabilities. The latter typically uses molecular beams from thermally evaporated elemental sources aiming at the substrate in an ultrahigh-vacuum chamber. In this case, rapid on/off control of the beams is achieved by using shutters in front of the sources. Finally, it should be mentioned that many metals (molybdenum, tantalum, titanium, and tungsten) can also be deposited using LPCVD. These are attractive for their low resistivities and their ability to form silicides with silicon. Due to its application in new interconnect technologies, copper CVD is an active area of research.

Physical Vapor Deposition (Evaporation and Sputtering)

In physical deposition systems the material to be deposited is transported from a source to the wafers, both being in the same chamber. Two physical principles are used to do this: evaporation and sputtering.

In evaporation, the source is placed in a small container with tapered walls, called the crucible, and is heated up to a temperature where evaporation occurs. Various techniques are utilized to reach the high temperatures needed, including the induction of high currents with coils wound around the crucible and the bombardment of the material surface with an electron beam (e-beam evaporators). This process is mainly used to deposit metals, although dielectrics can also be evaporated. In a typical system the crucible is located at the bottom of a vacuum chamber whereas the wafers are placed lining the dome-shaped ceiling of the chamber (Fig. 3.7). The main characteristic of this process is very poor step coverage, including shadow effects as illustrated in Fig. 3.8. As will be explained in subsequent sections, some microfabrication techniques utilize these effects to pattern the deposited layer. One way to improve the step coverage is by rotating and/or heating the wafers during deposition.

In sputtering, a target of the material to be deposited is bombarded with high-energy inert ions (usually argon). The outcome of the bombardment is that

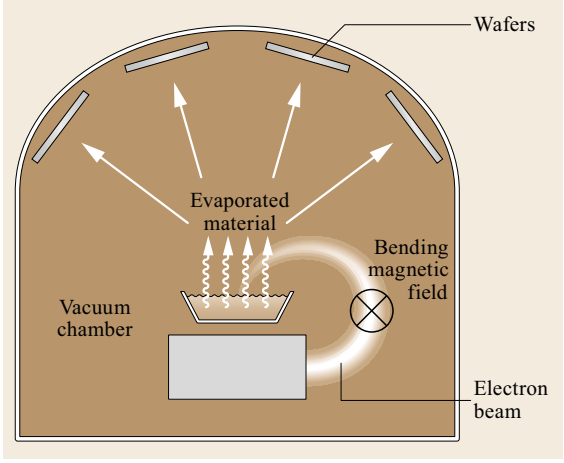


Fig. 3.7 Schematic representation of an e-beam deposition system

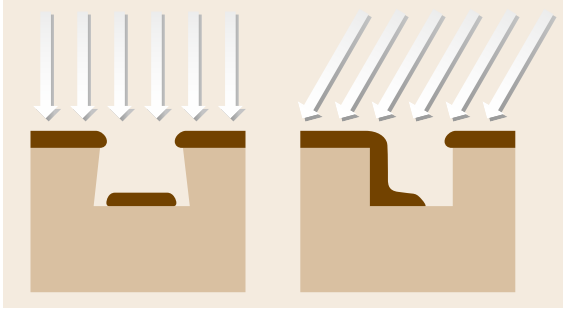


Fig. 3.8 Shadow effects observed in evaporated films. *Arrows* show the trajectory of the material atoms being deposited

individual atoms or clusters are removed from the surface and ejected towards the wafer. The physical nature and simplicity of this process allows its use with virtually any existing material. Examples of interesting materials for microfabrication that are frequently sputtered include metals, dielectrics, alloys (such as shape memory alloys), and all kinds of compounds (for example, piezoelectric lead zirconate titanate (PZT)). The inert ions bombarding the target are produced in direct-current (DC) or RF plasma. In a simple parallel-plate system the top electrode is the target and the wafers are placed horizontally on top of the bottom electrode. In spite of its lower deposition rate, step coverage in sputtering is much better than in evaporation. However, the films obtained with this deposition process are nonconformal. Figure 3.9 illustrates successive sputtering profiles in a trench.

Both evaporation and sputtering systems are often able to deposit more than one material simultaneously or sequentially. This capability is very useful to obtain alloys and multilayer films (e.g., multilayer magnetic recording heads are sputtered). For certain low-reactivity metals such as Au and Pt the previous deposition of a thin layer of another metal is needed to improve adhesion. Ti and Cr are two frequently used adhesion promoters. Stress in evaporated or sputtered layers is typically tensile. The deposition rates are much higher than for most CVD techniques. However, due to stress accumulation and cracking, thickness beyond 2 μm is rarely deposited with these processes. For thicker deposition a technique described in the next section is sometimes used.

Electroplating

Electroplating (or electrodeposition) is a process typically used to obtain thick (tens of μm) metal structures. The sample to be electroplated is introduced into a solution containing a reducible form of the ion of the desired metal and is maintained at a negative potential (cathode) relative to a counterelectrode (anode). The ions are reduced at the sample surface and the insoluble metal atoms are incorporated into the surface. As an example, copper electrodeposition is frequently done in copper sulfide-based solutions. The reaction taking place at the surface is $Cu^{2+} + 2e^- \rightarrow Cu_{(s)}$. Recommended current densities for electrodeposition processes are on the order of $5-100\,mA/cm^2$.

As can be deduced from the process mechanism, the surface to be electroplated has to be electrically conductive, and preferably of the same material as the deposited one if good adhesion is desired. In order to electrodeposit metals on top of an insulator (the most frequent case) a thin film of the same metal, called the seed layer, is previously deposited on the surface. Masking of the seed layer with a resist permits selective electroplating of the patterned areas. Figure 3.10 illustrates a typical sequence of the steps required to obtain isolated metal structures.

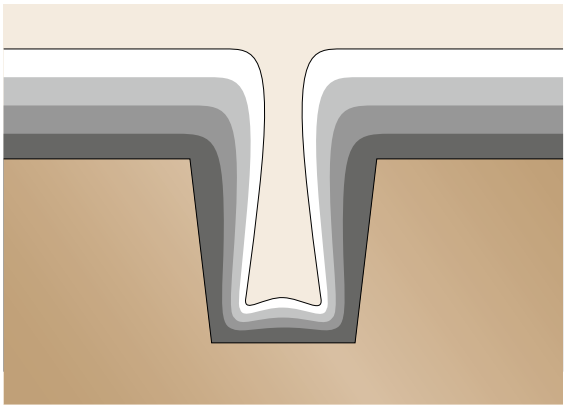


Fig. 3.9 Typical cross section evolution of a trench while being filled with sputter deposition

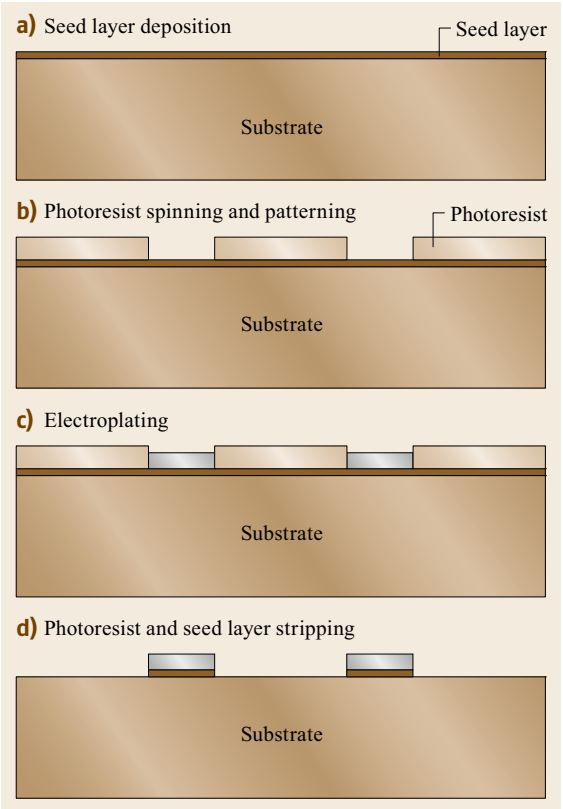


Fig. 3.10a–d Formation of isolated metal structures by electroplating through a mask: **(a)** seed layer deposition, **(b)** photoresist spinning and patterning, **(c)** electroplating, and **(d)** photoresist and seed layer stripping

Pulsed Laser Deposition and Atomic Layer Deposition

Pulsed laser and atomic layer deposition techniques have attracted a considerable amount of attention recently. These two techniques offer several unique advantages compared with other thin-film deposition methods that are particularly useful for next-generation nanoscale device fabrication. Pulsed laser deposition (PLD) is a simple technique that uses an intense (1 GW within 25 ns) UV laser (e.g., a KrF excimer) to ablate a target material [3.6]. Plasma is subsequently formed from the target and is deposited on the substrate. Multitarget systems with Auger and reflection high-energy electron diffraction (RHEED) spectroscopes are commercially available. Figure 3.11 shows a typical PLD deposition setup. The main advantages of the PLD are its simplicity and ability to deposit complex materials with preserved stoichiometry (so-called stoichiometry transfer). In addition, fine control over film thickness is also possible by controlling the number of pulses. The stoichiometry-transfer property of the PLD allows many complex targets such as ferroelectrics, superconductors, and magnetostrictives to be deposited. Other deposited materials include oxides, carbides, polymers, and metallic systems (e.g., FeNdB).

Atomic layer deposition (ALD) is a gas-phase self-limiting deposition method capable of depositing atomic layer thin films with excellent large-area uniformity and conformality [3.7]. It enables simple and accurate control over film composition and thickness at the atomic layer level (typical growth rates of a few Å/cycle). Although most of the attention recently has been directed towards depositing high-k dielectric materials (Al_2O_3, and HfO_2) for next-generation complementary metal–oxide–semiconductor (CMOS) electronics, other materials can also be deposited. These include transition metals (Cu, Co, Fe, and Ni), metal oxides, sulfides, nitrides, and fluorides. Atomic-level control over film thickness and composition are also attractive features for applications in MEMS such as conformal three-dimensional (3-D) packaging and air-gap structures. ALD is a modification of the CVD process and is based on two or more vapor-phase reactants that are introduced into the deposition chamber in a sequential manner. One growth cycle consists of four steps. First, a precursor vapor is introduced into the chamber, resulting in the deposition of a self-limiting monolayer on the surface of the substrate. Then, the extra unreacted vapor is pumped out and a vapor dose of a second reactant is introduced. This reacts with the precursor on the surface in a self-limiting fashion. Finally, the extra unreacted vapor is pumped out and the cycle is repeated.

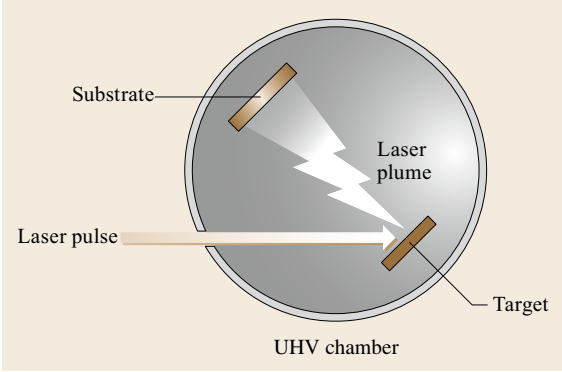


Fig. 3.11 A typical PLD deposition setup

3.1.3 Etching and Substrate Removal

Thin-film and bulk substrate etching is another fabrication step that is of fundamental importance to both VLSI processes and micro/nanofabrication. In the VLSI area, various conducting and dielectric thin films deposited for passivation or masking purposes need to be removed at some point or another. In micro/nanofabrication, in addition to thin-film etching, very often the substrate (silicon, glass, GaAs, etc.) also needs to be removed in order to create various mechanical micro/nanostructures (beams, plates, etc.). Two important figures of merit for any etching process are selectivity and directionality. Selectivity is the degree to which the etchant can differentiate between the masking layer and the layer to be etched. Directionality has to do with the etch profile under the mask. In an isotropic etch, the etchant attacks the material in all directions at the same rate, hence creating a semicircular profile under the mask (Fig. 3.12a). In contrast, in an anisotropic etch,

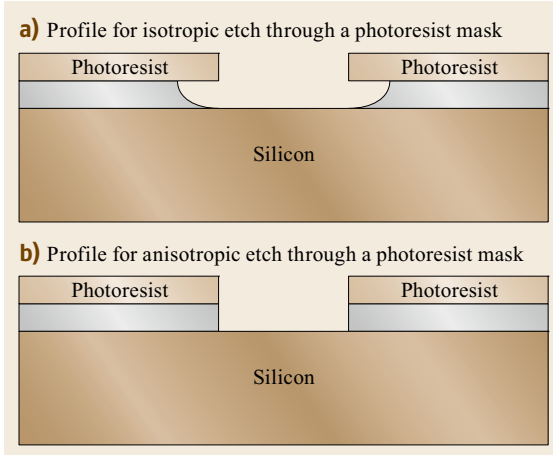


Fig. 3.12a,b Profile for (a) isotropic and (b) anisotropic etching through a photoresist mask

the dissolution rate depends on specific directions and one can obtain straight side-walls or other noncircular profiles (Fig. 3.12b). One can also divide the various etching techniques into wet and dry categories. In this chapter, we will use this classification and discuss different wet etchants first followed by dry etching techniques used most often in the micro/nanofabrication.

Wet Etching

Historically, wet etching techniques preceded the dry ones. These still constitute an important group of etchants for micro/nanofabrication in spite of their less frequent application in the VLSI processes. Wet etchants are by and large isotropic and show superior selectivity to the masking layer as compared with various dry techniques. In addition, due to the lateral undercut, the minimum feature size achievable with wet etchants is limited to $> 3\,\mu m$. Silicon dioxide is commonly etched in dilute (6 : 1, 10 : 1, or 20 : 1 by volume) or buffered HF (BHF: $HF + NH_4F$) solutions (etch rate of $\approx 1000\,Å/min$ in BHF). Photoresist and silicon nitride are the two most common masking materials for the wet oxide etch. The wet etchant for silicon nitride is hot (140–200 °C) phosphoric acid with silicon oxide as the masking material. Nitride wet etch is not very common (except for blanket etch) due to the masking difficulty and nonrepeatable etch rates. Metals can be etched using various combinations of acid and base solutions. There are also many commercially available etchant formulations for aluminum, chromium, and gold which can easily be used. A comprehensive table of various metal etchants can be found in [3.8].

Anisotropic and isotropic wet etching of crystalline (silicon and gallium arsenide) and noncrystalline (glass) substrates is an important topic in micro/nanofabrication [3.9–12]. In particular, the realization of the possibility of anisotropic wet etching of silicon is considered to mark the beginning of the micromachining and MEMS discipline. Isotropic etching of silicon using $HF/HNO_3/CH_3COOH$ (various different formulations have been used) dates back to the 1950s and is still frequently used to thin down the silicon wafer. The etch mechanism for this combination has been elucidated and is as follows: HNO_3 is used to oxidize the silicon, which is subsequently dissolved away in the HF. The acetic acid is used to prevent the dissociation of HNO_3 (the etch works as well without the acetic acid). For short etch times, silicon dioxide can be used as the masking material; however, one needs to use silicon nitride if a longer etch time is desired. This etch also shows dopant selectivity, with the etch rate dropping at lower doping concentrations ($< 10^{17}\,cm^{-3}$ n- or p-type). Although this effect can potentially be used as an etch-stop mechanism in order to fabricate microstructures, the difficulty in masking has prevented widespread application of this approach. Glass can also be isotropically etched using the HF/HNO_3 combination with the etch surfaces showing considerable roughness. This has been extensively used in fabricating microfluidic components (mainly channels). Although Cr/Au is usually used as the masking layer, long etch times require a more robust mask (bonded silicon has been used for this purpose).

Silicon anisotropic wet etch constitutes an important technique in bulk micromachining. The three most important silicon etchants in this category are potassium hydroxide (KOH), ethylenediamine pyrocatechol (EDP), and tetramethyl ammonium hydroxide (TMAH). These are all anisotropic etchants which attack silicon along preferred crystallographic directions. In addition, they all show marked reduction of the etch rate in heavily ($> 5 \times 10^{19}\,cm^{-3}$) boron-doped ($p^{++}$) regions. The chemistry behind the action of these etchants is not yet very clear but it seems that silicon atom oxidation at the surface and its reaction with hydroxyl ions (OH^-) is responsible for the formation of a soluble silicon complex ($SiO_2(OH)^{2-}$). The etch rate depends on the concentration and temperature and is usually around $1\,\mu m/min$ at temperatures of 85–115 °C. Common masking materials for anisotropic wet etchants are silicon dioxide and nitride, with the latter being superior for longer etch times. The crystallographic plane which shows the slowest etch rate is the (111) plane. Although the lower atomic concentration along these planes has been speculated to be the reason for this phenomena, the evidence is inconclusive and other factors must be included to account for this remarkable etch-stop property. The anisotropic behavior of these etchants with respect to the (111) planes has been extensively used to create beams, membranes, and other mechanical and structural components. Figure 3.13 shows the typical cross sections of (100) and (110) silicon wafers etched with an anisotropic wet etchant. As can be seen, the (111) slow planes are exposed in both situations, one creating 54.7° sloped side-walls in the (100) wafer and the other creating vertical side-walls in the (110) wafer. Depending on the dimensions of the mask opening, a V-groove or a trapezoidal trench is formed in the (100) wafer. A large enough opening will allow the silicon to be etched all the way through the wafer, thus creating a thin dielectric membrane on the other side. It should be mentioned that exposed convex corners have a higher etch rate than concave ones, resulting in an undercut which can be used to create dielectric (e.g., nitride) cantilever beams. Figure 3.14 shows a cantilever beam fabricated using the convex corner undercut on a (100) wafer.

The three above-mentioned etchants show different directional and dopant selectivities. KOH has the best (111) selectivity (400/1), followed by TMAH and EDP. However, EDP has the highest selectivity with respect to deep boron diffusion regions. Safety and CMOS compatibility are other important criteria for choosing a particular anisotropic etchant. Among the three mentioned etchants TMAH is the most benign, whereas EDP is extremely corrosive and carcinogenic. Silicon can be dissolved in TMAH in order to improve its selectivity with respect to aluminum. This property has made TMAH very appealing for post-CMOS micromachining where aluminum lines have to be protected. Finally, it should be mentioned that one can modulate the etch rate using a reversed-biased p–n junction (electrochemical etch stop). Figure 3.15 shows the setup commonly used to perform electrochemical etching. The silicon wafer under etch consists of an n-epi region on a p-type substrate. Upon the application of a reverse-bias voltage to the structure (p-substrate is in contact with the solution and n-epi is protected using a watertight fixture), the p-substrate is etched away. When the n-epi regions are exposed to the solution an oxide passivation layer is formed and etching is stopped. This technique can be used to fabricate single-crystalline silicon membranes for pressure sensors and other mechanical transducers.

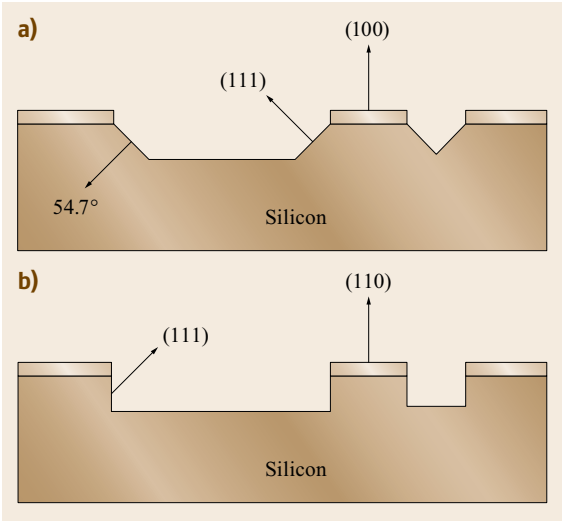


Fig. 3.13a,b Anisotropic etch profiles for: **(a)** (100) and **(b)** (110) silicon wafers

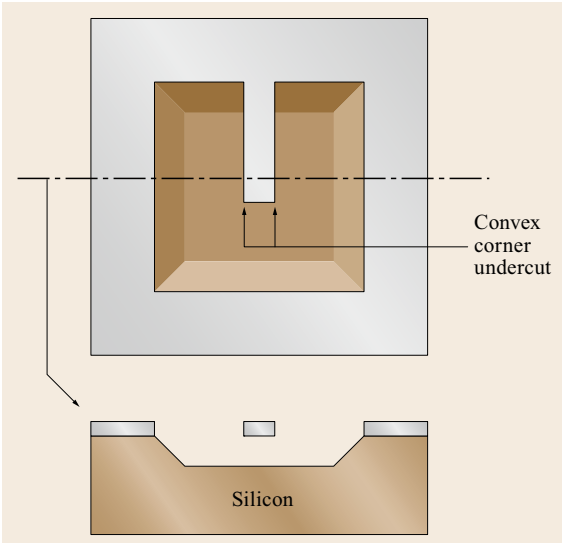


Fig. 3.14 Top view and cross section of a dielectric cantilever beam fabricated using convex corner undercut

Dry Etching

Most of the dry etching techniques are plasma based. They have several advantages when compared with wet etching. These include smaller undercut (allowing smaller lines to be patterned) and higher anisotropy (allowing high-aspect-ratio vertical structures), especially, as compared to some wet etching instances. However, the selectivity of dry etching techniques is lower than that of wet etchants, and one must take into account the finite etch rate of the masking materials. The three basic dry etching techniques, namely high-pressure plasma etching, reactive-ion etching (RIE), and ion milling, utilize different mechanisms to obtain directionality.

Ion milling is a purely physical process which utilizes accelerated inert ions (e.g., Ar^{+}) striking perpendicular to the surface to remove the material ($p \approx 10^{-4}$–10^{-3} Torr) (Fig. 3.16a). The main characteristics of this technique are very low etch rates (in the order of a few nm/min) and poor selectivity (close to 1 : 1 for most materials); hence it is generally used to etch very thin layers. In high-pressure (0.1–5 Torr) plasma etchers highly reactive species are created that react with

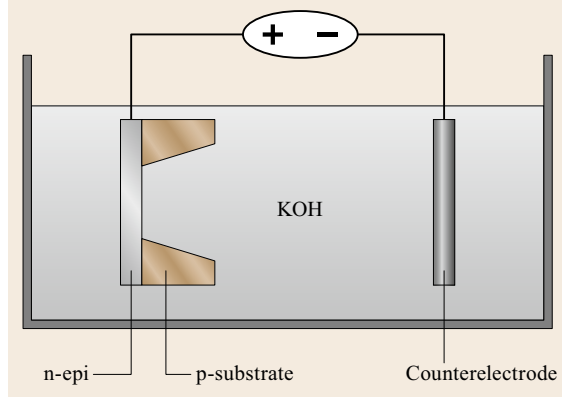


Fig. 3.15 Electrochemical etch setup

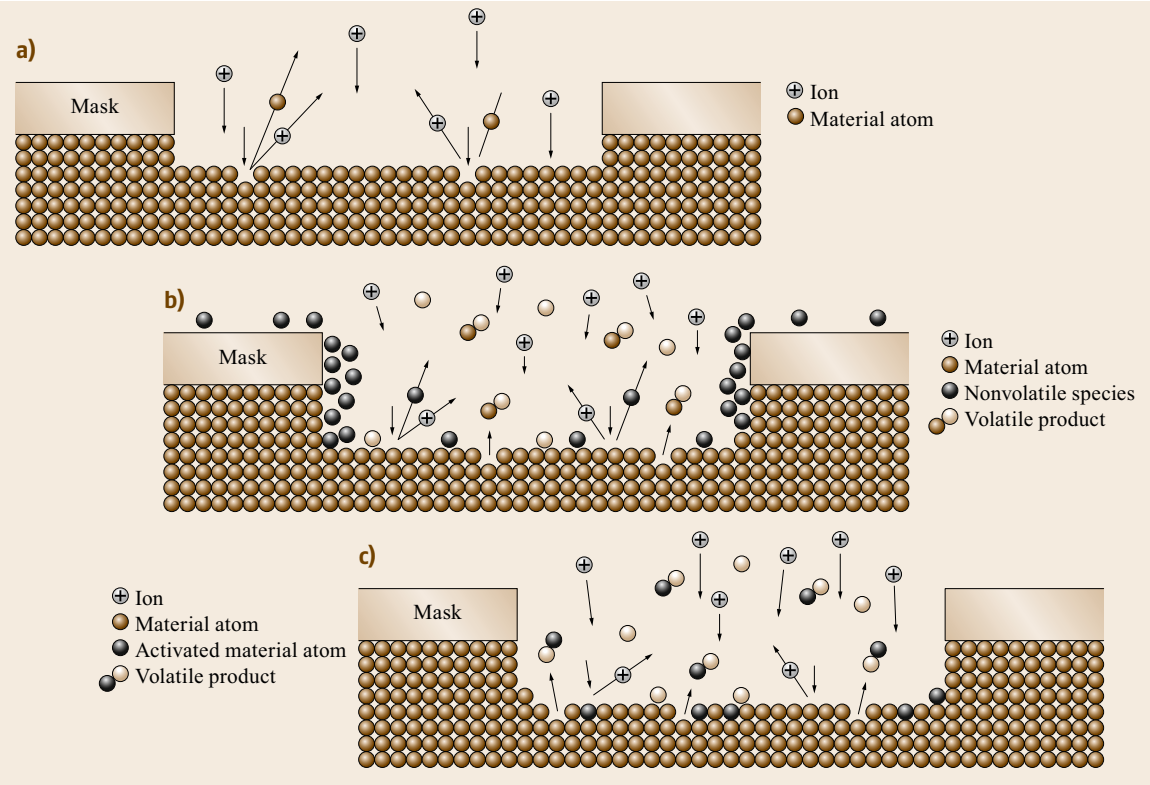


Fig. 3.16a–c Simplified representation of etching mechanisms for: **(a)** ion milling, **(b)** high-pressure plasma etching, and **(c)** RIE

the material to be etched. The products of the reaction are volatile so that they diffuse away and new material is exposed to the reactive species. Directionality can be achieved, if desired, with the side-wall passivation technique (Fig. 3.16b). In this technique nonvolatile species produced in the chamber deposit on and passivate the surfaces. The deposit can only be removed by physical collision with incident ions. Because the movement of the ions has a vertical directionality the deposit is removed mainly on horizontal surfaces, while vertical walls remain passivated. In this fashion, the vertical etch rate becomes much higher than the lateral one.

RIE etching, also called ion-assisted etching, is a combination of physical and chemical processes. In this technique the reactive species react with the material only when the surfaces are *activated* by the collision of incident ions from the plasma (e.g., by breaking bonds at the surface). As in the previous technique, the directionality of the ion's velocity produces much more collisions on the horizontal surfaces than on the walls, thus generating faster etching rates in the vertical direction (Fig. 3.16c). To increase the etch anisotropy further, in some cases side-wall passivation methods are also used. An interesting case is the deep reactive-ion etching (DRIE) technique, capable of achieving aspect ratios of 30 : 1 and silicon etching rates of 2–3 μm/min (through wafer etch is possible). In this technique, the passivation deposition and etching steps are performed sequentially in a two-step cycle, as shown in Fig. 3.17. In commercial silicon DRIE etchers SF_6/Ar is typically used for the etching step and a combination of Ar and a fluoropolymer (nCF_2) for the passivation step. A Teflon-like polymer about 50 nm thick is deposited during the latter step, covering only the side-walls (Ar^+ ion bombardment removes the Teflon on the horizontal surfaces). Due to the cyclic nature of this process, the side-walls of the etched features show a periodic *wave-shaped* roughness in the range of 50–400 nm. More recently, *Aimi* et al., reported on a similar method for deep etching of titanium. In this case titanium oxide was used as a side-wall passivation layer [3.13].

Dry etching can also be performed in nonplasma equipment if the etching gases are reactive enough. The so-called vapor-phase etching (VPE) processes can be carried out in a simple chamber with gas feeding and pumping capabilities. Two examples of VPE are xenon difluoride (XeF_2) etching of silicon and HF vapor

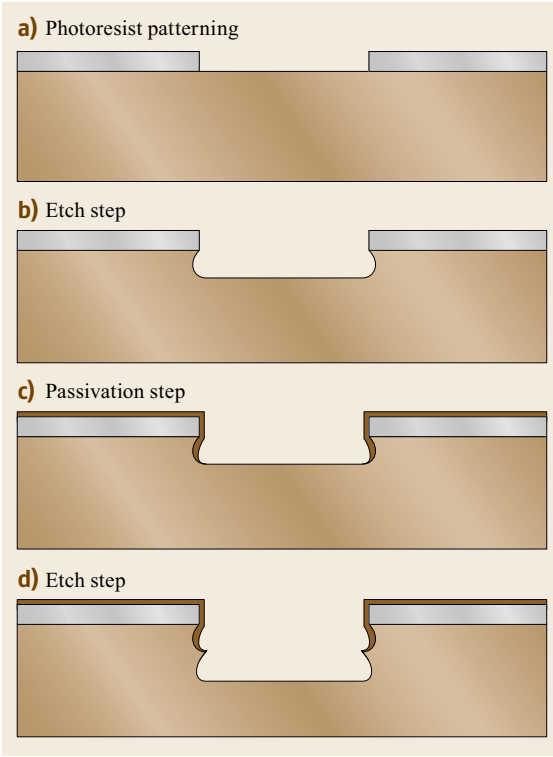


Fig. 3.17a–d DRIE cyclic process: **(a)** photoresist patterning, **(b)** etch step, **(c)** passivation step, and **(d)** etch step

etching of silicon dioxide. Due to its isotropic nature, these processes are typically used for etching sacrificial layers and releasing structures while avoiding stiction problems (Sects. 3.2.1 and 3.2.2).

Most important materials can be etched with the aforementioned techniques, and for each material a variety of chemistries are available. Table 3.1 lists some of the most common materials along with selected etch recipes [3.14]. For each chemistry the etch rate, directionality, and selectivity with respect to the mask materials depend on parameters such as the flow rates of the gases entering the chamber, the working pressure, and the RF power applied to the plasma.

Table 3.1 Typical dry etch chemistries

Si	CF_4/O_2, CF_2Cl_2, CF_3Cl, $SF_6/O_2/Cl_2$, $Cl_2/H_2/C_2F_6/CCl_4$, C_2ClF_5/O_2, Br_2, SiF_4/O_2, NF_3, ClF_3, CCl_4, $C_3Cl_3F_5$, C_2ClF_5/SF_6, C_2F_6/CF_3Cl, CF_3Cl/Br_2
SiO_2	CF_4/H_2, C_2F_6, C_3F_8, CHF_3/O_2
Si_3N_4	$CF_4/O_2/H_2$, C_2F_6, C_3F_8, CHF_3
Organics	O_2, CF_4/O_2, SF_6/O_2
Al	BCl_3, BCl_3/Cl_2, $CCl_4/Cl_2/BCl_3$, $SiCl_4/Cl_2$
Silicides	CF_4/O_2, NF_3, SF_6/Cl_2, CF_4/Cl_2
Refractories	CF_4/O_2, NF_3/H_2, SF_6/O_2
GaAs	BCl_3/Ar, $Cl_2/O_2/H_2$, $CCl_2F_2/O_2/Ar/He$, H_2, CH_4/H_2, $CClH_3/H_2$
InP	CH_4/H_2, C_2H_6/H_2, Cl_2/Ar
Au	$C_2Cl_2F_4$, Cl_2, $CClF_3$

3.1.4 Substrate Bonding

Substrate (wafer) bonding (silicon–silicon, silicon–glass, and glass–glass) is among the most important fabrication techniques in microsystem technology [3.15, 16]. It is frequently used to fabricate complex 3-D structures both as a functional unit and as a part of the final microsystem package and encapsulation. The two most important bonding techniques are silicon–silicon fusion (or silicon direct bonding) and silicon–glass electrostatic (or anodic) bonding. In addition to these techniques, several other alternative methods which utilize an intermediate layer (eutectic, adhesive, and glass frit) have also been investigated. All these techniques can be used to bond the substrates at the wafer level. In the present Handbook edition, as an explanation of each of the mentioned main bonding techniques (Si direct bonding, anodic bonding and bonding with intermediate layers) cannot be included, we refer to the following relevant literature [3.16–19].

3.2 MEMS Fabrication Techniques

In this section, we will discuss various important MEMS fabrication techniques commonly used to build various microdevices (microsensors and microactuators) [3.9–12]. The dimensional spectrum of the microstructures that can be fabricated using these techniques spans from 1 mm to 1 μm. As mentioned in the introduction, we will mostly emphasize the more important techniques and will not discuss specialized methods. Nonetheless, currently there is a trend in manufacturing from MEMS towards nanoelectromechanical systems (NEMS), as they provide superior performance, e.g., operation at higher frequencies). Downscaling can be addressed by to distinctive approaches, as it will be briefly referred in the next section (Sect. 3.3). Following the methods described for MEMS is an evolutionary technology, where top-down fabrication of the mechanical element down to the nanoscale is achieved with further reduction of the feature

size, i. e., based on advanced lithography/patterning, thin film deposition or etching methods. Alternatively, a revolutionary route (disruptive in some cases) replaces the (monolithic) mechanical element by new entities, such as novel nanomaterials (e.g., graphene) of nanostructured objects (e.g., CVD Si nanowires (NWs)), which can be integrated, typically, by adaptations in the fabrication sequence of the same processing of conventional planar-technology. Detailed explanations on MEMS/NEMS concepts and applications can be found in Part C and D of this Handbook.

3.2.1 Bulk Micromachining

Bulk micromachining is the oldest MEMS technology and hence probably one of the more mature ones [3.20]. It is currently by far the most commercially successful one, helping to manufacture devices such as pressure sensors and inkjet printheads. Although there are many different variations, the basic concept behind bulk micromachining is selective removal of the substrate (silicon, glass, GaAs, etc.). This allows the creation of various micromechanical components such as beams, plates, and membranes which can be used to fabricate a variety of sensors and actuators. The most important microfabrication techniques used in bulk micromachining are wet and dry etching and substrate bonding. Although one can use various criteria to categorize bulk micromachining techniques, we will use a historical timeline for this purpose. Starting with the more traditional wet etching techniques, we will proceed to discuss the more recent ones using deep RIE and wafer bonding.

Bulk Micromachining Using Wet Etch and Wafer Bonding

The use of anisotropic wet etchants to remove silicon can be marked as the beginning of the micromachining era. Back-side etch was used to create movable structures such as beams, membranes, and plates (Fig. 3.18). Initially, etching was timed in order to create a specified thickness. However, this technique proved to be inadequate for the creation of (relatively) thin structures ($< 20\,\mu m$). Subsequent use of various etch-stop techniques allowed the creation of thinner membranes in a more controlled fashion. As was mentioned in *High-Aspect Ratio Micromachining*, heavily boron-doped regions and electrochemical bias can be used to slow down the etch process drastically and hence create controllable thickness microstructures. Figure 3.19a,b shows the cross section of two piezoresistive pressure sensors fabricated using electrochemical and p^{++} etch-stop techniques. The use of the p^{++} method requires epitaxial growth of a lightly doped region on top of a p^{++} etch-stop layer. This layer is subsequently used for the placement of piezoresistors. However, if no active component is required one can simply use the p^{++} region to create a thin membrane (Fig. 3.19c).

The p^{++} etch-stop technique can also be used to create isolated thin silicon structures through the dissolution of the entire lightly doped region [3.21]. This technique was successfully used to fabricate silicon recording and stimulating electrodes for biomedical applications. Figure 3.20 shows the cross section of

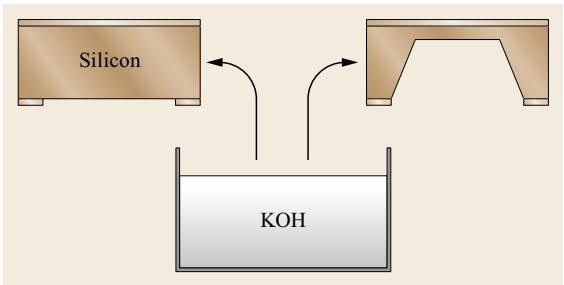


Fig. 3.18 Wet anisotropic silicon back-side etch

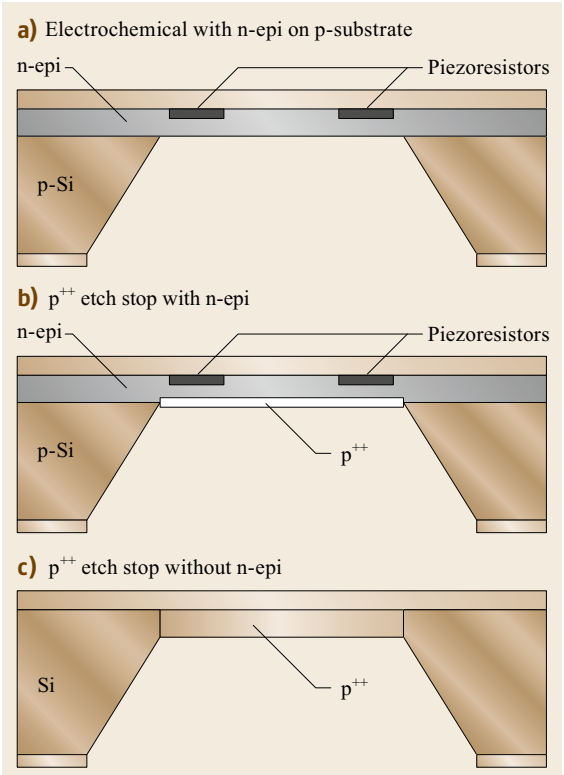


Fig. 3.19a–c Wet micromachining etch-stop techniques: **(a)** electrochemical with n-epi on p-substrate, **(b)** p^{++} etch stop with n-epi, and **(c)** p^{++} etch stop without n-epi

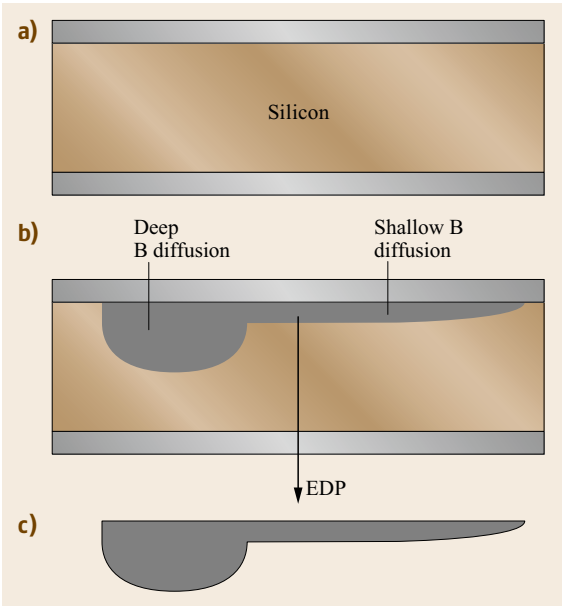


Fig. 3.20a–c Free-standing microstructure fabrication using deep and shallow boron diffusion and EDP release **(a)** silicon wafer, **(b)** deep and shallow boron diffusion, and **(c)** EDP etch

such a process which relies on deep (15–20 μm) and shallow boron (2–5 μm) diffusion steps to create microelectrodes with flexible connecting ribbon cables. An extension of this process which uses a combination of p^{++} etch-stop layers and silicon–glass anodic bonding has also been developed. This process is commonly known as the dissolved wafer process and has been used to fabricate a variety of microsensors and microactuators [3.22]. Figure 3.21 shows the cross section of this process. Figure 3.22 shows a scanning electron microscopy (SEM) photograph of a microaccelerometer fabricated using the dissolved wafer process.

Importantly, it is also possible to merge wet bulk micromachining and microelectronics fabrication processes to build micromechanical components on the same substrate as the integrated circuits (CMOS, bipolar, or bipolar complementary metal oxide semiconductor (BiCMOS)) [3.23]. This is very appealing since it allows the integration of interface and signal-processing circuitry with MEMS structures on a single chip. However, important fabrication issues such as process compatibility and yield have to be carefully considered. Among the most popular techniques in this category is postprocessing of CMOS integrated circuits by front-side etching in TMAH solutions. As was mentioned previously, silicon-rich TMAH does not attack aluminum and therefore can be used to undercut

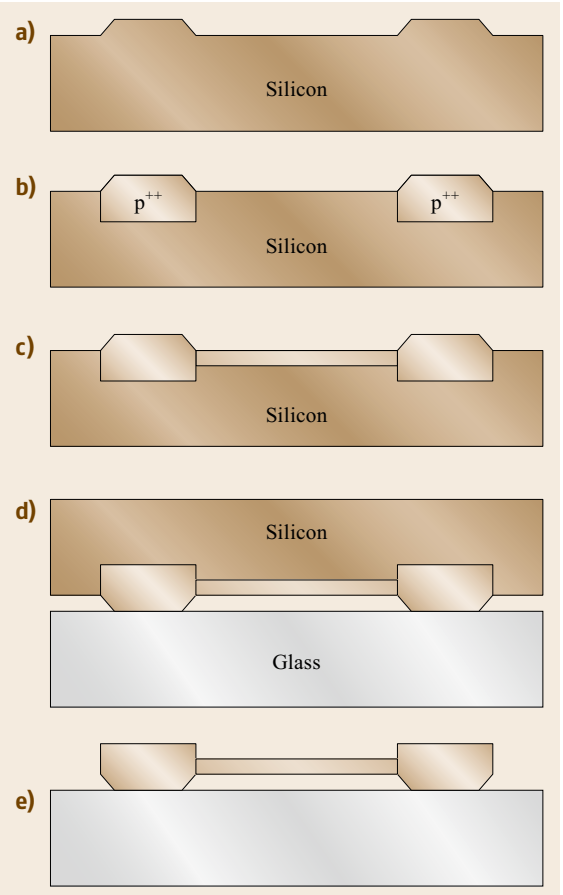


Fig. 3.21a–e Dissolved wafer process sequence: **(a)** KOH etch, **(b)** deep B diffusion, **(c)** shallow B diffusion, **(d)** silicon–glass anodic bond, and **(e)** release in EDP

microstructures in an already processed CMOS chip. Figure 3.23 shows a schematic of such a process in which a front-side wet etch and electrochemical etch stop are used to produce suspended beams. This technique has been extensively used to fabricate a variety of microsensors (e.g., humidity, gas, chemical, and pressure). Figure 3.24 shows a photograph of a post-CMOS-processed chemical sensor.

Bulk Micromachining Using Dry Etch

Bulk silicon micromachining using dry etching is a very attractive alternative to the wet techniques described in the previous section. These techniques were developed during the mid 1990s subsequent to successful efforts geared towards the development of processes for anisotropic dry silicon etch. More recent advances in deep silicon RIE and the availability of SOI wafers with a thick top silicon layer have increased the applicability of these techniques. These techniques allow the fabrica-

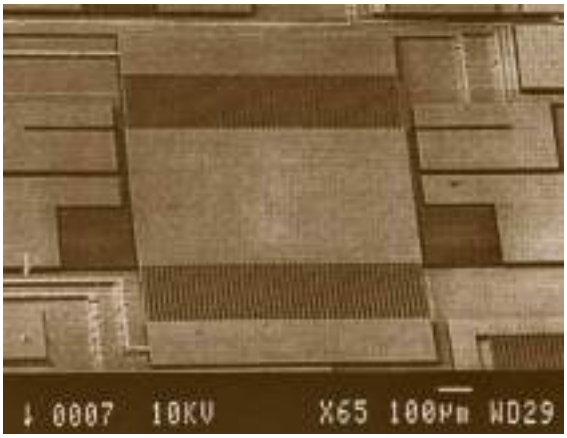


Fig. 3.22 SEM image of a microaccelerometer fabricated using the dissolved wafer process (after [3.22])

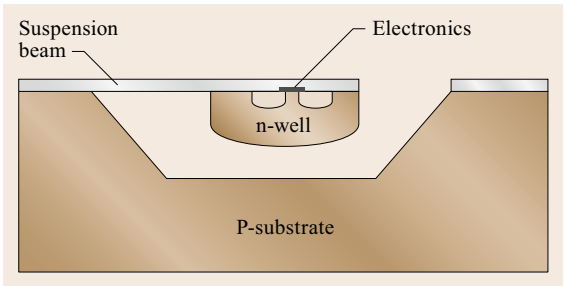


Fig. 3.23 Suspended island created on a prefabricated CMOS chip using front-side wet etch and electrochemical etch stop

tion of high-aspect-ratio vertical structures in isolation or along with on-chip electronics. Process compatibility with active microelectronics is less of a concern in dry methods since many of them do not damage the circuit or its interconnect.

The simplest dry bulk micromachining technique relies on front-side undercut of microstructures using a XeF_2 vapor-phase etch [3.24]. As was mentioned before, this, however, is an isotropic etch and therefore has limited applications. A combination of isotropic/anisotropic dry etch is more useful and can be used to create a variety of interesting structures. Two successful techniques using this combination are single-crystal reactive etching and metallization (SCREAM) [3.25] and post-CMOS dry release using aluminum/silicon dioxide laminate [3.26]. The first technique relies on the combination of isotropic/anisotropic dry etch to create single-crystalline suspended structures. Figure 3.25 shows the cross section of this process. If electrostatic actuation is desired, a metal can be sputtered to cover the top and side-wall of the microstructure and bottom of the cavity formed below it (Fig. 3.25f). Figure 3.26

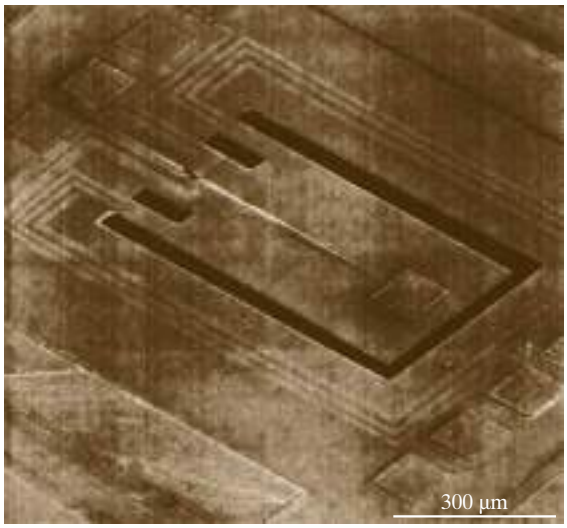


Fig. 3.24 Photograph of a post-CMOS-processed cantilever beam resonator for chemical sensing (after [3.23])

shows an SEM photograph of a comb-drive actuator fabricated using SCREAM technology.

The second dry release technique relies on the masking capability of aluminum interconnect lines in a CMOS integrated circuit to create suspended microstructures. Figure 3.27 shows a cross section of this process. As can be seen the third level Al of a prefabricated CMOS chip is used as a mask to etch the underlying oxide layers anisotropically all the way to the silicon (CHF_3/O_2) (Fig. 3.27b). This is followed by an anisotropic silicon etch to create a recess in the silicon, which will be used in the final step to facilitate the undercut and release (Fig. 3.27c). Finally, an isotropic silicon etch is used to undercut and release the structures (Fig. 3.27d). Figure 3.28 shows an SEM photograph of a comb-drive actuator fabricated using this technology.

In addition to the methods described above, recent advancements in the development of deep reactive-ion etching of silicon (DRIE) have created new opportunities for dry bulk micromachining techniques (Sect. 3.2.3). One of the most important ones uses thick silicon SOI wafers which are commercially available in various top silicon thicknesses [3.27]. Figure 3.29 shows a cross section of a typical process using DRIE and SOI wafers. The top silicon layer is patterned and etched all the way to the buried oxide (Fig. 3.29b). The oxide is subsequently removed in HF, hence releasing suspended single-crystalline microstructures (Fig. 3.29c). In a modification of this process, the substrate can also be removed from the back-side, allowing easy access from both sides (which allows easier release and prevents stiction) (Fig. 3.29d).

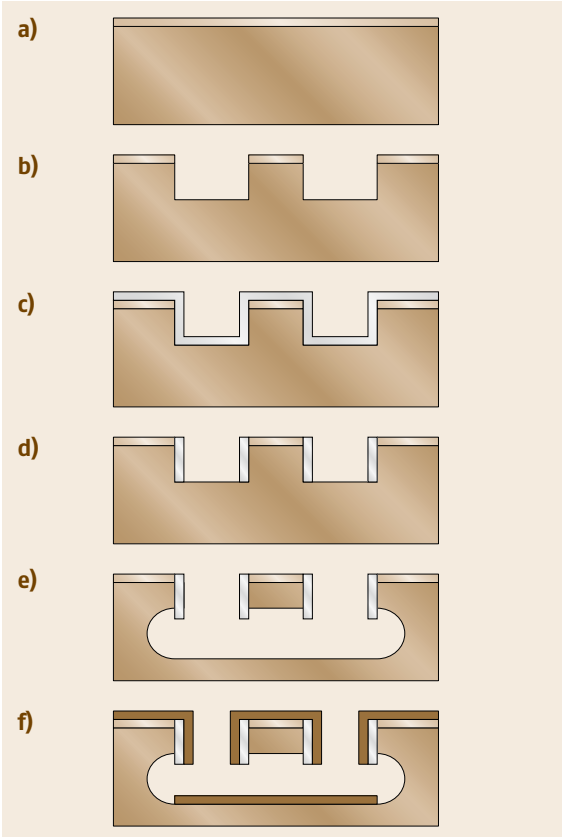


Fig. 3.25a–f Cross section of the SCREAM process **(a)** silicon wafer, **(b)** anisotropic silicon etch, **(c)** conformal passivation, **(d)** anisotropic etching of the passivation (hence protecting the sidewall), **(e)** isotropic silicon etch, and **(f)** metal deposition

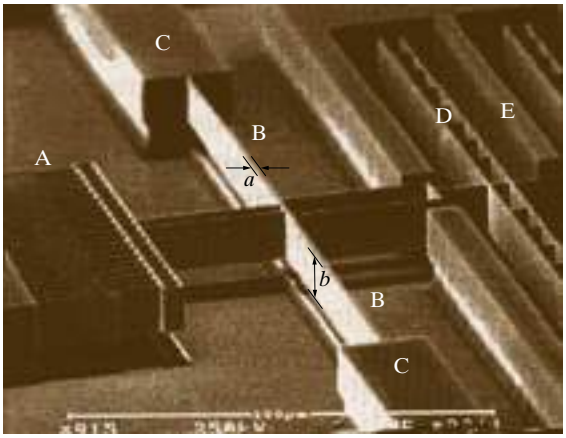


Fig. 3.26 SEM image of a structure fabricated using the SCREAM process: *A* comb-drive actuator, *B* suspended spring, *C* spring support, *D* moving suspended capacitor plate, and *E* fixed capacitor plate (after [3.28])

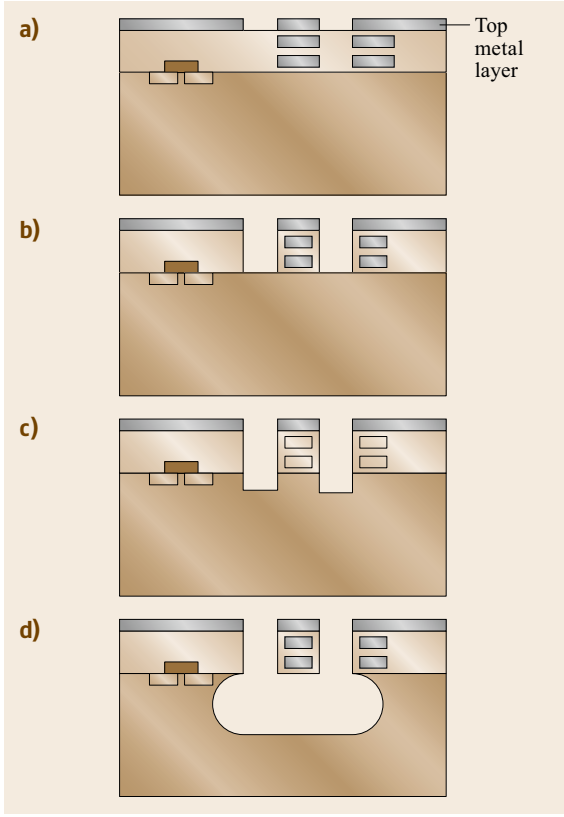


Fig. 3.27a–d Cross section of the process flow for post-CMOS dry microstructure fabrication. **(a)** Starting structure. **(b)** Anisotropic etching. **(c)** Anisotropic etching of silicon. **(d)** Isotropic etching of silicon

3.2.2 Surface Micromachining

Surface micromachining is another important MEMS microfabrication technique which can be used to create movable microstructures on top of a silicon substrate [3.30]. This technique relies on the deposition of structural thin films on a sacrificial layer which is subsequently etched away, resulting in movable micromechanical structures (beams, membranes, plate, etc.). The main advantage of surface micromachining is that extremely small sizes can be obtained. In addition, it is relatively easy to integrate the micromachined structures with on-chip electronics for increased functionality. However, due to the increased surface nonplanarity with any additional layer, there is a limit to the number of layers that can be deposited. Although one of the earliest reported MEMS structures was a surface-micromachined resonant gate transistor [3.31], material-related difficulties resulted in the termination of efforts in this area. In the mid 1980s, improvements in the field of thin-film deposition rekindled

interest in surface micromachining [3.32]. Later in the same decade polysilicon surface micromachining was introduced which opened the door to the fabrication of a variety of microsensors (accelerometers, gyroscopes, etc.) and microactuators (micromirrors, RF switches, etc.). In this section, we will concentrate on the key process steps involved in surface-micromachining fabrication and the various materials used. In addition, monolithic integration of CMOS with MEMS structures and 3-D surface micromachining are also discussed. For NEMS on CMOS integration, a review on cantilever beams fabrication can be found in [3.33].

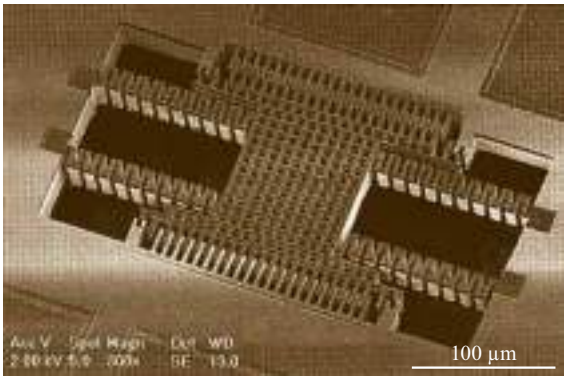


Fig. 3.28 SEM image of a comb-drive actuator fabricated using aluminum-mask post-CMOS dry release (after [3.29])

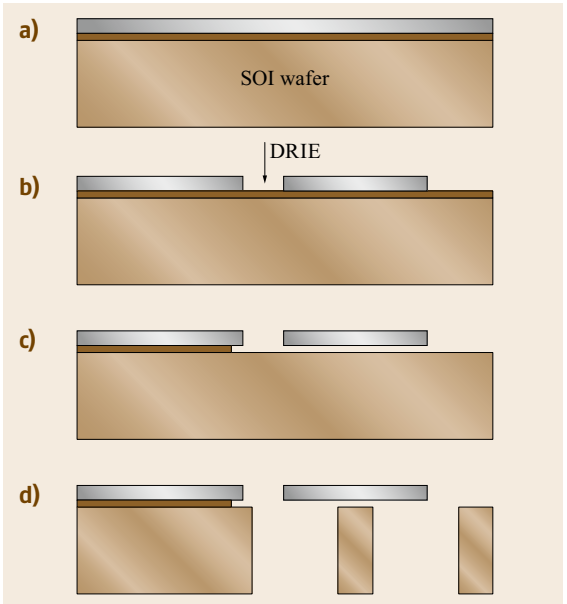


Fig. 3.29a–d DRIE processes using SOI wafers. **(a)** Starting SOI wafer. **(b)** DRIE of silicon. **(c)** Etching sacrificial layer. **(d)** Back etching

Basic Surface-Micromachining Processes

The basic surface-micromachining process is illustrated in Fig. 3.30. The process begins with a silicon substrate, on top of which a sacrificial layer is grown and patterned (Fig. 3.30a). Subsequently, the structural material is deposited and patterned (Fig. 3.30b). As can be seen the structural material is anchored to the substrate through the openings created in the sacrificial layer during the previous step. Finally, the sacrificial layer is removed, resulting in the release of the microstructures (Fig. 3.31c). In wide structures, it is usually necessary to provide access holes in the structural layer for fast sacrificial layer removal. It is also possible to seal microcavities created by the surface-micromachining technique [3.11]. This can be done at the wafer level and is a big advantage in applications such as pressure sensors which require a sealed cavity. Figure 3.31 shows two different techniques that can be used for this purpose. In the first technique, following the etching of the sacrificial layer, a LPCVD dielectric layer (oxide or nitride) is deposited to cover and seal the etch holes in the structural material (Fig. 3.32a). Since the LPCVD deposition is performed at reduced pressures, a subatmospheric pillbox microcavity can be created. In the second technique, also called reactive sealing, the polysilicon structural material is oxidized following the sacrificial layer removal (Fig. 3.32b). If access holes are small enough the grown oxide can seal the cavity. Due to the consumption of oxygen during the growth process, in this case also the cavity is subatmospheric.

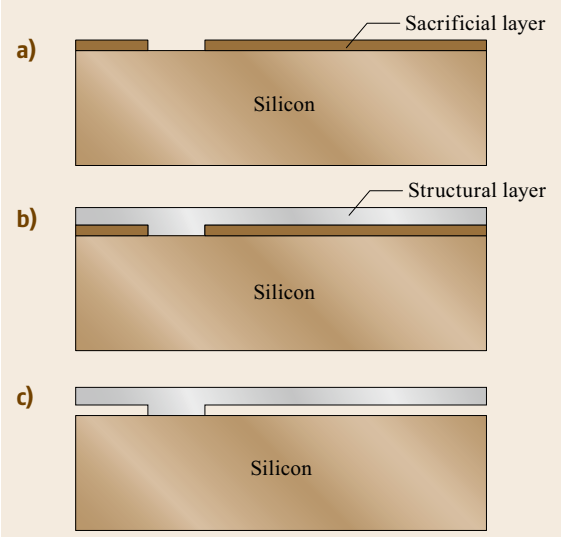


Fig. 3.30a–c Basic surface-micromachining fabrication process **(a)** silicon wafer with patterned sacrificial layer, **(b)** deposition and patterning of the structural layer, and **(c)** removal of the sacrificial layer

The most common sacrificial and structural materials are phosphosilicate glass (PSG) and polysilicon, respectively (low-temperature oxide, LTO, is also frequently used as the sacrificial layer). However, there are several other sacrificial/structural combinations that have been used to create a variety of surface-micromachined structures. Important design issues related to the choice of the sacrificial layer are:

1. Quality (pinholes, etc.)
2. Ease of deposition
3. Deposition rate
4. Deposition temperature
5. Etch difficulty and selectivity (sacrificial layer etchant should not attack the structural layer).

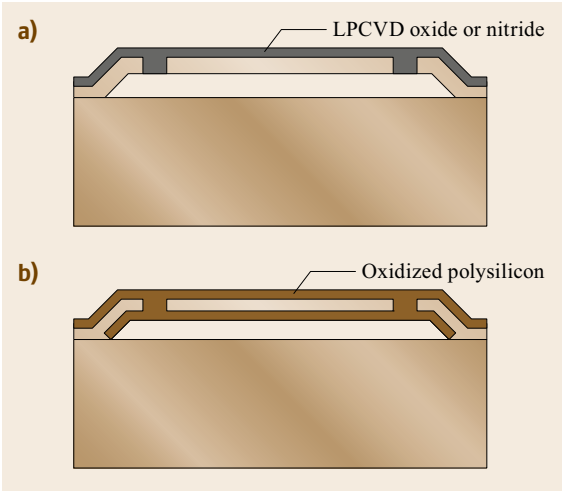


Fig. 3.31a,b Two sealing techniques for cavities created by surface micromachining

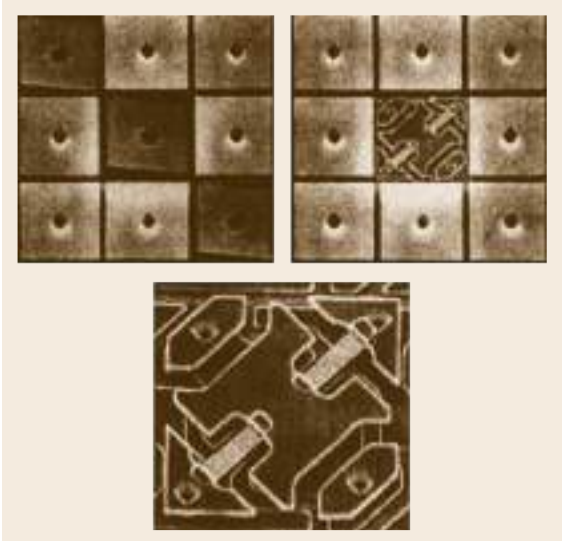


Fig. 3.32 SEM images of the Texas Instrument micromirror array (after [3.30])

The particular choice of material for the structural layer depends on the desired properties and specific application. Several important requirements are:

1. Ease of deposition
2. Deposition rate
3. Step coverage
4. Mechanical properties (internal stress, stress gradient, Young's moduli, fracture strength, and internal damping)
5. Etch selectivity
6. Thermal budget and history
7. Electrical conductivity
8. Optical reflectivity.

Two examples of commercially available surface-micromachined devices illustrate various successful sacrificial/structural combinations. The Texas Instruments (TI) deformable mirror display (DMD) spatial light modulator uses aluminum as the structural material (good optical reflectivity) and photoresist as the sacrificial layer (easy dry etch and low processing temperatures, allowing easy post-IC integration with CMOS) [3.34] (Fig. 3.32), whereas the Analog Devices microgyroscope uses polysilicon structural material and a PSG sacrificial layer (Fig. 3.33). Two additions to the collection of available structural layers

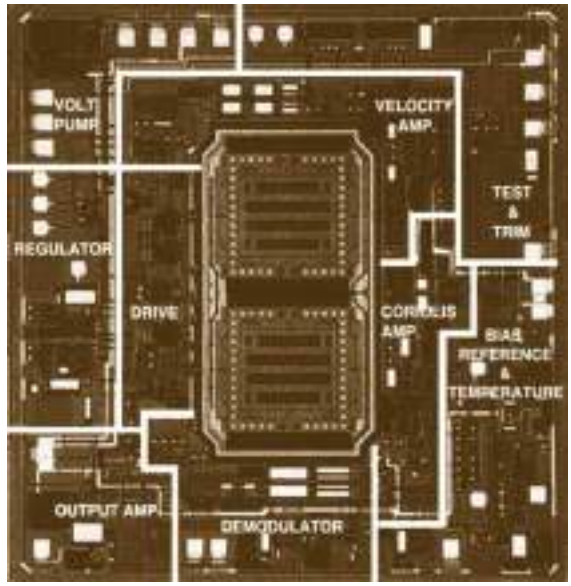


Fig. 3.33 SEM image of the Analog Devices gyroscope (after [3.35])

are polysilicon–germanium and polygermanium [3.36, 37]. These are intended as substitutes for polysilicon in applications where the high polysilicon deposition temperature (around 600 °C) is prohibitive (e.g., CMOS integration). Unlike in LPCVD of polysilicon, polygermanium (poly-Ge) and polysilicon–germanium (poly-$Si_{1-x}Ge_x$) can be deposited at temperatures as low as 350 °C (poly-Ge deposition temperature is usually lower than that for poly-SiGe). Table 3.2 summarizes important surface-micromachined sacrificial/structural combinations.

An important consideration in the design and processing of surface-micromachined structures is the issue of stiction [3.11, 38, 39], as the size of the mechanical element is reduced (NEMS) the risk is more accute. Stiction can happen during the release step if a wet etchant is used to remove the sacrificial layer or during the device lifetime. The reason for stiction during release is the surface tension of the liquid etchant, which can hold the microstructure down and cause stiction. This usually happens when the structure is compliant and does not possess enough spring constant to overcome the surface tension force of the rinsing liquid (i. e., water). There are several ways one can alleviate the release-related stiction problem. These include:

1. The use of dry or vapor phase etchant
2. The use of solvents with lower surface tension
3. Geometrical modifications
4. CO_2 critical drying
5. Freeze-drying
6. Self-assembled monolayer (SAM) or organic thin-film surface modification.

The first technique prevents stiction by not using a wet etchant, although in the case of vapor-phase release, condensation is still possible and can cause some stiction. The second method uses rinsing solvents such as methanol with a lower surface tension than water. This is usually followed by rapid evaporation of the solvent on a hot-plate. However, this technique is not optimum and many structures still stick. The third technique is geometrical, providing dimples in the structural layer in order to reduce the contact surface area and hence reduce the attractive force. The fourth and fifth techniques rely on phase change (in one case CO_2 and the other butyl alcohol) which avoids the liquid phase altogether by directly going to the gas phase. The last technique uses SAMs or organic thin films to coat the surfaces with a hydrophobic layer. The stiction that occurs during the operating lifetime of the device (in-use stiction) is due to condensation of moisture on the surfaces, electrostatic charge accumulation, or direct chemical bonding. Surface passivation using SAMs or organic thin films can be used to reduce the surface energy and reduce or eliminate capillary forces and direct chemical bonding. These organic coatings also reduce electrostatic forces if a thin layer is applied directly to the semiconductor (without an intervening oxide layer). Commonly used organic coatings include fluorinated fatty acids (TI aluminum micromirrors), silicone polymeric layers (Analog devices accelerometers), and siloxane SAMs.

Surface-Micromachining Integration with Active Electronics

Integration of surface-micromachined structures with on-chip circuitry can increase performance and simplify packaging. However, issues related to process compatibility and yield have to be carefully considered. The two most common techniques are MEMS-first and MEMS-last techniques. In the MEMS-last technique, the integrated circuit is first fabricated and surface-micromachined structures are subsequently built on top of the silicon wafer. An aluminum structural layer with a sacrificial photoresist layer is an attractive combination due to the low thermal budget of the process (TI micromirror array). However, in applications where mechanical properties of Al are not adequate, polysilicon structural material with an LTO or PSG sacrificial layer must be used. Due to the rather high deposition temperature of polysilicon, this combination requires special attention with regard to the thermal budget. For example, aluminum metallization must be avoided and substituted with refractory metals such as tungsten. This can only be achieved at the cost of greater process complexity and lower transistor performance.

The MEMS-first technique alleviates these difficulties by fabricating the microstructures at the very beginning of the process. However, if the microstructures are processed first, they have to be buried in a sealed trench to eliminate the interference of microstructures

Table 3.2 Several important surface-micromachined sacrificial–structural combinations

System	Sacrificial layer	Structural layer	Structural layer etchant	Sacrificial layer etchant
1	PSG or LTO	Poly-Si	RIE	Wet or vapor HF
2	Photoresist, polyimide	Metals (Al, Ni, Co, Ni-Fe)	Various metal etchants	Organic solvents, plasma O_2
3	Poly-Si	Nitride	RIE	KOH
4	PSG or LTO	Poly-Ge	H_2O_2 or RCA1	Wet or vapor HF
5	PSG or LTO	Poly-Si-Ge	H_2O_2 or RCA1	Wet or vapor HF

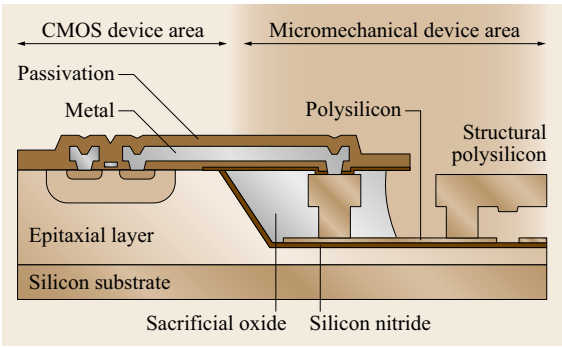


Fig. 3.34 Cross section of the Sandia MEMS-first integrated fabrication process

with subsequent CMOS processes. Figure 3.34 shows a cross section of a MEMS-first fabrication process developed at the Sandia National Laboratory [3.40]. The process starts with shallow anisotropic etching of trenches in a silicon substrate to accommodate the height of the polysilicon structures fabricated later on. A silicon nitride layer is then deposited to provide isolation at the bottom of the trenches. Next, several layers of polysilicon and sacrificial oxide are deposited and patterned in a standard surface-micromachining process. Subsequently, the trenches are completely filled with sacrificial oxide and the wafers are planarized with chemical-mechanical polishing (this avoids complication in the following lithographic steps). After an annealing step, the trenches are sealed with a nitride cap. At this point, a standard CMOS fabrication process is performed. At the end of the CMOS process the nitride cap is etched and the buried structures released by etching the sacrificial oxide.

Three-Dimensional Microstructures Using Surface Micromachining

Three-dimensional (3-D) surface microstructures can be fabricated using surface micromachining. The fabrication of hinges for the vertical assembly of MEMS was a major advance towards achieving 3-D microstructures [3.41]. Optical microsystems have greatly benefited from surface-micromachined 3-D structures. These microstructures are used as passive or active components (micromirror, Fresnel lens, optical cavity, etc.) on a silicon optical bench (silicon microphotonics). An example is a Fresnel lens that has been surface micromachined in polysilicon and then erected using hinge structures and locked in place using micromachined tabs, thus liberating the structure from the horizontal plane of the wafer [3.42, 43]. Various microactuators (e.g., comb drive, and vibromotors) have been used to move these structures out of the silicon plane and into position. Figure 3.35 shows an SEM

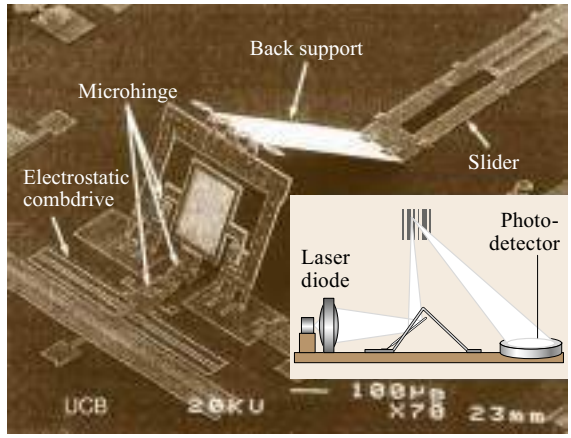


Fig. 3.35 Silicon pin-and-sample hinge scanner with 3-D surface-micromachined structures (after [3.43])

photograph of a bar-code microscanner using a silicon optical microbench with 3-D surface-micromachined structures.

3.2.3 High-Aspect-Ratio Micromachining

The bulk and surface micromachining technologies presented in the previous sections fulfill the requirements of a large group of applications. Certain applications, however, require the fabrication of high-aspect-ratio structures that is not possible with the aforementioned technologies. In this section, we describe three technologies, capable of producing structures with vertical dimensions much larger than their lateral dimensions; these are x-ray lithography (LIGA) and DRIE etching (HEXSIL and HARPSS).

LIGA

LIGA is a high-aspect-ratio micromachining process which relies on x-ray lithography and electroplating (in German: *Lithographie*, *Galvanoformung*, *Abformung*) [3.44, 45]. We already introduced the concept of the plating-through-mask technique in *Surface Micromachining* (Fig. 3.10). With standard UV photolithography and photoresists, the maximum thickness achievable is on the order of a few tens of microns and the resulting metal structures show tapered walls. LIGA is a technology based on the same plating-through-mask idea but can be used to fabricate metal structures of thickness ranging from a few microns to a few millimeters with almost vertical side-walls. This is achieved using x-ray lithography and special photoresists. Due to their short wavelength, x-rays are able to penetrate through a thick photoresist layer with no scattering and define features with lateral dimensions down to 0.2 µm (aspect ratio $> 100:1$).

The photoresists used in LIGA should comply with certain requirements, including sensitivity to x-rays, resistance to electroplating chemicals, and good adhesion to the substrate. Based on such requirements poly-(methyl methacrylate) (PMMA) is considered to be an optimal choice for the LIGA process. Application of the thick photoresist on top of the substrate can be performed by various techniques such as multiple spin-coating, precast PMMA sheets, and plasma polymerization coating. The mask structure and materials for x-ray lithography must also comply with certain requirements. The traditional masks based on glass plates with a patterned chrome thin layer are not suitable because x-rays are not absorbed by the chromium layer and the glass plate is not transparent enough. Instead, x-ray lithography uses a silicon nitride mask with gold as the absorber material (typically formed by electroplating gold to a thickness of 10–20 μm). The nitride membrane is supported by a silicon frame which can be fabricated using bulk micromachining techniques. Once the photoresist is exposed to the x-rays and developed, the process proceeds with electroplating of the desired metal. Ni is the most commonly used, although other metals and metallic compounds such us Cu, Au, NiFe, and NiW are also electroplated in LIGA processes. Good agitation of the plating solution is the key to obtaining a uniform and repeatable result during this step. A paddle plating cell, based on a windshield-wiper-like device moving only a millimeter away from the substrate surface, provides extremely reproducible agitation. Figure 3.36 shows an SEM of a LIGA microstructure fabricated with electroplating nickel.



Fig. 3.36 SEM of assembled LIGA-fabricated nickel structures (after [3.45])

Due to the high cost of the x-ray sources (synchrotron radiation), LIGA technology was initially intended for the fabrication of molds that could be used many times in hot-embossing or injection-molding processes. However, it has been also used in many

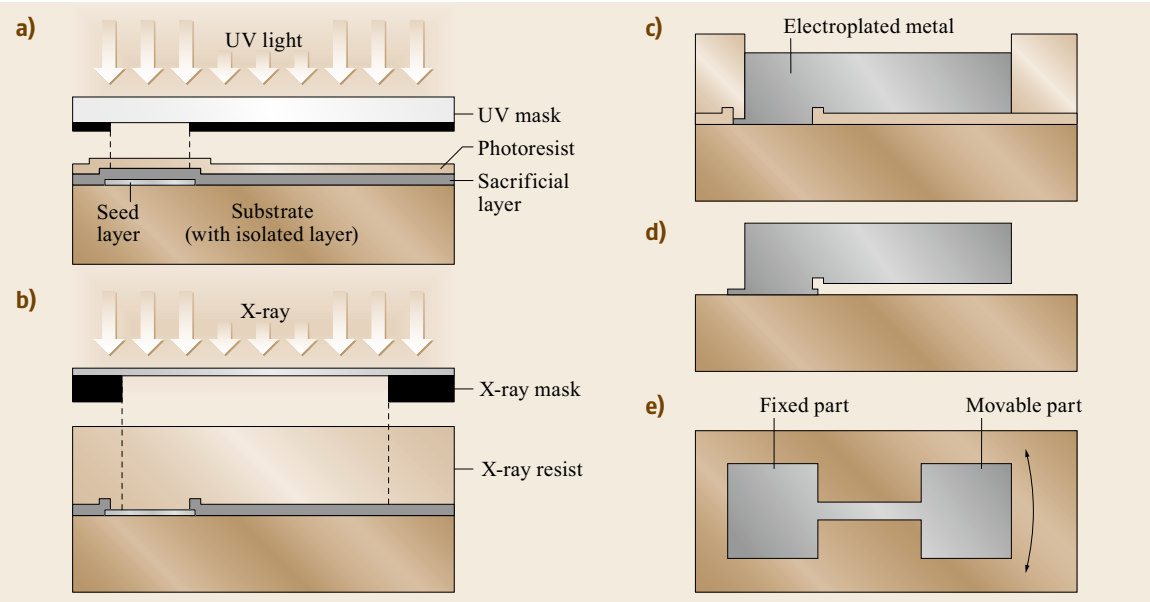


Fig. 3.37a–e Sacrificial LIGA process: **(a)** UV lithography for sacrificial layer patterning, **(b)** x-ray lithography, **(c)** electroplating, **(d)** structure releasing, and **(e)** top view of the movable structure

applications to directly form high-aspect-ratio metal structures on top of a substrate. A cheaper alternative to the LIGA process (with somehow poorer qualities) called UV-LIGA or *poor man's LIGA* has been proposed [3.46, 47]. This process uses SU-8 negative photoresists (available for spin-coating at various thickness ranging from 1 to 500 μm) and standard contact lithography equipment. Using this technique, aspect ratios larger than 20 : 1 have been demonstrated. A major problem of this alternative is the removal of the SU-8 photoresist after plating. Various methods, showing different degrees of success, have been proposed. These include: wet etching with special solvents, burning at high temperatures (600 °C), dry etching, use of a release layer, and high-pressure water-jet etching.

A variation of the basic LIGA process, shown in Fig. 3.37, permits the fabrication of electrically isolated movable structures, and thus opens more possibilities for sensor and actuator design using this technology [3.48]. This so-called sacrificial LIGA (SLIGA) starts with the patterning of the seed layer. Subsequently a sacrificial layer (e.g., titanium) is deposited and patterned. The process then proceeds as in standard LIGA until the last step, when the sacrificial layer is removed. The electroplated structures that overlap with the sacrificial layer are released in this step.

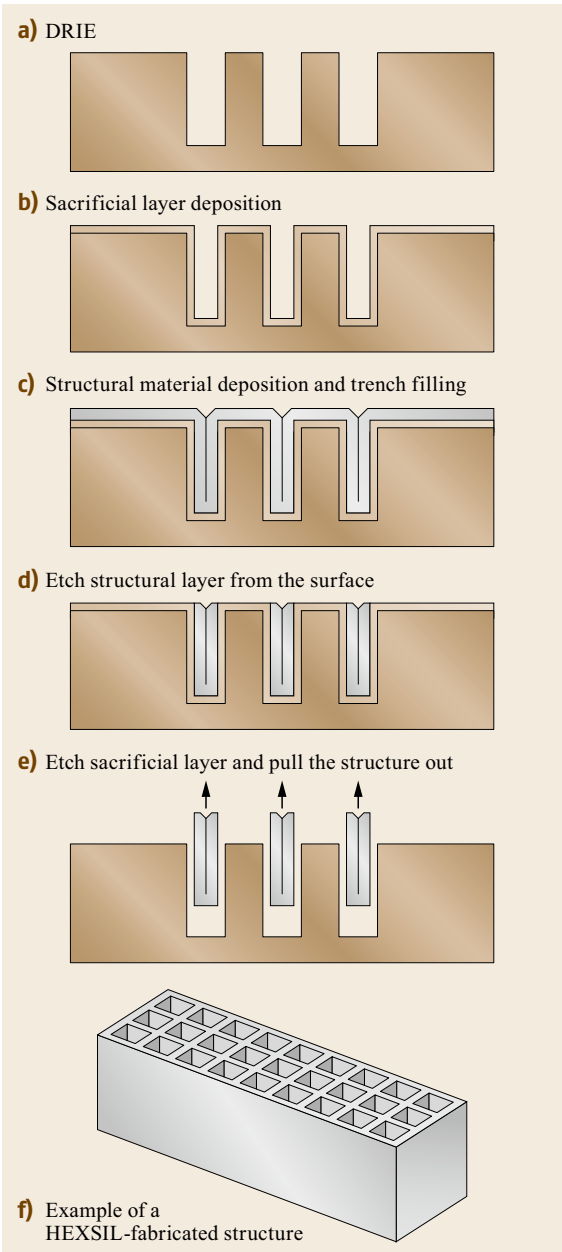


Fig. 3.38a–f HEXSIL process flow: **(a)** DRIE, **(b)** sacrificial layer deposition, **(c)** structural material deposition and trench filling, **(d)** etch structural layer from the surface, **(e)** etch sacrificial layer and pull the structure out, and **(f)** example of a HEXSIL-fabricated structure

HEXSIL

The second method for fabricating high-aspect-ratio structures, which is based on a template replication technology, is hexagonal honeycomb polysilicon (HEXSIL) [3.50]. Figure 3.38 shows a simplified process flow. A high-aspect-ratio template is first formed in a silicon substrate using DRIE. Next, a sacrificial multilayer is deposited to allow the final release of the structures. The multilayer is composed of one or more PSG nonconformal layers to provide fast etch release (≈ 20 μm/min in 49% HF) alternated with conformal

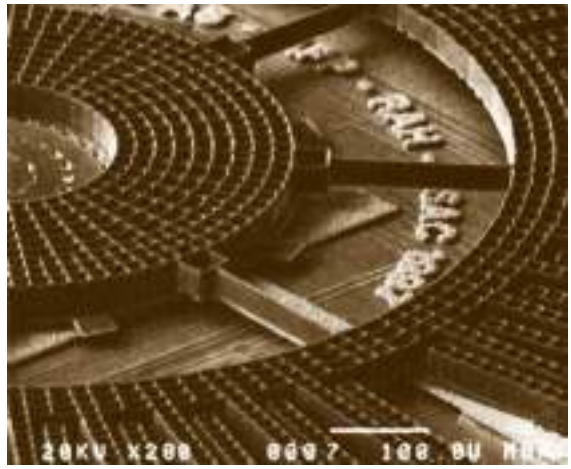


Fig. 3.39 SEM micrograph of an angular microactuator fabricated using the HEXSIL process (after [3.49])

layers of either oxide or nitride to provide enough thickness for proper release of the structures. The total thickness of the sacrificial layer has to be larger than the shrinkage or elongation of the structures caused by the relaxation of the internal (compressive or tensile) stress during the release step. Otherwise the structures will clamp themselves to the walls of the template and their retrieval will not be possible. Any material that can be conformally deposited and yet not damaged during the HF release step is suitable for the structural layer. Structures made of polysilicon, nitride, and electroless nickel [3.51] have been reported. Nickel can only be deposited in combination with polysilicon since a conductive surface is needed for the deposition to occur. After deposition of structural materials a blanket etch (poly-Si or nitride) or a mechanical lapping (nickel) is performed to remove the excess materials from the surface. Finally, a 49% HF with surfactant is used to dissolve the sacrificial layers. The process can be repeated many times using the same template, thus considerably lowering fabrication costs. Figure 3.39 shows an SEM photograph of a microactuator fabricated using the HEXSIL process.

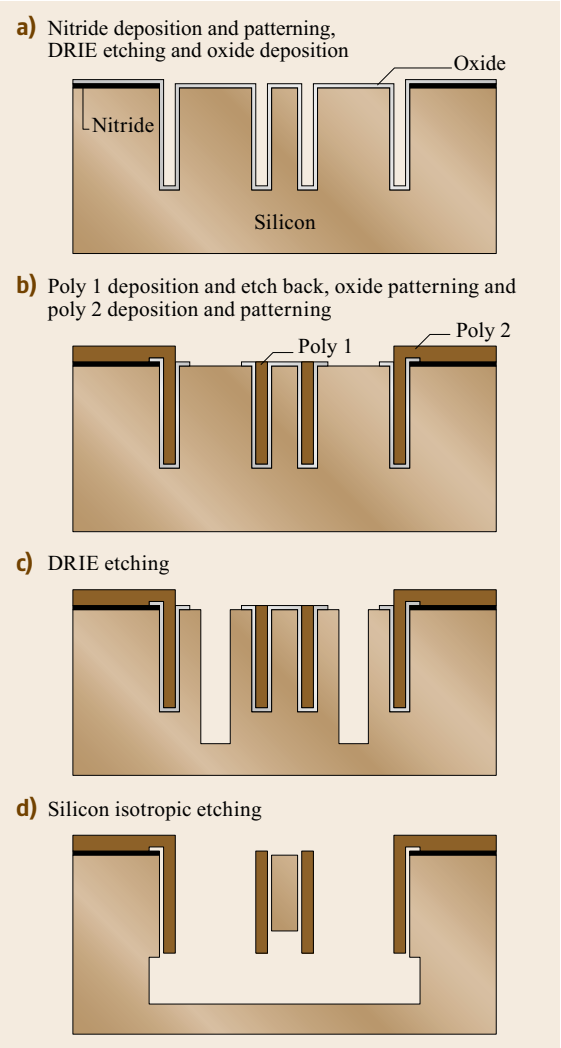


Fig. 3.40a–d HARPSS process flow: **(a)** nitride deposition and patterning, DRIE etching and oxide deposition, **(b)** poly 1 deposition and etch back, oxide patterning, and poly 2 deposition and patterning, **(c)** DRIE etching, and **(d)** silicon isotropic etching

HARPSS

The high-aspect-ratio combined poly- and single-crystal silicon (HARPSS) technology is another technique capable of producing high-aspect-ratio electrically isolated polycrystalline and single-crystal silicon microstructures with capacitive air gaps ranging from submicrometer to tens of micrometers [3.52]. The structures, tens to hundreds of micrometers thick, are defined by trenches etched with DRIE and filled with oxide and poly layers. The release of the microstructures is done at the end by means of a directional silicon etch followed by an isotropic etch. The small vertical gaps and thick structures possible with this technology find application in the fabrication of a variety of MEMS devices, particularly inertial sensors [3.53] and RF beam resonators [3.54]. Figure 3.40 shows the process flow in a cross section of a single-crystal silicon beam resonator. The HARPSS process starts with deposition and patterning of a silicon nitride layer that will be used to isolate the poly structure's connection pads from the substrate. High-aspect-ratio trenches ($\approx 5\,\mu m$ wide) are then etched into the substrate using a DRIE etch. Next, a conformal oxide layer (LPCVD) is deposited. This layer has two functions, to:

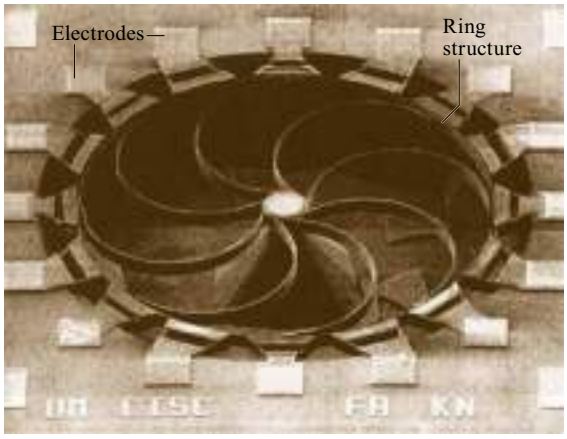


Fig. 3.41 SEM photograph of a microgyroscope fabricated using the HARPSS process (after [3.55])

1. Protect the structures during the dry etch release.
2. Define the submicrometer gap between silicon and polysilicon structures.

Following the oxide deposition, the trenches are completely filled with LPCVD polysilicon. The polysilicon is etched back and the underlying oxide is patterned to provide anchor points for the structures. A second layer of polysilicon is then deposited and patterned. Finally, the structures are released using a DRIE step followed by an isotropic silicon etch through a photoresist mask that exposes only the areas of silicon substrate surrounding the structures. It should be noted that single-crystal silicon structures are not protected at the bottom during the isotropic etch. This causes the single-crystal silicon structures to be etched vertically from the bottom, and thus be shorter than the polysilicon structures. Figure 3.41 shows an SEM photograph of a microgyroscope fabricated using the HARPSS process.

3.3 Nanofabrication Techniques

The microfabrication techniques discussed so far were mostly geared towards fabricating devices in the 1 mm to 1 μm dimensional range (submicrometer dimensions being possible in certain techniques such as HARPSS using a dielectric sacrificial layer). Last decades, a tremendous demand in fabricating submicro- (1 μm–100 nm) and nanostructures (100−1 nm range) [3.56] has led the field. This interest arises from both practical and fundamental viewpoints. At the more scientific and fundamental level, nanostructures provide an interesting tool for studying the electrical, magnetic, optical, thermal, and mechanical properties of matter at the nanometer scale. These include important quantum-mechanical phenomena (e.g., conductance quantization, bandgap modification, coulomb blockade, etc.) arising from confinement of charged carriers in structures such as quantum wells, wires, and dots (Fig. 3.42). On the practical side, nanostructures can provide significant improvements in the performance of electronic/optical devices and sensors. Furthermore some nanomaterials exhibit novel phenomena which would ultimately provide even completely new device concepts. Initially, in the electronic device area investigators have been mostly interested in fabricating nm-sized transistors in anticipation of technical difficulties forecasted in extending Moore's law beyond 32 nm resolution. In addition, optical sources and detectors having nm-size dimensions exhibit improved characteristics not achievable in larger devices (e.g., lower threshold current, improved dynamic behavior, and improved emission line width in quantum dot lasers). These improvements create novel possibilities for next-generation computation and communication devices. In the sensors area, shrinking dimensions beyond conventional optical lithography can provide major improvements in sensitivity and selectivity. Beyond, microfabrication, the so-called More Moore approach relies on higher patterning resolution, either based on charged beams (e.g., energetic electron beams) or advanced masks (such as in directed self assembly (DSA) of block copolymers (BCPs)). Recent trends go beyond these evolutionary approaches. The so-called More than Moore approach includes the idea to become truly disruptive, technologically, and therefore very innovative, for applications. Among other ways, they incorporate novel materials such as carbon nanotubes or nanoscale forms, such as in GaAs or ZnO NWs. Importantly in nanoscale (electronics) fabrication, driven by the pursuit of integration of electronic devices and sensors on flexible susbtrates, as well as system requirements for energy and information autonomy, nanofabrication techniques have been forcedly diversified from the conventional semiconductor processing techniques, such as represented by the rise of ink jet techniques [3.57].

Simply presented, one can broadly divide various nanofabrication techniques into top-down and bottom-

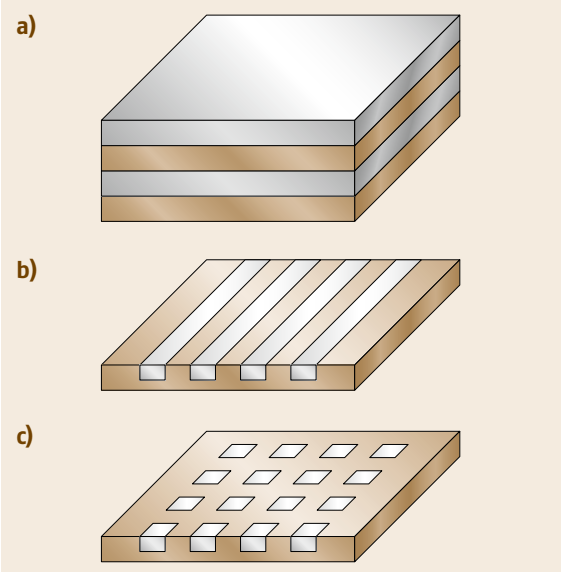


Fig. 3.42a–c Several important quantum confinement structures: **(a)** quantum well, **(b)** quantum wire, and **(c)** quantum dot

up categories. Fabrication sequences often combine both approaches. The first approach starts with a bulk or thin-film material and removes selective regions to fabricate nanostructures; in essence similar to micromachining techniques. The second method relies on integrating nanomaterials, molecular recognition and self-assembly to fabricate nanostructures out of smaller building blocks (molecules, colloids, and clusters). As can be anticipated, the top-down approach is an off-shoot of standard lithography and micromachining techniques (More Moore). On the other hand, the bottom-up approach has a more chemical engineering and material science flavor and relies on fundamentally different principles (towards More than Moore). In this chapter, we will shortly discuss several nanofabrication techniques that are not covered in other chapters of this Handbook. These include:

1. E-beam nanofabrication
2. Epitaxy and strain engineering
3. Scanning-probe techniques
4. Self-assembly and template manufacturing.

3.3.1 E-Beam Nanofabrication

In previous sections, we discussed several important lithography techniques used commonly in MEMS and microfabrication. These included various forms of UV (regular, deep, and extreme) and x-ray lithography. However, due to the lack of resolution (in case of the UV) or difficultly in manufacturing mask and radiation sources (x-ray), these techniques are not suitable for nm-scale fabrication. E-beam lithography (EBL) is an alternative and attractive technique for fabricating nanostructures [3.58]. It uses an electron beam to expose an electron-sensitive resist such as poly(methyl methacrylate) (PMMA) dissolved in trichlorobenzene (positive) or polychloromethylstyrene (negative) [3.59]. Safer commercial formulations are done, e.g., in anisole. The e-beam gun is usually part of a scanning electron microscope (SEM), although transmission electron microscopes (TEM) can also be used. Nowadays, dedicated equipments have been developed which include interferometric stages, allowing automated patterning with ultraprecise alignment and stitching at the wafer scale; i. e., so called mix and match (combination of photolithography and EBL) is possible. Although electron wavelengths of the order of 1 Å can be easily achieved, phenomena such as electron scattering in the resist limits the attainable resolutions to > 10 nm. It is a direct writing technique where beam control and pattern generation is achieved through a computer interface, providing full freedom for features design. Importantly, EBL is serial and hence has low throughput. Although this is not a major concern in fabricating devices used in studying fundamental micro/nanophysics, it severely limits industrial, large-scale nanofabrication. EBL in conjunction with processes such as lift-off, etching, and thin film deposition can be used to fabricate countless nanostructures and nanodevices based on its great pattern flexibility and aligment capability. Undoubtedly, the main application of EBL is as a tool for prototyping, where the possibility to interface nano-sized objects and novel nanomaterials largely contributed to their fundamental study and functionality development. Examples can be found in this Handbook, for instance, in Part A and B.

3.3.2 Epitaxy and Strain Engineering

Atomic-precision deposition techniques such as molecular-beam epitaxy (MBE) and metalorganic chemical vapor deposition (MOCVD) have proven to be effective tools in fabricating a variety of quantum confinement structures and devices (quantum well lasers, photodetector, resonant tunneling diodes, etc.) [3.60–62]. Although quantum wells and superlattices are the structures that lend themselves most easily to these techniques (Fig. 3.42a), quantum wires and dots have also been fabricated by adding subsequent steps such as etching and selective growth. Fabrication of quantum well and superlattice structures using epitaxial growth is a mature and well-developed field and therefore will not be discussed in this chapter. As we have concentrated the description to Si-base substrates, we only provide references on quantum wire and dot nanostructure fabrication using basic epitaxial techniques [3.63, 64]. Examples for InGaAs (on GaAs substrate) are found, e.g., in [3.65], as well as examples based in regrowth of epitaxial layers over nonplanar surfaces in [3.64, 66], strain-induced self-assembly [3.64, 67], where the most commonly used material combination is the $In_xGa_{1-x}As$/GaAs system [3.68, 69].

3.3.3 Scanning Probe Microscope Techniques

The invention of scanning probe microscopy (SPM) in the 1980s revolutionized atomic-scale imaging and spectroscopy. In particular scanning tunneling and atomic force microscopes (STM and AFM) have found widespread applications in physics, chemistry, material science, and biology. The possibility of atomic-scale manipulation, lithography, and nanomachining using such probes was considered from the beginning and has matured considerably over the past decade. In this section after a brief introduction to scanning probe

microscopes, we will discuss in detail several nanolithography and machining techniques which have been used to create nm-sized structures, mainly for their foundational relevance in nanopatterning.

SPM systems are capable of controlling the movement of an atomically sharp tip in close proximity to or in contact with a surface with subnanometer accuracy. Piezoelectric positioners are typically used in order to achieve such accuracy. High-resolution images can be acquired by raster scanning the tips over a surface while simultaneously monitoring the interaction of the tip with the surface. In scanning tunneling microscope systems a bias voltage is applied to the sample and the tip is positioned close enough to the surface that a tunneling current develops through the gap (Fig. 3.43a). Because this current is extremely sensitive to the distance between the tip and the surface, scanning the tip in the x–y-plane while recording the tunnel current permits the mapping of the surface topography with resolution on the atomic scale. In a more common mode of operation the amplified current signal is connected to the z-axis piezoelectric positioner through a feedback loop so that the current and therefore the distance are kept constant throughout the scanning. In this configuration the picture of the surface topography is obtained by recording the vertical position of the tip at each x–y-position.

The STM system only works for conductive surfaces because of the need to establish a tunneling current. The atomic force microscopy was developed as an alternative for imaging either conducting or nonconducting surfaces. In AFM the tip is attached to a flexible cantilever and is brought into contact with the surface (Fig. 3.43b). The force between the tip and the surface is detected by sensing the cantilever deflection. A topographic image of the surface is obtained by plotting the deflection as a function of the x–y-position. In a more common mode of operation a feedback loop is used to maintain a constant deflection while the topographic information is obtained from the cantilever vertical displacement. Some scanning probe systems use a combination of the AFM and the STM modes, i. e., the tip is mounted on a cantilever with electrical connection so that both the surface forces and tunneling currents can be controlled or monitored. STM systems can be operated in ultrahigh vacuum (UHV STM) or in air, whereas AFM systems are typically operated in air. When a scanning probe system is operated in air, water adsorbed onto the sample surface accumulates underneath the tip, forming a meniscus between the tip and the surface. This water meniscus plays an important role in some of the scanning probe techniques described below.

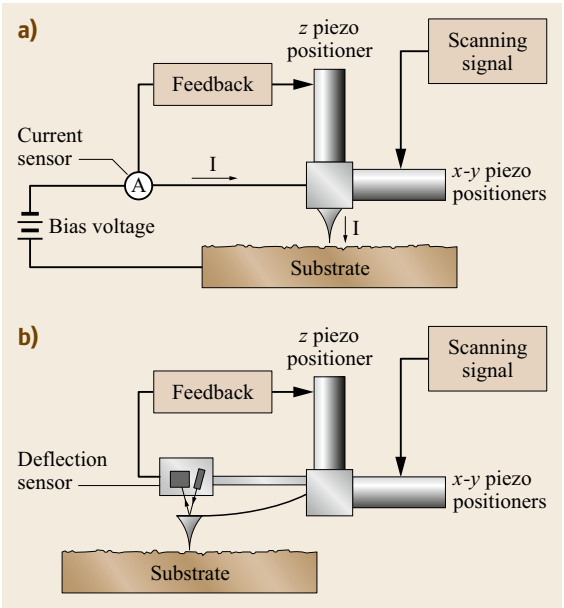


Fig. 3.43a,b Scanning probe systems: **(a)** STM and **(b)** AFM

Scanning-Probe-Induced Oxidation

Nanometer-scale local oxidation of various materials can be achieved using scanning probes operating in air and biased at a sufficiently high voltage (Fig. 3.44). A tip bias of -2 to -10 V is normally used, with a writing speed of $0.1-100\,\mu m/s$ in ambient humidity of 20–40%. It is believed that the water meniscus formed at the contact point serves as an electrolyte such that the biased tip anodically oxidizes a small region of the surface [3.71]. The most common application of this principle is the oxidation of hydrogen-passivated

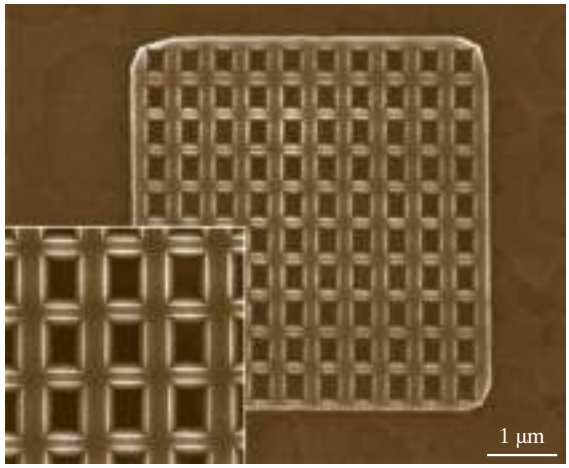


Fig. 3.44 SEM image of an inverted truncated pyramid array fabricated on a silicon SOI wafer by SPM oxidation and subsequent etch in TMAH (pitch is 500 nm) (after [3.70])

silicon. A dip in HF solution is typically used to passivate the silicon surfaces with hydrogen atoms. Patterns of oxide *written* on a silicon surface can be used as a mask for wet or dry etching. Patterns with 10 nm line width have been successfully transferred to a silicon substrate in this fashion [3.72]. Various metals have also been locally anodized using this approach, such us aluminum or titanium [3.73]. An interesting variation of this process is anodization of deposited amorphous silicon [3.74]. Amorphous silicon can be deposited at low temperature on top of many materials. The deposited silicon layer can be patterned and used as, for example, the gate of a 0.1 μm CMOS transistor [3.75], or it can be used as a mask to pattern an underlying film. The major drawback of this technique is poor reproducibility due to tip wear during the anodization. However, using AFM in noncontact mode has proved to overcome this problem [3.71].

Probe Resist Exposure and Lithography

Electrons emitted from a biased SPM tip can be used to expose a resist in the same way e-beam lithography that does (Fig. 3.45) [3.75]. Various systems have been used for this lithographic technique; these include constant-current STM, noncontact AFM, and AFM with constant tip–resist force and constant current. The systems using AFM cantilevers have the advantage that they can perform imaging and alignment tasks without exposing the resist. Resists well characterized for e-beam lithography (e.g., PMMA or SAL601) have been used with scanning probe lithography to achieve reliable sub-100 nm lithography. The procedure for this process is as follows. The wafers are cleaned and the native oxide (in the case of silicon or poly) is removed with a HF dip. Subsequently 35–100 nm-thick resist is spin-coated on top of the surface. The exposure is done by moving the SPM tip over the surface while applying a bias voltage sufficiently high to produce emission of electrons from the tip (a few tens of volts). Development of the resist is performed in standard solutions following the exposure. Features below 50 nm in width have been achieved with this procedure.

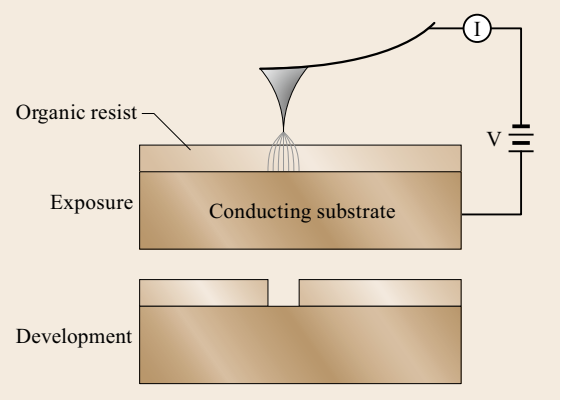


Fig. 3.45 Scanning probe lithography with organic resist

Dip-Pen Nanolithography

In dip-pen nanolithography (DPN) the tip of an AFM operated in air is *inked* with a chemical of interest and brought into contact with a surface. The ink molecules flow from the tip onto the surface as with a fountain pen. The water meniscus that naturally forms between the tip and the surface enables the diffusion and transport of the molecules, as shown in Fig. 3.46. Inking can be done by dipping the tip into a solution containing a low concentration of the molecules followed by a drying step (e.g., blow-drying with compressed difluoroethane). Line widths down to 12 nm with spatial resolution of 5 nm have been demonstrated with this technique [3.76]. Species patterned with DPN include conducting polymers, gold, dendrimers, deoxyribonucleic acid (DNA), organic dyes, antibodies, and alkanethiols. Alkanethiols have been also used as an organic monolayer mask to etch a gold layer and subsequently etch the exposed silicon substrate. One can also use a heated AFM cantilever to control the deposition of a solid organic ink. This technique was recently reported by *Sheehan* et al. in which 100 nm lines of octadecylphosphonic acid (melting point 100 °C) were written using a heated AFM probe [3.77].

Other Scanning Probe Nanofabrication Techniques

A great variety of nanofabrication techniques using scanning probe systems have been demonstrated. A good review on scanning probe lithography highlighting the potential and diversity of materials, working principles and environments, and their potential, can be found in [3.78]. Some of these are proof-of-concept demonstrations and their utility as viable and repeatable fabrication processes continue to be questioned; their direct commercial impact, simply discarded. Regardless of these limitations, one could mention, for

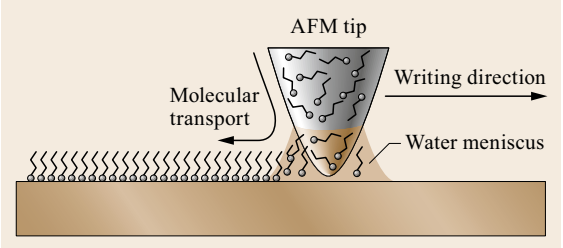


Fig. 3.46 Schematic representation of the working principle of dip-pen nanolithography

example, a STM/AFM tip acting as a plow or engraving tool for mechanical machining [3.79], which can be facilitated by heating the tip above the glass-transition temperature of a polymeric substrate material [3.80].

Atomic emission by local electric fields from the same tip has been used to form mounds (10–20 nm) of metals such as Au, Ag, and Pt [3.81] or to extract single atoms from a semiconductor surface [3.82]. Manipulation of nanoparticles, molecules, and single atoms on top of a surface has also been achieved by simply pushing or sliding them with the SPM tip [3.83]. Metals can also be locally deposited by the STM CVD technique [3.84]. Local electrochemical etching [3.85] and electrodeposition [3.86] is also possible using SPM systems.

Using a single tip to serially produce the desired modification in a surface leads to very slow fabrication processes that are impractical for mass production. Many of the scanning probe techniques developed so far, however, could also be performed by an array of tips, which would increase throughput and make them more competitive compared with other parallel nanofabrication processes. This approach has been demonstrated for imaging, lithography [3.87], and data storage [3.88] using both one- and two-dimensional arrays of scanning probes. With the development of larger arrays with advances in individual control of force, vertical position, and current, we could have seen these techniques being incorporated as standard fabrication processes in the industry.

3.3.4 Self-Assembly and Template Manufacturing

Self-assembly is a nanofabrication technique that involves aggregation of colloidal nanoparticles into the final desired structure [3.89]. This aggregation can be either spontaneous (entropic) and due to the thermodynamic minima (energy minimization) constraints, or chemical and due to the complementary binding of organic molecules and supramolecules (molecular self-assembly) [3.90]. Molecular self-assembly is one of the most important techniques used in biology for the development of complex functional structures. Since these techniques require that the target structures be thermodynamically stable, it tends to produce structures that are relatively defect-free and self-repairing. Self-assembly is by no means limited to molecules or the nanodomain and can be carried out on just about any scale, making it a powerful bottom-up assembly and manufacturing method (multiscale ordering). Another attractive feature of this technique relates to the possibility of combining self-assembly properties of organic molecules with the electronic, magnetic, and photonic properties of inorganic components. Template manufacturing is another bottom-up technique which utilizes material deposition (electroplating, CVD, etc.) into nanotemplates in order to fabricate nanostructures. Due to the simplicity and flexibility of electrochemistry for plating and surface finishing of a broad range of materials, its principle has recently been widely used for electrochemical fabrication of various metallic nanostructures based on various templates. For example, electrochemical deposition has been used to deposit large arrays of nanostructures in nanoporous templates, such as porous alumina. This template-based deposition typically provides metal nanowires as small as 25 nm in diameter and a few micrometers in length [3.91].

The nanotemplates used to fabricate nanostructures are usually prepared using self-assembly techniques. In the following sections, we will discuss various important self-assembly and template manufacturing techniques, which have been heavily investigated. A most sophisticated version of nanofabrication is DSA of BCPs. DSA of BCPs currently holds the highest expectations for More Moore approach, as it is easily compatible with conventional planar technology. Introductory and comprehensive works can be consulted, respectively, in [3.92] and [3.93].

Physical and Chemical Self-Assembly

The central theme behind the self-assembly process is spontaneous (physical) or chemical aggregation of colloidal nanoparticles [3.94]. Spontaneous self-assembly exploits the tendency of monodispersed nano or submicro colloidal spheres to organize into a face-centered cubic (fcc) lattice. The force driving this process is the desire of the system to achieve a thermodynamically stable state (minimum free energy). In addition to spontaneous thermal self-assembly, gravitational, convective, and electrohydrodynamic forces can also be used to induce aggregation into complex 3-D structures. Chemical self-assembly requires the attachment of a single-molecular organic layer (self-assembled monolayer or SAM) to the colloidal particles (organic or inorganic) and subsequent self-assembly of these components into complex structures using molecular recognition and binding.

Physical Self-Assembly

This is an entropy-driven method that relies on spontaneous organization of colloidal particles into a relatively stable structure through noncovalent interactions; for example, colloidal polystyrene spheres can be assembled into a 3-D structure on a substrate which is held vertically in the colloidal solution (Fig. 3.47) [3.95, 96]. Upon the evaporation of the solvent, the spheres aggregate into a hexagonal close-packed (hcp) structure. The

interstitial pore size and density are determined by the polymer sphere size. The polymer spheres can be etched into smaller sizes after forming the hcp arrays, thereby altering the template pore separations [3.97]. This technique can fabricate large patterned areas in a quick, simple, and cost-effective way. A classic example is the natural assembly of on-chip silicon photonic-bandgap crystals [3.95] which are capable of reflecting the light arriving in any direction in a certain wavelength range [3.98]. In this method, a thin layer of silica colloidal spheres is assembled on a silicon substrate. This is achieved by placing a silicon wafer vertically in a vial containing an ethanolic suspension of silica spheres. A temperature gradient across the vial aids the flow of silica spheres. Figure 3.48 shows a cross-sectional SEM image of a thin planar opal template assembled directly on a Si wafer from 855 nm spheres. Once such a template is prepared, LPCVD can be used to fill the interstitial spaces with Si, so that the high refractive index of silicon provides the necessary bandgap.

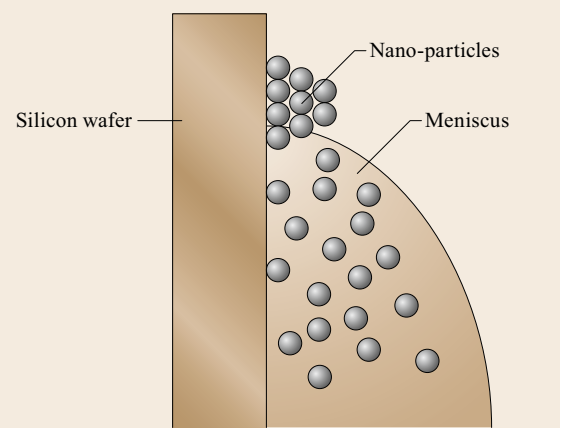


Fig. 3.47 Colloidal particle self-assembly onto solid substrates upon drying in vertical position

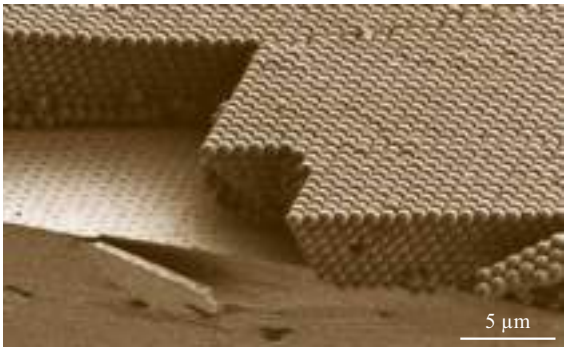


Fig. 3.48 Cross-sectional SEM image of a thin planar opal silica template (spheres 855 nm in diameter) assembled directly on a Si wafer (after [3.95])

One can also deposit colloidal particles onto a patterned substrate (template-assisted self-assembly, TASA) [3.99, 100]. This method is based on the principle that, when an aqueous dispersion of colloidal particles is allowed to dewet from a solid surface which is already patterned, the colloidal particles are trapped by the recessed regions and assemble into aggregates with shapes and sizes determined by the geometric confinement provided by the template. The patterned arrays of templates can be fabricated using conventional contact-mode photolithography which gives control over the shape and dimensions of the templates, thereby allowing the assembly of complex structures from colloidal particles. The cross-sectional view of a fluidic cell used in TASA is shown in Fig. 3.49. This method can be used to fabricate a variety of polygonal and polyhedral aggregates which are difficult to generate [3.101].

Chemical Self-Assembly

Organic and supramolecular SAMs play a critical role in colloidal particle self-assembly. SAMs are robust organic molecules which are chemically adsorbed onto solid substrates [3.102]. Most often they have a hydrophilic (polar) head which can be bonded to various solid surfaces and a long hydrophobic (nonpolar) tail which extends outward. SAMs are formed by the immersion of a substrate in a dilute solution of the molecule in an organic solvent or water (liquid phase) or by exposure to an atmosphere containing such a molecule (gas phase). The resulting film is a dense organization of molecules arranged to expose the end group. The durability of the SAM is highly dependent on the effectiveness of the anchoring to the surface of the substrate. SAMs have been widely studied because the end group can be functionalized to form precisely arranged molecular arrays for various applications ranging from simple, ultrathin insulators and lubricants to complex biological sensors.

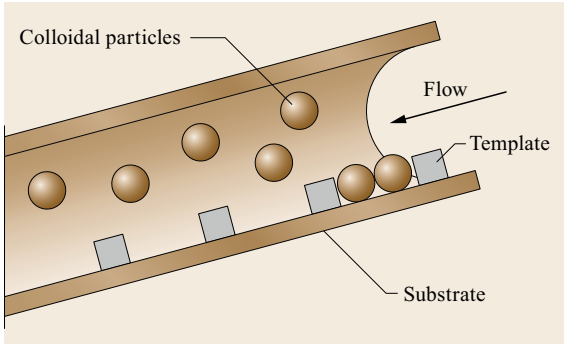


Fig. 3.49 A cross-sectional view of the fluidic cell used for template-assisted self-assembly

Chemical self-assembly uses organic or supramolecular SAMs as the binding and recognition sites for fabricating complex 3-D structures from colloidal nanoparticles. The most commonly used organic monolayers include:

1. Organosilicon compounds on glass and oxidized silicon
2. Alkanethiols, dialkyl disulfides, and dialkyl sulfides on gold
3. Fatty acids on alumina and other metal oxides
4. DNA.

Octadecyltrichlorosilane (OTS) is the most common organosilane used in the formation of SAMs, mainly because of the fact that it is simple, readily available, and forms good, dense layers [3.103, 104]. Alkyltrichlorosilane monolayers can be prepared on clean silicon wafers whose surface is SiO_2 (with almost 5×10^{14} SiOH groups/cm^2). Figure 3.50 shows a schematic representation of the formation of alkylsiloxane monolayers by adsorption of alkyltrichlorosilane from solution onto Si/SiO_2 substrates. Since the silicon–chloride bond is susceptible to hydrolysis, a limited amount of water has to be present in the system in order to obtain good-quality monolayers. Monolayers made of methyl- and vinyl-terminated alkylsilanes are autophobic to the hydrocarbon solution and hence emerge uniformly dry from the solution, whereas monolayers made of ester-terminated alkylsilanes emerge wet from the solution used in their formation. The disadvantage of this method is that, if the alkyltrichlorosilane in the solvent adhering to the substrate is exposed to water, a cloudy film is deposited on the surface due to the formation of a gel of polymeric siloxane. One solution to this problem is the use of alkyldimethylchlorosilanes, which have a single anchoring point, and thus cannot form polymers. Chlorosilanes are sometimes preferred over alkoxysilanes because of their higher reactivity. However, the reactivity of chlorosilanes severely limits the range of functional groups that can be introduced at the end of the hydrocarbon tail. On the contrary, methoxysilanes and ethoxysilanes are commonly available with many functional groups including amino, mercapto, epoxy, and thiocyanate groups, which are often necessary for subsequent binding of colloidal particles and biomolecules. Gas-phase deposition of these molecules yields more uniform layers compared with liquid-phase procedures [3.105]. Soft-lithography-like molds have been used to obtain reactive silane patterns from the gas phase by taking advantage of the characteristic permeability of polydimethylsiloxane (PDMS) to volatile molecules [3.106].

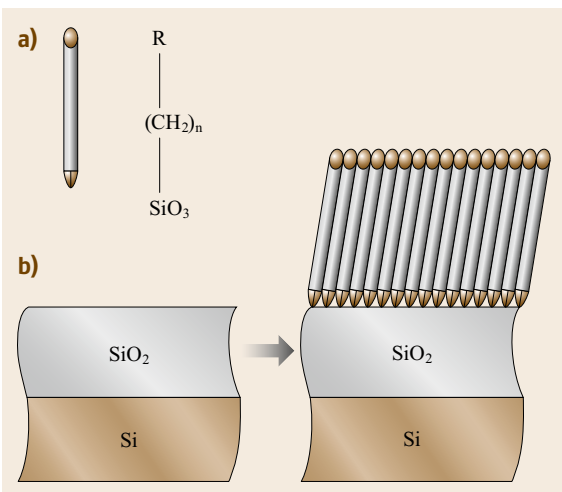


Fig. 3.50 **(a)** Alkylsiloxane formed from the adsorption of alkyltrichlorosilane on Si/SiO_2 substrates. **(b)** Schematic representation of the process

Another important organic SAM system is the alkanethiols ($X(CH_2)_nSH$, where X is the end group) on gold [3.102, 107–109]. A major advantage of using gold as the substrate material is that it does not have a stable oxide and thus can be handled in ambient conditions. When a fresh, clean, hydrophilic gold substrate is immersed (several min to several h) into a dilute solution (10^{-3} M) of the organic sulfur compound (alkanethiols) in an inorganic solvent a close-packed, oriented monolayers can be obtained. Sulfur is used as the head group because of its strong interaction with the gold substrate (44 kcal/mol), resulting in the formation of a close-packed, ordered monolayer. The end group of alkanethiol can be modified to render the adsorbed layer hydrophobic or hydrophilic. Another method for depositing alkanethiol SAM is soft lithography. This technique is based on inking a PDMS stamp with alkanethiol and its subsequent transfer to planar or nonplanar substrates. Alkanethiol-functionalized surfaces (planar, nonplanar, spherical) can also be used to self-assemble a variety of intricate 3-D structures [3.110].

Carboxylic acid derivatives self-assemble on surfaces (e.g., glass, Al_2O_3, and Ag_2O) through an acid–base reaction, giving rise to monolayers of fatty acids [3.111]. The time required for the formation of a complete monolayer increases with decreasing concentration. Higher concentration of the carboxylic acid is required to form a monolayer on gold as compared with on Al_2O_3. This is due to differences in the affinity of the COOH groups (more affinity to Al_2O_3 and glass than gold) and also the surface concentration of the salt-forming oxides of the two substrates. In the case

of amorphous metal oxide surfaces, the chemisorption of alkanoic acids is not unique. For example, on Ag_2O, the two oxygen atoms of the carboxylate bind to the substrate in a nearly symmetrical manner, thus resulting in ordered monolayers with a chain tilt angle of 15–25° from the surface normal. However, on CuO and Al_2O_3, the oxygen atoms bind themselves symmetrically and the chain tilt angle is close to 0°. The structure of the monolayers is thus a balance of the various interactions taking place in the polymer chains.

Deoxyribonucleic acid (DNA), the framework on which all life is built, can be used to self-assemble nanomaterials into useful macroscopic aggregates that display a number of desirable physical properties [3.112]. DNA consists of two strands, which are coiled around each other to form a double helix. When the two strands are uncoiled singular strands of nucleotides are left. These nucleotides consist of a sugar (pentose ring) a phosphate (PO_4), and a nitrogenous base. The order and architecture of these components is essential for the proper structure of a nucleotide. There are typically four nucleotides found in DNA: adenine (A), guanine (G), cytosine (C), and thymine (T). A key property of the DNA structure is that the nucleotides described bind specifically to another nucleotide when arranged in the two-strand double helix (A to T, and C to G). This specific bonding capability can be used to assemble nanophase material and nanostructures [3.113]. For example, nucleotide-functionalized nanogold particles have been assembled into complex 3-D structures by attaching DNA strands to the gold via an enabler or linker [3.114]. In a separate work DNA was used to assemble nanoparticles into macroscopic materials. This method uses alkane dithiol as the linker molecule to connect the DNA template to the nanoparticle. The thiol groups at each end of the linker molecule covalently attach themselves to the colloidal particles to form aggregate structures [3.115].

Template Manufacturing

Template manufacturing refers to a set of techniques that can be used to fabricate 3-D organic or inorganic structures from a nanotemplate. These templates differ in material, pattern, feature size, overall template size, and periodicity. Although nanotemplates can be fabricated using EBL, the serial nature of this technique prohibits its widespread application. Self-assembly is the preferred technique to produce large-area nanotemplates in a massively parallel fashion. Several nanotemplates have been investigated for use in template manufacturing. These include polymer colloidal spheres, alumina membranes, and nuclear-track etched membranes. Colloidal spheres can be deposited in a regular 3-D array using the techniques described in the previous section (Figs. 3.47– 3.49). Porous aluminum oxide membranes can be fabricated by the anodic oxidation of aluminum [3.116]. The oxidized film consists of columnar arrays of hexagonal close-packed pores with separation comparable to the pore size. By controlling the electrolyte species, temperature, anodizing voltage, and time, different pore sizes, densities, and heights can be obtained. The pore size and depth can further be adjusted by etching the oxide in an appropriate acid. Templates of porous polycarbonate or mica membranes can be fabricated by nuclear-track etched membranes [3.117]. This technique is based on the passage of high-energy decay fragments from a radioactive source through a dielectric material. The particles leave behind chemically active damaged tracks which can subsequently be etched to create pores through the thickness of the membrane [3.118, 119]. Unlike other methods, the pore separation and hence the pore density is independent of the pore size. The pore density is only determined by the irradiation process. More recently, electrochemical deposition of metallic nanowires on the step edges of highly oriented pyrolytic graphite (HOPG) templates have also been demonstrated, which produces metallic nanowires with diameters as small as 15 nm. This method has been successfully used for metals such as Cu, Ni, Au, and Pd [3.120, 121].

Subsequent to template fabrication, the interstitial spaces (in the case of colloidal spheres) or pores (in the case of alumina and polycarbonate membranes) in the template are filled with the desired material [3.97, 122]. This can be done by using a variety of deposition techniques such as electroplating and CVD. The final structure can be a composite of nanotemplate and deposited material or the template can be selectively etched, resulting in an air-filled 3-D complex structure. For example, Ni [3.123], Fe [3.124], and Co [3.125] nanowires have been electrochemically grown into porous template matrices. Three-dimensional photonic crystals have been fabricated by electrochemical deposition of CdSe and silicon into polystyrene and silica colloidal assembly templates [3.95, 126]. An interesting example of template-assisted manufacturing is the synthesis of nanometer metallic barcodes [3.127]. These nanobarcodes are prepared by electrochemical reduction of metallic ions into the pores of an aluminum oxide membrane, followed by their release through etching of the template [3.97].

3.4 Summary and Conclusions

In this chapter, we have discussed various micro/nanofabrication techniques used to manufacture structures of a wide range of dimensions (mm–nm). Starting with some of the most common microfabrication techniques (lithography, deposition, and etching), we have presented an array of micromachining and MEMS technologies which can be used to fabricate microstructures down to $\approx 1\,\mu$m. These techniques have attained an adequate level of maturity to enable a variety of MEMS-based commercial products (pressure sensors, accelerometers, gyroscopes, etc.). More recently, nm-size structures have attracted an enormous amount of interest. This is mainly due to their unique electrical, magnetic, optical, thermal, and mechanical properties, leading to a variety of electronic, photonic, and sensing devices with a superior performance compared with their macro counterparts. Subsequent to our discussion on MEMS and micromachining, we presented several important or succesful nanofabrication techniques. Although e-beam and other high-resolution lithography techniques can be used to fabricate nm-size structures, in a top-down fashion, their serial nature and/or cost preclude their widespread application. This has forcedly multiplied the number of nanofabrication processes used by investigators. Efforts to develop alternative and potentially superior techniques such as strain engineering, self-assembly, and nanoimprint lithography have been made. More recently, the focus on novel nanomaterials and bottom-up nanostructured materials synthesis are promoting the quest for a more diverse and flexible combination of the existing methods in (micro and) nanofabrication, beyond conventional semiconductor technology. A paradigmatic example comes from the requirements of werable electronics, boosting the revisiting of, e.g., injet techniques for flexible electronics, while maximum portability requires signal and energy autonomy, therefore, radically advanced device concepts, design and systems.

References

3.1 S.A. Campbell: *The Science and Engineering of Microelectronic Fabrication* (Oxford Univ. Press, New York 2001)

3.2 C.J. Jaeger: *Introduction to Microelectronic Fabrication* (Prentice Hall, New Jersey 2002)

3.3 J.D. Plummer, M.D. Deal, P.B. Griffin: *Silicon VLSI Technology* (Prentice Hall, New Jersey 2000)

3.4 J.E. Bjorkholm: EUV lithography: The successor to optical lithography, Intel Technol. J. **2**, 1–8 (1998)

3.5 S. Owa, H. Nagasaka: Advantage and feasibility of immersion lithography, J. Microlithogr. Microfabr. Microsyst. **3**, 97–103 (2004)

3.6 H.U. Krebs, M. Störmer, J. Faupel, E. Süske, T. Scharf, C. Fuhse, N. Seibt, H. Kijewski, D. Nelke, E. Panchenko, M. Buback: Pulsed laser deposition (PLD) – A versatile thin film technique, Adv. Solid State Phys. **43**, 505–517 (2003)

3.7 M. Leskela, M. Ritala: Atomic layer deposition chemistry: recent developments and future challenges, Angew. Chem. Int. Ed. **42**, 5548–5554 (2003)

3.8 J.L. Vossen: *Thin Film Processes* (Academic, New York 1976)

3.9 M. Gad-el-Hak (Ed.): *The MEMS Handbook* (CRC, Boca Raton 2002)

3.10 T.-R. Hsu: *MEMS and Microsystems Design and Manufacture* (McGraw Hill, New York 2002)

3.11 G.T.A. Kovacs: *Micromachined Transducers Sourcebook* (McGraw Hill, New York 1998)

3.12 P. Rai-Choudhury (Ed.): *Handbook of Microlithography, Micromachining and Microfabrication. Vol. 2: Micromachining and Microfabrication* (SPIE/IEE, Bellingham, Washington/London 1997)

3.13 M.F. Aimi, M.P. Rao, N.C. MacDonald, A.S. Zuruzi, D.P. Bothman: High-aspect-ratio bulk micromachining of titanium, Nat. Mater. **3**, 103–105 (2004)

3.14 T.J. Cotler, M.E. Elta: Plasma-etch technology, IEEE Circuits Devices Mag. **6**, 38–43 (1990)

3.15 U. Gosele, Q.Y. Tong: Semiconductor wafer bonding, Annu. Rev. Mater. Sci. **28**, 215–241 (1998)

3.16 Q.Y. Tong, U. Gosele: *Semiconductor Wafer Bonding: Science and Technology* (Wiley, New York 1999)

3.17 F. Niklaus, P. Enoksson, E. Kalveston, G. Stemme: Void-free full-wafer adhesive bonding, J. Micromech. Microeng. **11**, 100–107 (2000)

3.18 C.A. Harper: *Electronic Packaging and Interconnection Handbook* (McGraw-Hill, New York 2000)

3.19 W.H. Ko, J.T. Suminto, G.J. Yeh: Bonding techniques for microsensors. In: *Micromachining and Micropackaging for Transducers*, ed. by W.H. Ko (Elsevier, Amsterdam 1985)

3.20 G.T.A. Kovacs, N.I. Maluf, K.A. Petersen: Bulk micromachining of silicon, Proc. IEEE **86**(8), 1536–1551 (1998)

3.21 K. Najafi, K.D. Wise, T. Mochizuki: A high-yield IC-compatible multichannel recording array, IEEE Trans. Electron Devices **32**, 1206–1211 (1985)

3.22 A. Selvakumar, K. Najafi: A high-sensitivity z-axis capacitive silicon microaccelerometer with a tortional suspension, J. Microelectromech. Syst. **7**, 192–200 (1998)

3.23 H. Baltes, O. Paul, O. Brand: Micromachined thermally based CMOS microsensors, Proc. IEEE **86**(8), 1660–1678 (1998)

3.24 B. Eyre, K.S.J. Pister, W. Gekelman: Multi-axis microcoil sensors in standard CMOS. In: *Proc. SPIE*

Conf. Micromach. Devices Compon., Austin (1995) pp. 183–191
3.25 K.A. Shaw, Z.L. Zhang, N.C. MacDonnald: SCREAM: A single mask, single-crystal silicon process for microelectromechanical structures. In: *Proc. IEEE Workshop Micro Electro Mech. Syst., Fort Lauderdale* (1993) pp. 155–160
3.26 G.K. Fedder, S. Santhanam, M.L. Reed, S.C. Eagle, D.F. Guillo, M.S.C. Lu, L.R. Carley: Laminated high-aspect-ratio microstructures in a conventional CMOS process. In: *Proc. IEEE Workshop Micro Electro Mech. Syst., San Diego* (1996) pp. 13–18
3.27 B.P. Van Drieenhuizen, N.I. Maluf, I.E. Opris, G.T.A. Kovacs: Force-balanced accelerometer with mG resolution fabricated using silicon fusion bonding and deep reactive ion etching. In: *Proc. Int. Conf. Solid-State Sens. Actuators, Chicago* (1997) pp. 1229–1230
3.28 N.C. MacDonald: SCREAM MicroElectroMechanical Systems, Microelectron. Eng. **32**, 51–55 (1996)
3.29 X. Huikai, L. Erdmann, Z. Xu, K.J. Gabriel, G.K. Fedder: Post-CMOS processing for high-aspect-ratio integrated silicon microstructures, J. Microelectromech. Syst. **11**, 93–101 (2002)
3.30 J.M. Bustillo, R.S. Muller: Surface micromachining for microelectromechanical systems, Proc. IEEE **86**(8), 1552–1574 (1998)
3.31 H.C. Nathanson, W.E. Newell, R.A. Wickstrom, J.R. Davis: The resonant gate transistor, IEEE Trans. Electron Devices **14**, 117–133 (1967)
3.32 R.T. Howe, R.S. Muller: Polycrystalline silicon micromechanical beams. In: *Proc. Electrochem. Soc. Spring Meet., Montreal* (1982) pp. 184–185
3.33 G. Rius, F. Pérez Murano: Nanocantilever beam fabrication techniques in silicon. In: *Nanocantilever Beams: Modeling, Fabrication, and Applications*, ed. by I. Voiculescu, M. Zaghloul (CRC, Boca Raton 2015) pp. 3–36
3.34 P.F. Van Kessel, L.J. Hornbeck, R.E. Meier, M.R. Douglass: A MEMS-based projection display, Proc. IEEE **86**(8), 1687–1704 (1998)
3.35 J.A. Geen, S.J. Sherman, J.F. Chang, S.R. Lewis: Single-chip surface-micromachined integrated gyroscope with 50°/h root Allan deviation, IEEE J. Solid-State Circuits **37**, 1860–1866 (2002)
3.36 A.E. Franke, D. Bilic, D.T. Chang, P.T. Jones, R.T. Howe, G.C. Johnson: Post-CMOS integration of germanium microstructures. In: *Proc. Micro Electro Mech. Syst., Orlando* (1999) pp. 630–637
3.37 S. Sedky, P. Fiorini, M. Caymax, S. Loreti, K. Baert, L. Hermans, R. Mertens: Structural and mechanical properties of polycrystalline silicon germanium for micromachining applications, J. Microelectromech. Syst. **7**, 365–372 (1998)
3.38 N. Tas, T. Sonnenberg, H. Jansen, R. Legtenberg, M. Elwenspoek: Stiction in surface micromachining, J. Micromech. Microeng. **6**, 385–397 (1996)
3.39 R. Maboudian, R.T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B **15**(1), 1–20 (1997)
3.40 J.H. Smith, S. Montague, J.J. Sniegowski, J.R. Murray, P.J. McWhorter: Embedded micromechanical devices for the monolithic integration of MEMS with CMOS. In: *Proc. Int. Electron Devices Meet., Washington* (1995) pp. 609–612
3.41 K.S.J. Pister, M.W. Judy, S.R. Burgett, R.S. Fearing: Microfabricated hinges: 1 mm vertical features with surface micromachining. In: *Proc. 6th Int. Conf. Solid-State Sens. Actuators, San Francisco* (1991) pp. 647–650
3.42 L.Y. Lin, S.S. Lee, M.C. Wu, K.S.J. Pister: Micromachined integrated optics for free space interconnection. In: *Proc. IEEE Micro Electro Mech. Syst. Workshop, Amsterdam* (1995) pp. 77–82
3.43 R.S. Muller, K.Y. Lau: Surface-micromachined microoptical elements and systems, Proc. IEEE **86**(8), 1705–1720 (1998)
3.44 E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Munchmeyer: Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic molding (LIGA process), Microelectron. Eng. **4**, 35–56 (1986)
3.45 H. Guckel: High-aspect-ratio micromachining via deep x-ray lithography, Proc. IEEE **86**(8), 1586–1593 (1998)
3.46 K.Y. Lee, N. LaBianca, S.A. Rishton, S. Zolgharnain, J.D. Gelorme, J. Shaw, T.H.P. Chang: Micromachining applications of a high resolution ultra-thick photoresist, J. Vac. Sci. Technol. B **13**, 3012–3016 (1995)
3.47 K. Roberts, F. Williamson, G. Cibuzar, L. Thomas: The fabrication of an array of microcavities utilizing SU-8 photoresist as an alternative 'LIGA' technology. In: *Proc. 13th Bienn. Univ./Gov./Ind. Microelectron. Symp. (IEEE), Minneapolis* (1999) pp. 139–141
3.48 C. Burbaum, J. Mohr, P. Bley, W. Ehrfeld: Fabrication of capacitive acceleration sensors by the LIGA technique, Sens. Actuators A Phys. **A27**, 559–563 (1991)
3.49 D.A. Horsley, M.B. Cohn, A. Singh, R. Horowitz, A.P. Pisano: Design and fabrication of an angular microactuator for magnetic disk drives, J. Microelectromech. Syst. **7**, 141–148 (1998)
3.50 C.G. Keller, R.T. Howe: Hexsil bimorphs for vertical actuation. In: *Dig. Tech. Pap. 8th Int. Conf. Solid-State Sens. Actuators Eurosens. IX, Stockholm* (1995) pp. 99–102
3.51 C.G. Keller, R.T. Howe: Nickel-filled hexsil thermally actuated tweezers. In: *Dig. Tech. Pap. 8th Int. Conf. Solid-State Sens. Actuators Eurosens. IX, Stockholm* (1995) pp. 376–379
3.52 F. Ayazi, K. Najafi: High aspect-ratio combined poly and single-crystal silicon (HARPSS) MEMS technology, J. Microelectromech. Syst. **9**, 288–294 (1998)
3.53 F. Ayazi, K. Najafi: A HARPSS polysilicon vibrating ring gyroscope, J. Microelectromech. Syst. **10**, 169–179 (2001)
3.54 Y.S. No, F. Ayazi: The HARPSS process for fabrication of nano-precision silicon electromechanical resonators. In: *Proc. 2001 1st IEEE Conf. Nanotechnol., Maui* (2001) pp. 489–494
3.55 N. Yazdi, F. Ayazi, K. Najafi: Micromachined inertial sensors, Proc. IEEE **86**, 1640–1659 (1998)
3.56 G. Timp: *Nanotechnology* (Springer, New York 1998)

3.57 S. Khan, L. Lorenzelli, R.S. Dahiya: Technologies for printing sensors and electronics over large flexible substrates: A review, IEEE Sens. J. **15**(6), 3164–3185 (2015) doi:10.1109/JSEN.2014.2375203

3.58 P. Rai-Choudhury (Ed.): *Handbook of Microlithography, Micromachining and Microfabrication. Vol. 1: Microlithography* (SPIE/IEE, Bellingham, Washington/London 1997)

3.59 L. Ming, C. Bao-qin, Y. Tian-Chun, Q. He, X. Qiuxia: The sub-micron fabrication technology. In: *Proc. 6th Int. Conf. Solid-State Integr.-Circuit Technol., San Francisco*, Vol. 1 (2001) pp. 452–455

3.60 M.A. Herman: *Molecular Beam Epitaxy: Fundamentals and Current Status* (Springer, New York 1996)

3.61 J.S. Frood, G.J. Davis, W.T. Tsang: *Chemical Beam Epitaxy and Related Techniques* (Wiley, New York 1997)

3.62 S. Mahajan, K.S. Sree Harsha: *Principles of Growth and Processing of Semiconductors* (McGraw Hill, New York 1999)

3.63 S. Kim, M. Razegi: Advances in quantum dot structures. In: *Processing and Properties of Compound Semiconductors*, ed. by K. Willardson, E.R. Weber (Academic, New York 2001)

3.64 D. Bimberg, M. Grundmann, N.N. Ledentsov: *Quantum Dot Heterostructures* (Wiley, New York 1999)

3.65 G. Seebohm, H.G. Craighead: Lithography and patterning for nanostructure fabrication. In: *Quantum Semiconductor Devices and Technologies*, ed. by T.P. Pearsall (Kluwer, Boston 2000)

3.66 E. Kapon: Lateral patterning of quantum well heterostructures by growth on nonplanar substrates. In: *Epitaxial Microstructures*, ed. by A.C. Gossard (Academic, New York 1994)

3.67 F. Guffarth, R. Heitz, A. Schliwa, O. Stier, N.N. Ledentsov, A.R. Kovsh, V.M. Ustinov, D. Bimberg: Strain engineering of self-organized InAs quantum dots, Phys. Rev. B **64**, 085305-1–085305-7 (2001)

3.68 M. Sugawara: *Self-Assembled InGaAs/GaAs Quantum Dots* (Academic, New York 1999)

3.69 B.C. Lee, S.D. Lin, C.P. Lee, H.M. Lee, J.C. Wu, K.W. Sun: Selective growth of single InAs quantum dots using strain engineering, Appl. Phys. Lett. **80**, 326–328 (2002)

3.70 F.S.S. Chien, W.F. Hsieh, S. Gwo, A.E. Vladar, J.A. Dagata: Silicon nanostructures fabricated by scanning-probe oxidation and tetra-methyl ammonium hydroxide etching, J. Appl. Phys. **91**, 10044–10050 (2002)

3.71 M. Calleja, J. Anguita, R. Garcia, K. Birkelund, F. Perez-Murano, J.A. Dagata: Nanometer-scale oxidation of silicon surfaces by dynamic force microscopy: Reproducibility, kinetics and nanofabrication, Nanotechnology **10**, 34–38 (1999)

3.72 E.S. Snow, P.M. Campbell, F.K. Perkins: Nanofabrication with proximal probes, Proc. IEEE **85**, 601–611 (1997)

3.73 H. Sugimura, T. Uchida, N. Kitamura, H. Masuhara: Tip-induced anodization of titanium surfaces by scanning tunneling microscopy: A humidity effect on nanolithography, Appl. Phys. Lett. **63**, 1288–1290 (1993)

3.74 N. Kramer, J. Jorritsma, H. Birk, C. Schonenberger: Nanometer lithography on silicon and hydrogenated amorphous silicon with low energy electrons, J. Vac. Sci. Technol. B **13**, 805–811 (1995)

3.75 H.T. Soh, K.W. Guarini, C.F. Quate: *Scanning Probe Lithography* (Kluwer Academic, Boston 2001)

3.76 C.A. Mirkin: Dip-pen nanolithography: automated fabrication of custom multicomponent, sub-100-nanometer surface architectures, MRS Bull. **26**, 535–538 (2001)

3.77 P.E. Sheehan, L.J. Whitman, W.P. King, B.A. Nelson: Nanoscale deposition of solid inks via thermal dip pen nanolithography, Appl. Phys. Lett. **85**, 1589–1591 (2004)

3.78 R. Garcia, A.W. Knoll, E. Riedo: Advanced scanning probe lithography, Nat. Nanotechnol. **9**, 577–587 (2014)

3.79 L.L. Sohn, R.L. Willett: Fabrication of nanostructures using atomic-force-microscope-based lithography, Appl. Phys. Lett. **67**, 1552–1554 (1995)

3.80 H.J. Mamin, B.D. Terris, L.S. Fan, S. Hoen, R.C. Barrett, D. Rugar: High-density data storage using proximal probe techniques, IBM J. Res. Dev. **39**, 681–699 (1995)

3.81 K. Bessho, S. Hashimoto: Fabricating nanoscale structures on Au surface with scanning tunneling microscope, Appl. Phys. Lett. **65**, 2142–2144 (1994)

3.82 I.W. Lyo, P. Avouris: Field-induced nanometer- to atomic-scale manipulation of silicon surfaces with the STM, Science **253**, 173–176 (1991)

3.83 M.F. Crommie, C.P. Lutz, D.M. Eigler: Confinement of electrons to quantum corrals on a metal surface, Science **262**, 218–220 (1993)

3.84 A. de Lozanne: Pattern generation below 0.1 micron by localized chemical vapor deposition with the scanning tunneling microscope, Jpn. J. Appl. Phys. **33**, 7090–7093 (1994)

3.85 L.A. Nagahara, T. Thundat, S.M. Lindsay: Nanolithography on semiconductor surfaces under an etching solution, Appl. Phys. Lett. **57**, 270–272 (1990)

3.86 T. Thundat, L.A. Nagahara, S.M. Lindsay: Scanning tunneling microscopy studies of semiconductor electrochemistry, J. Vac. Sci. Technol. A **8**, 539–543 (1990)

3.87 S.C. Minne, S.R. Manalis, A. Atalar, C.F. Quate: Independent parallel lithography using the atomic force microscope, J. Vac. Sci. Technol. B **14**, 2456–2461 (1996)

3.88 M. Lutwyche, C. Andreoli, G. Binnig, J. Brugger, U. Drechsler, W. Häberle, H. Rohrer, H. Rothuizen, P. Vettiger: Microfabrication and parallel operation of 5×5 2-D AFM cantilever arrays for data storage and imaging, Proc. MEMS **98**, 8–11 (1998)

3.89 G.M. Whitesides, B. Grzybowski: Self-assembly at all scales, Science **295**, 2418–2421 (2002)

3.90 P. Kazmaier, N. Chopra: Bridging size scales with self-assembling supramolecular materials, MRS Bull. **25**, 30–35 (2000)

3.91 G.A. Gelves, Z.T.M. Murakami, M.J. Krantz, J.A. Haber: Multigram synthesis of copper nanowires using ac electrodeposition into porous aluminium oxide templates, J. Mater. Chem. **16**, 3075–3083 (2006)

3.92 M.A. Morris: Directed self-assembly of block copolymers for nanocircuitry fabrication, Microelectron. Eng. **132**(25), 207–217 (2015) doi:10.1016/j.mee.2014.08.009

3.93 R. Gronheid, P. Nealey (Eds.): *Directed Self-Assembly of Block Copolymers for Nanomanufacturing*, Woodhead Publishing Series in Electronic and Optical Materials, Vol. 83 (Elsevier, Cambridge 2015)

3.94 R. Plass, J.A. Last, N.C. Bartelt, G.L. Kellogg: Self-assembled domain patterns, Nature **412**, 875 (2001)

3.95 Y.A. Vlasov, X.-Z. Bo, J.G. Sturm, D.J. Norris: On-chip natural self-assembly of silicon photonic bandgap crystals, Nature **414**, 289–293 (2001)

3.96 C. Gigault, K. Dalnoki-Veress, J.R. Dutcher: Changes in the morphology of self-assembled polystyrene microsphere monolayers produced by annealing, J. Colloid Interface Sci. **243**, 143–155 (2001)

3.97 J.C. Hulteen, C.R. Martin: A general template-based method for the preparation of nanomaterials, J. Mater. Chem. **7**, 1075–1087 (1997)

3.98 J.D. Joannopoulos, P.R. Villeneuve, S. Fan: Photonic crystals: putting a new twist on light, Nature **386**, 143–149 (1997)

3.99 T.D. Clark, R. Ferrigno, J. Tien, K.E. Paul, G.M. Whitesides: Template-directed self-assembly of 10-μm-sized hexagonal plates, J. Am. Chem. Soc. **124**, 5419–5426 (2002)

3.100 S.A. Sapp, D.T. Mitchell, C.R. Martin: Using template-synthesized micro- and nanowires as building blocks for self-assembly of supramolecular architectures, Chem. Mater. **11**, 1183–1185 (1999)

3.101 Y. Yin, Y. Lu, B. Gates, Y. Xia: Template assisted self-assembly: A practical route to complex aggregates of monodispersed colloids with well-defined sizes, shapes and structures, J. Am. Chem. Soc. **123**, 8718–8729 (2001)

3.102 J.L. Wilbur, G.M. Whitesides: Self-assembly and self-assembles monolayers in micro- and nanofabrication. In: *Nanotechnology*, ed. by G. Timp (Springer, New York 1999)

3.103 S.R. Wasserman, Y.T. Tao, G.M. Whitesides: Structure and reactivity of alkylsiloxane monolayers formed by reaction of alkyltrichlorosilanes on silicon substrates, Langmuir **5**, 1074–1087 (1989)

3.104 C.P. Tripp, M.L. Hair: An infrared study of the reaction of octadecyltrichlorosilane with silica, Langmuir **8**, 1120–1126 (1992)

3.105 S. Fiorilli, P. Rivolo, E. Descrovi, C. Ricciardi, L. Pasquardini, L. Lunelli, L. Vanzetti, C. Pederzolli, B. Onida, E. Garrone: Vapor-phase self-assembled monolayers of aminosilane on plasma-activated silicon substrates, J. Colloid Interface Sci. **321**, 235–241 (2008)

3.106 R. de la Rica, A. Baldi, E. Mendoza, A. San Paulo, A. Llobera, C. Fernandez-Sanchez: Silane nanopatterns via gas-phase soft lithography, Small **4**(8), 1076–1079 (2008)

3.107 D.R. Walt: Nanomaterials: Top-to-bottom functional design, Nature **1**, 17–18 (2002)

3.108 J. Noh, T. Murase, K. Nakajima, H. Lee, M. Hara: Nanoscopic investigation of the self-assembly processes of dialkyl disulfides and dialkyl sulfides on Au(111), J. Phys. Chem. B **104**, 7411–7416 (2000)

3.109 M. Himmelhaus, F. Eisert, M. Buck, M. Grunze: Self-assembly of n-alkanethiol monolayers: A study by IR-visible sum frequency spectroscopy (SFG), J. Phys. Chem. **104**, 576–584 (1999)

3.110 A.K. Boal, F. Ilhan, J.E. DeRouchey, T. Thurn-Albrecht, T.P. Russell, V.M. Rotello: Self-assembly of nanoparticles into structures spherical and network aggregates, Nature **404**, 746–748 (2000)

3.111 A. Ulman: *An Introduction to Ultrathin Organic Films: From Langmuir–Blodgett to Self-Assembly* (Academic, New York 1991)

3.112 E. Winfree, F. Liu, L.A. Wenzler, N.C. Seeman: Design and self-assembly of two-dimensional DNA crystals, Nature **394**, 539–544 (1998)

3.113 J.H. Reif, T.H. LaBean, N.C. Seeman: Programmable assembly at the molecular scale: Self-assembly of DNA lattices. In: *Proc. 2001 IEEE Int. Conf. Robot. Autom., Seoul* (2001) pp. 966–971

3.114 A.P. Alivisatos, K.P. Johnsson, X. Peng, T.E. Wilson, C.J. Loweth, M.P. Bruchez Jr, P.G. Schultz: Organization of 'nanocrystal molecules' using DNA, Nature **382**, 609–611 (1996)

3.115 H. Cao, Z. Yu, J. Wang, J.O. Tegenfeldt, R.H. Austin, E. Chen, W. Wu, S.Y. Chou: Fabrication of 10 nm enclosed nanofluidic channels, Appl. Phys. Lett. **81**, 174–176 (2002)

3.116 H. Masuda, H. Yamada, M. Satoh, H. Asoh: Highly ordered nanochannel-array architecture in anodic alumina, Appl. Phys. Lett. **71**, 2770–2772 (1997)

3.117 R.L. Fleischer: *Nuclear Tracks in Solids: Principles and Applications* (Univ. of California Press, Berkeley 1976)

3.118 R.E. Packard, J.P. Pekola, P.B. Price, R.N.R. Spohr, K.H. Westmacott, Y.Q. Zhu: Manufacture observation and test of membranes with locatable single pores, Rev. Sci. Instrum. **57**, 1654–1660 (1986)

3.119 L. Sun, P.C. Searson, C.L. Chien: Electrochemical deposition of nickel nanowire arrays in single-crystal mica films, Appl. Phys. Lett. **74**, 2803–2805 (1999)

3.120 E.C. Walter, M.P. Zach, F. Favier, B.J. Murray, K. Inazu, J.C. Hemminger, R.M. Penner: Metal nanowire arrays by electrodeposition, ChemPhysChem **4**(2), 131–138 (2003)

3.121 M.Z. Atashbar, D. Banerji, S. Singamaneni, V. Bliznyuk: Deposition of parallel arrays of palladium nanowires and electrical characterization using microelectrode contacts, Nanotechnology **15**(3), 374–378 (2004)

3.122 Y. Du, W.L. Cai, C.M. Mo, J. Chen, L.D. Zhang, X.G. Zhu: Preparation and photoluminescence of alumina membranes with ordered pore arrays, Appl. Phys. Lett. **74**, 2951–2953 (1999)

3.123 M. Guowen, C. Anyuan, C. Ju-Yin, A. Vijayaraghavan, J.J. Yung, M. Shima, P.M. Ajayan: Ordered Ni nanowire tip arrays sticking out of the an-

odic aluminum oxide template, J. Appl. Phys. **97**, 64303–1–64303–5 (2005)
3.124 S. Yang, H. Zhu, D. Yu, Z. Jin, S. Tang, Y. Du: Preparation and magnetic property of Fe nanowire array, J. Magn. Magn. Mater. **222**, 97–100 (2000)
3.125 M. Sun, G. Zangari, R.M. Metzger: Cobalt island arrays with in-plane anisotropy electrodeposited in highly ordered alumina, IEEE Trans. Magn. **36**, 3005–3008 (2000)
3.126 P.V. Braun, P. Wiltzius: Electrochemically grown photonic crystals, Nature **402**, 603–604 (1999)
3.127 S.R. Nicewarner-Pena, R.G. Freeman, B.D. Reiss, L. He, D.J. Pena, I.D. Walton, R. Cromer, C.D. Keating, M.J. Natan: Submicrometer metallic barcodes, Science **294**, 137–141 (2001)

4. 3-D Nanostructure Fabrication by Focused-Ion Beam, Electron- and Laser Beam

Shinji Matsui, Hiroaki Misawa, Quan Sun

In this chapter, we describe three-dimensional (3-D) nanostructure fabrication techniques using focused-ion-beam (FIB)-induced chemical vapor deposition (CVD), electron-beam (EB)-induced CVD, and femtosecond laser (fs-laser) techniques. We first describe 30 keV Ga^+ FIB-induced CVD using a *phenanthrene* ($C_{14}H_{10}$) source gas as the precursor. A diamond-like amorphous carbon film is deposited during this process; it has a Young's modulus exceeding 600 GPa, making it potentially highly desirable for various applications. A three-dimensional pattern generator system has been developed to make arbitrary three-dimensional nanostructures. We also discuss microstructure plastic art, which is a new field that has been made possible by microbeam technology, and we present examples of such art, including a *micro wine glass* with an external diameter of 2.75 μm and a height of 12 μm. We then discuss free-space nanowiring and show by using a mixture of $C_{14}H_{10}$ and $W(CO)_6$ that the electrical properties indicate an increase in metal content results in a lower resistivity. We also demonstrate that a *Morpho* butterfly scale quasistructure fabricated by FIB-induced CVD has almost the same optical characteristics as a real *Morpho* butterfly scale. We then discuss three-dimensional nanostructure fabrication using EB-induced CVD. Because of the nanometer resolution, EB-induced CVD is now indispensable for mask repair techniques for the 193 nm node. According to real-time observations by transmission electron microscopy, the W clusters, as the initial growth stage, are formed first followed by the W layer which forms as W clusters coalesce due to EB irradiation. We go on to discuss photonic crystals and Smith–Purcell electron optics as examples of three-dimensional nanostructure applications using EB-induced CVD. Finally, we describe femtosecond-laser-assisted micro/nanofabrication which has been recognized as a promising technique to fabricate three-dimensional structures inside transparent materials. The spatial resolution can reach submicrometer levels and even tens of nanometers owing to suppression of the involved heat diffusion and nonlinear adsorption. We discuss three-dimensional femtosecond laser nanofabrication using the direct laser writing technique and multiple beam interference lithography and describe the fabrication of photonic crystals in a photoresist.

4.1 **Various 3-D Nanostructure Fabrication Techniques** 88
4.2 **3-D Nanostructure Fabrication by Focused Ion Beam** 88
4.2.1 Focused-Ion-Beam-Induced Chemical Vapor Deposition 88
4.2.2 3-D Fabrication Process 89
4.2.3 Mechanical Properties 91
4.2.4 Free-Space Nanowiring 93
4.2.5 Electrical Properties 95
4.2.6 Brilliant Blue Observation from a *Morpho* Butterfly Scale Quasistructure 96
4.3 **3-D Nanostructure Fabrication by Electron Beam** 97
4.3.1 Electron-Beam-Induced Chemical Vapor Deposition 97
4.3.2 3-D Nanofabrication by Electron-Beam-Induced CVD 102
4.4 **3-D Nanostructure Fabrication by Laser** 103
4.4.1 3-D Nanostructure Fabrication by Femtosecond Laser 103
4.4.2 3-D Nanodevices: Fabrication of Photonic Crystals 105
References 109

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_4

4.1 Various 3-D Nanostructure Fabrication Techniques

Recent years have witnessed a number of investigations concerning nanostructure technology. The objective of research on nanostructure technology is to explore the basic physics, technology, and applications of ultrasmall structures and devices with dimensions in the sub-10 nm regime. Electron-beam (EB) lithography is the most widely used and versatile lithography tool for fabricating nanostructure devices and photo masks. EB can focus to diameters of less than 5 nm. The minimum beam diameter of the scanning electron microscope (SEM) and scanning transmission electron microscope (STEM) are 1.5 and 0.5 nm, respectively. While the focused ion beam (FIB) can be focused close to 5 nm. EB and FIB have been used to fabricate various two-dimensional (2-D) nanostructure devices such as single-electron transistors and MOS (metal oxide semiconductor) transistors with nanometer gate-lengths. 10 nm structures can be formed by using a commercially available EB or FIB system with 5−10 nm-diameter beams and a high-resolution resist [4.1]. 2-D nanostructure fabrication is therefore already an established process. Conventional devices have been fabricated by a 2-D processing technique. Developing a three-dimensional (3-D) processing technique presents the possibility of realizing a device with higher functionality. In this chapter, we discuss 3-D nanostructure fabrications using EB-induced chemical vapor deposition (CVD), FIB-induced CVD, and femtosecond laser (fs-laser) techniques.

Beam-induced chemical vapor deposition (CVD) is widely used in the electrical device industry for the repair of chips and masks. This type of deposition is mainly performed on 2-D pattern features. Work has been done on 3-D fabrication by laser, EB, or FIB using beam-induced CVD. In 1990, *Wagner* et al. demonstrated pillars and walls with high aspect ratios achieved using FIB-induced CVD [4.2]. In 1991, *Lehman* et al. demonstrated 3-D microfabrication by laser-induced CVD [4.3]. In 1994, *Koops* et al. demonstrated some applications such as an AFM (atomic force microscopy) tip and a field emitter that were realized with EB-induced CVD [4.4]. In 2000, *Matsui* et al. demonstrated 3-D nanostructure fabrication using FIB-induced CVD [4.5]. FIB- and EB-induced CVD are superior to laser-induced CVD in terms of spatial resolution and beam-scan control. To demonstrate 3-D nanostructure fabrication using a laser, in 1999, *Sun* et al. reported on the fs-laser two-photon-absorption photopolymerization of a 3-D photonic crystal (PhC) structure in a resist [4.6].

4.2 3-D Nanostructure Fabrication by Focused Ion Beam

4.2.1 Focused-Ion-Beam-Induced Chemical Vapor Deposition

Two commercially available FIB systems (SMI9200, SMI2050, SII Nanotechnology Inc., Tokyo, Japan) with a Ga^+ ion beam operating at 30 keV were used to study 3-D FIB-induced CVD. The FIB-induced CVD used a *phenanthrene* ($C_{14}H_{10}$) precursor as the source material. The beam diameter of SMI9200 was about 7 nm and that of SMI2050 was about 5 nm. The SMI9200 system was equipped with two gas sources in order to increase the gas pressure. The nozzles faced each other and were directed at the beam point. The nozzles were set a distance of 40 μm from each other and positioned about 300 μm above the substrate surface. The inside diameter of a nozzle was 0.3 mm. The *phenanthrene* gas pressure during pillar growth was typically 5×10^{-5} Pa in the specimen chamber, but the local gas pressure at the beam point was expected to be much higher. The crucible for the source was heated to 85 °C. Conversely, the SMI2050 system was equipped with a single gas nozzle. The FIB is scanned in order to be able to write the desired pattern via computer control, and the ion dose is adjusted to deposit a film of the desired thickness. The experiments were carried out at room temperature on a silicon substrate.

The deposited film was characterized by observing it with a transmission electron microscope (TEM) and by analyzing its Raman spectra [4.5]. A thin film of carbon (200 nm thick) was deposited on a silicon substrate by 30 keV Ga^+ FIB using a *phenanthrene* precursor gas. The cross-sections of the structures created and its electron diffraction patterns were observed by using a 300 kV TEM. There were no crystal structures in the TEM images and diffraction patterns. It was therefore concluded that the deposited film was amorphous carbon (a-C). Raman spectra of the a-C films were measured at room temperature with the 514.5 nm line of an argon ion laser. The Raman spectra were recorded using a monochromator equipped with a CCD (charge coupled device) multichannel detector. Raman spectra were measured at 0.1−1.0 mW to avoid thermal decomposition of the samples. A relatively sharp Raman band at 1550 cm^{-1} and a broad-shouldered band at around

$1400\,cm^{-1}$ were observed in the spectra excited by the 514.5 nm line. Two Raman bands were plotted after Gaussian line shape analysis. These Raman bands, located at 1550 and $1400\,cm^{-1}$, originate from the trigonal (sp^2) bonding structure of graphite and tetrahedral (sp^3) bonding structure of diamond. This result suggests that the a-C film deposited by FIB-induced CVD is diamond-like amorphous carbon (DLC), which has attracted attention due to its hardness, chemical inertness, and optical transparency.

The coordination of carbon atoms in the carbon-based material formed by CVD of *phenanthrene* assisted by Ga^+ FIB was investigated by the measurement of near-edge X-ray absorption fine structure spectra of the carbon *K*-edge over the excitation energy range 275–320 eV [4.7]. A novel peak observed at 289.0 eV was assigned to the $1\,s \rightarrow \sigma^*$ transition of carbon neighboring the residual Ga. The material formed by this method was found to be Ga-doped DLC which consists of a high sp^3 hybridized carbon. The atomic fraction of the FIB-induced CVD DLC film has been determined as $C : Ga : H = 87.4 : 3.6 : 9.0$ at.% using Rutherford backscattering spectrometry and elastic recoil detection analysis. The hydrogen content of the FIB-induced CVD DLC film was relatively speaking lower than that of DLC films formed by other CVD methods [4.8].

4.2.2 3-D Fabrication Process

A 3-D structure usually contains overhang structures and hollows. Gradual position scanning of the ion-beam during the CVD process causes the position of the growth region around the beam point to shift. When the beam point reaches the edge of the wall, secondary electrons appear at the side of the wall and just below the top surface. The DLC then starts to grow laterally; the width of the vertical growth is also about 80 nm. Therefore, combining the lateral growth mode with rotating beam scanning, it is possible to obtain 3-D structures with rotational symmetry like a wine glass.

The process of fabricating 3-D structures by FIB-induced CVD is illustrated in Fig. 4.1 [4.5]. In FIB-induced CVD processes, the beam is scanned in digital mode. First, a pillar is formed on the substrate by fixing the beam position (position 1). After that, the beam position is moved to within a diameter of the pillar (position 2) and then fixed until the deposited terrace thickness exceeds the range of the ions (a few 10 nm). This process is repeated to make 3-D structures. The key point in making 3-D structures is to adjust the beam scan speed so that the ion beam remains within the deposited terrace, which means that the terrace thickness always exceeds the range of the ions. Growth in the *x*- and *y*-directions is controlled by both beam deflectors. The growth in the *z*-direction is determined by the deposition rate; that is, the height of the structure is proportional to the irradiation time when a deposition rate is constant.

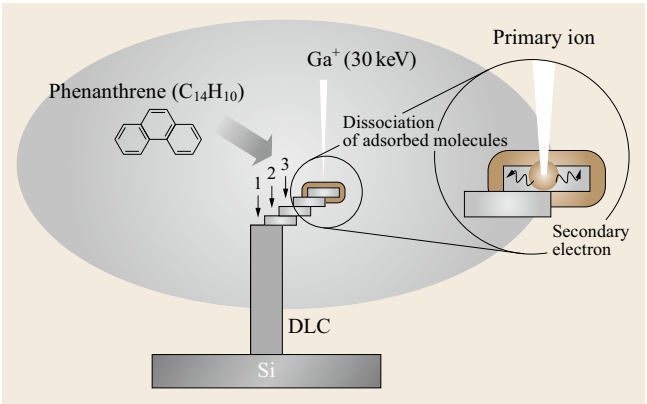


Fig. 4.1 Fabrication process for a three-dimensional nanostructure by FIB-induced CVD

A *micro wine glass* was created on a human hair as a work of microstructure plastic art as shown in the scanning ion microscope (SIM) image in Fig. 4.2. A micro wine glass with an external diameter of 2.75 µm and a height of 12 µm was created [4.5]. The fabrication time was 600 s with a beam current of

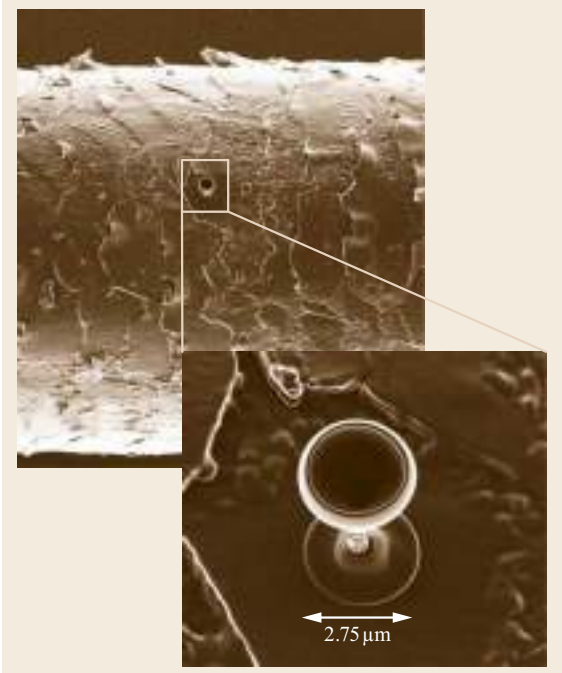


Fig. 4.2 Micro wine glass with an external diameter of 2.75 µm and a height of 12 µm on a human hair

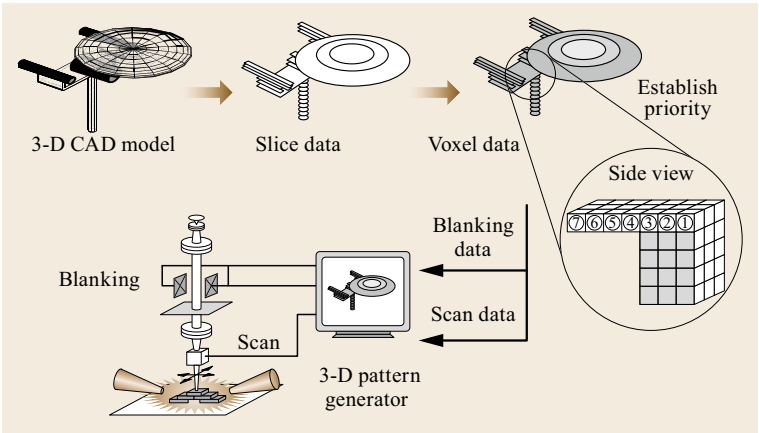


Fig. 4.3 Data flow of 3-D pattern-generating system for FIB-induced CVD

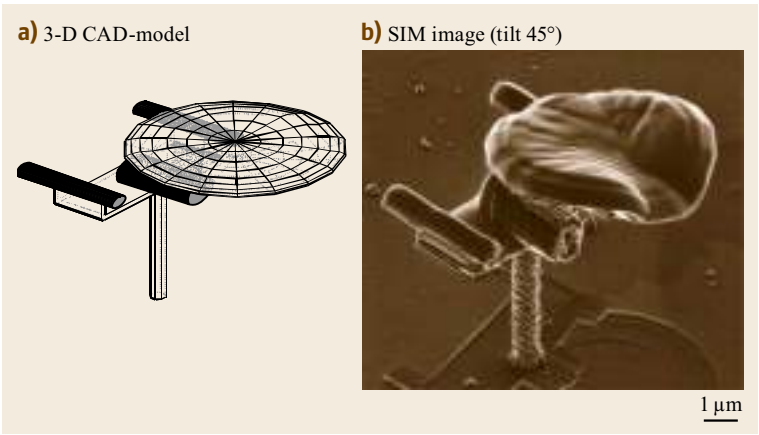


Fig. 4.4a,b *Micro starship Enterprise NCC-1701D*, 8.8 µm long

16 pA. This beautiful micro wine glass shows the potential of the field of microstructure plastic art.

A 3-D pattern generator system was developed to make arbitrary 3-D nanostructures [4.9]. A 3-D multilayer structure is built-up. In the first step of this 3-D pattern-generating system, a 3-D model of the structure, designed using a 3-D computer-aided design (CAD) system (3-D DXF format), is needed. In this case we realized a structure that was shaped like a pendulum. The 3-D CAD model, which is a surface model, is cut into several slices, as shown in Fig. 4.3. The thickness of the slices depends upon the resolution in the z-direction (the vertical direction). The x and y coordinates of the slices are then used to create the scan data (voxel data). To fabricate the overhanging structure, the ion beam must irradiate the correct positions in the correct order. If the ion beam irradiates a voxel located in mid-air without a support layer, the ions intended for the voxel will be deposited on the substrate. Therefore, the sequence of irradiation is determined, as shown in Fig. 4.3. The scan data and blanking signal therefore include the scan sequence, the dwell time, the interval time, and the irradiation pitch. These parameters are calculated from the beam diameter, x-y resolution and z resolution of fabrication. The z resolution is proportional to the dwell time and inversely proportional to the square of the irradiation pitch. The scan data are passed to the beam deflector of the FIB-induced CVD as are the blanking data. The blanking signal controls the dwell time and interval time of the ion beam. Figure 4.4 shows a 3-D CAD model and a SIM image of the starship Enterprise NCC-1701D (from the television series Star Trek) which was fabricated by FIB-induced CVD at $10 \approx 20$ pA [4.9]. The nanospaceship is 8.8 µm long and was realized at about a 1 : 100 000 000 scale on a silicon substrate. The dwell time, interval time, irradiation pitch, and total process time were 80 µs, 150 µs, 2.4 nm, and 2.5 h, respectively. The horizontal overhang structure was successfully fabricated. Figure 4.5 shows a *nano T4 bacteriophage*, which is an artificial version of the virus fabricated by FIB-induced CVD on a silicon surface. The size of the artificial nano *T4 bacteriophage* is about 10 times that of the real virus.



Fig. 4.5 T-4 bacteriophage

To fabricate an arbitrary 3-D nanostructure, it is necessary to understand the growth mechanism that occurs during FIB-induced CVD. With the aim of elucidating these details, in situ observation of 3-D nanostructure growth during FIB-induced CVD using SEM in a FIB–SEM dual-beam system was carried out. As a result, it was demonstrated that the 3-D nanostructure growth depended on the beam profile of the FIB as well as on the FIB irradiation time and scanning speed [4.10].

4.2.3 Mechanical Properties

An evaluation of the mechanical characteristics of such nanostructures is required for material physics. *Buks* and *Roukes* reported a simple but useful technique for measuring the resonant frequencies of nanoscale objects using an SEM [4.11]. The secondary electron detector in the SEM can detect frequencies up to around 4 MHz, so the sample vibration is measured as the oscillatory output signal of the detector. This technique was used to evaluate the Casimir attractive force between the two parallel beams fabricated at a nanoscale. We evaluated the mechanical characteristics of DLC pillars in terms of the Young's modulus, determined using resonant vibration and the SEM monitoring technique [4.12, 13].

The system set-up for monitoring mechanical vibration is shown in Fig. 4.6b. There were two ways of measuring the pillar vibrations. One was active measurement, where the mechanical vibration was induced by a thin piezoelectric device, 300 μm thick and 3 mm square. The piezo device was bonded to the sidewall of the SEM's sample holder with silver paste. The sample holder was designed to observe cross-sections in the SEM (S5000, Hitachi) system. Therefore, the pillar's vibration was observed as a side-view image, as shown in Fig. 4.6a. The range of vibration frequencies involved were 10 kHz up to 2 MHz, which is much faster than the SEM raster scanning speed. Thus, the resonant vibrations of the pillars can be taken as the trace of the pillar's vibration in the SEM image. The resonant frequency and amplitude were controlled by adjusting the power of the driving oscillator.

The other way to measure pillar vibrations is passive measurement using a spectrum analyzer (Agilent, 4395A), where most of the vibration seemed to derive from environmental noise due to rotary pumps and air conditioners. Some parts of the vibration result from the spontaneous vibration associated with thermal excitations [4.11]. Because of the excitation and residual

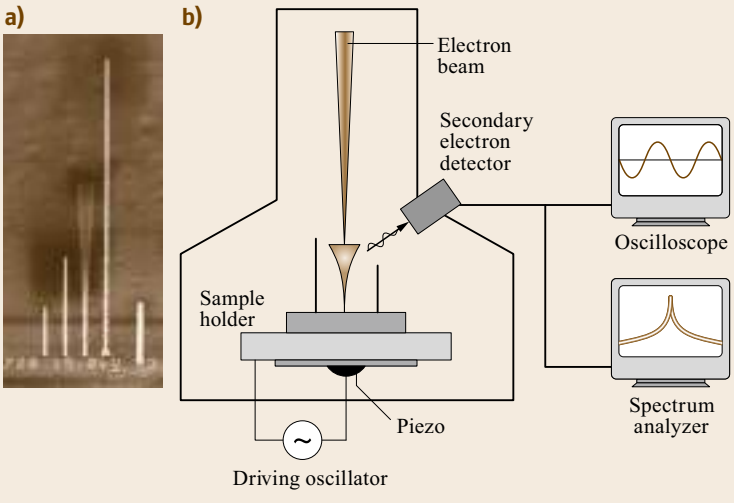


Fig. 4.6 **(a)** SEM image of the vibration. The resonant frequency was 1.21 MHz. **(b)** Schematic diagram of the vibration monitoring system

noise, the pillars on the SEM sample holder always vibrated at a fundamental frequency, even if noise isolation is enforced on the SEM system. The amplitude of these spontaneous vibrations was on the order of a few nanometers at the top of the pillar and high-resolution SEM can easily detect it at a magnification of 300 000.

We arranged several pillars that had varying diameters and lengths. The DLC pillars with the smallest diameter of 80 nm were grown using point irradiation. While we used two FIB systems for pillar fabrication, slight differences in the beam diameters of the two systems did not affect the diameters of the pillars. Larger diameter pillars were fabricated using an area-limited raster scan mode. Raster scanning a 160 nm^2 region produced a pillar with a cross-section of about 240 nm^2 and a 400 nm^2 scan resulted in a pillar with a cross-section of 480 nm^2. The typical SEM image taken during resonance is shown in Fig. 4.6a. The FIB-induced CVD pillars seemed very durable against mechanical vibration. This kind of measurement usually requires at least 30 min, including a spectrum analysis and photorecording, but the pillars still survived without any change in resonance characteristics. This durability of the DLC pillars should be useful in nanomechanical applications.

The resonant frequency f of the pillar is defined by (4.1) for a pillar with a square cross-section and (4.2) for a pillar with a circular cross-section

$$f_{\text{square}} = \frac{a\beta^2}{2\pi L^2}\sqrt{\frac{E}{12\rho}} \tag{4.1}$$

$$f_{\text{circular}} = \frac{a\beta^2}{2\pi L^2}\sqrt{\frac{E}{16\rho}}\,, \tag{4.2}$$

where a is the width of the square pillar and/or the diameter of the circular-shaped pillar, L is the length of the pillar, ρ is the density, and E is the Young's modulus. The coefficient β defines the resonant mode and $\beta = 1.875$ for the fundamental mode. We used (4.1) for pillars that were 240 nm and 480 nm wide and (4.2) for pillars grown by point-beam irradiations. The relationship of the resonant frequency to the Young's modulus, which depends on the ratio of the pillar diameter to the squared length, is summarized in Fig. 4.7. All of the pillars evaluated in this figure were fabricated using the SMI9200 FIB system under rapid growth conditions. Typical growth rates were about 3–5 μm/min for the 100 nm diameter and 240 μm wide pillars and 0.9 μm/min for the 480 nm wide pillars. When calculating the data shown in Fig. 4.7, we assumed that the density of the DLC pillars was about 2.3 g/cm^3, which is almost identical to that of graphite. The inclination of the line in Fig. 4.7 indicated the Young's modulus

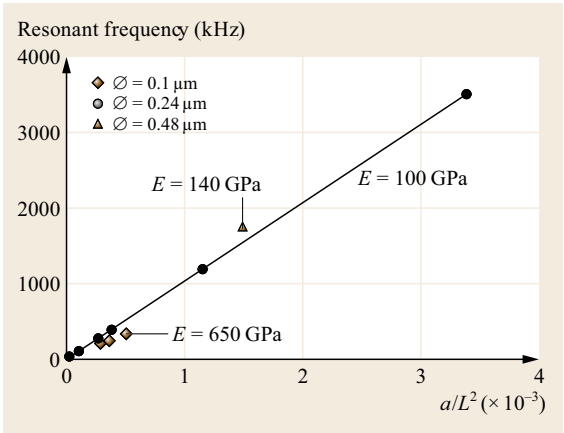


Fig. 4.7 Dependence of resonant frequency on the pillar length

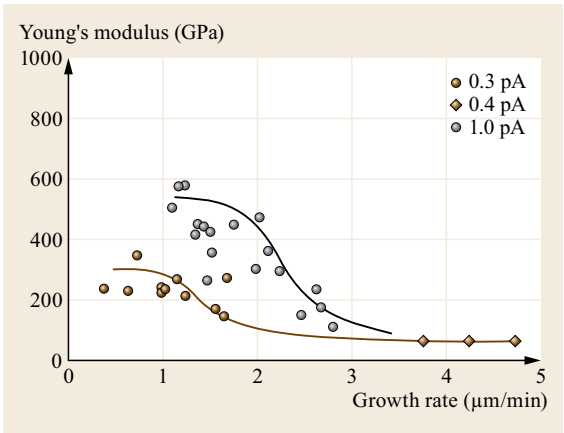


Fig. 4.8 Dependence of Young's modulus on the growth rate

for each pillar. The Young's moduli of the pillars were distributed over a range from 65–140 GPa, which is almost identical to that of normal metals. Wider pillars tended to have larger Young's moduli.

Figure 4.8 shows that the stiffness increases significantly as the local gas pressure decreases. While the absolute value of the local gas pressure at the beam point is very difficult to determine, we found that the growth rate can be a useful parameter for describing the dependence of pressure on Young's modulus. All data points indicated in Fig. 4.8 were obtained from pillars grown using point irradiation. Therefore, the pillar diameters did vary slightly from 100 nm but not by more than 5%. A relatively low gas pressure, with a good uniformity, was obtained by using a single gas nozzle and gas reflector. We use a cleaved side wall of Si tips as the gas reflector, which was placed 10–50 μm away from the beam point so as to be facing the gas

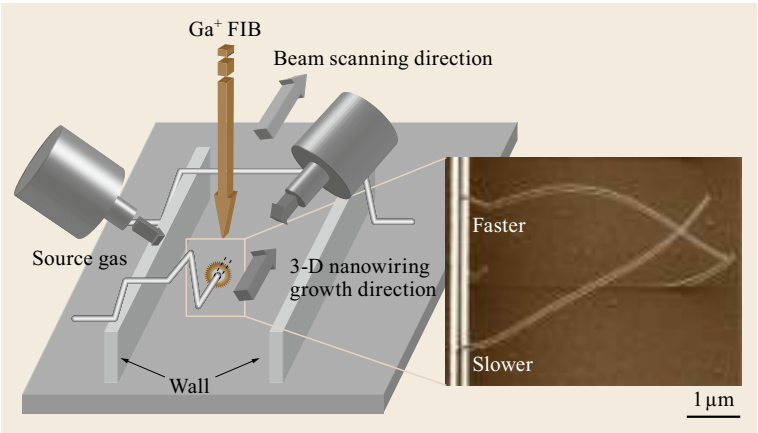


Fig. 4.9 Fabrication of DLC free-space wiring using both FIB-induced CVD and CPG

nozzle. The growth rate was controlled by changing the distance to the wall. While there is a large distribution of data points, the stiffness of the pillar tended to become greater as the growth rate decreased. The two curves in Fig. 4.8 represented data points obtained for a beam current of 0.3 pA (*open circles*) and 1 pA (*solid circles*), respectively. Both curves show the same tendency; the saturated upper levels of the Young's modulus are different for each ion current at low gas pressure (low growth rate). It should be noted that a few pillars' Young's moduli exceeded 600 GPa, which is of the same order as that of tungsten carbide. In addition, these estimations assume a pillar density of 2.3 g/cm^3, but a finite amount of Ga was incorporated with the pillar growth. If the calculation takes the increase in pillar density due to the Ga concentration into account, the Young's modulus exceeds 800 GPa. Such a high Young's modulus reaches that of carbon nanotubes and natural diamond crystals. We think that this high Young's modulus is due to surface modification caused by the direct ion impact. In contrast, when the gas pressure was high enough to achieve a growth rate of more than 3 µm/min, the pillars became soft but the change in the Young's modulus was small. The uniformity of the Young's modulus (as seen in Fig. 4.7) presumably results from the fact that the growth occurred in this insensitive region, where the low levels of source gas limit pillar growth.

Various 3-D mechanical parts can be made with this useful, high Young' modulus. *Kometani* et al. reported DLC electrostatic nanoactuators and bio-nanotools such as injectors and manipulators [4.14–18]. Nakamatsu et al. reported a DLC nanospring with a 130 nm spring section diameter, which showed behavior similar to that of a macroscale spring. The estimated shear modulus of the DLC nanospring was about 70 GPa. This value is very close to that of a conventional coil spring made of steel [4.19].

4.2.4 Free-Space Nanowiring

All experiments were carried out in a commercially available FIB system (SMI9200:SII NanoTechnology Inc.) using a beam of 30 kV Ga^+ ions. The beam was focused to a spot size of 7 nm at a beam current of 0.4 pA and it was incident perpendicular to the surface. The pattern-drawing system (compute pattern generator, CPG-1000: Crestec Co., Tokyo, Japan) was added to the FIB apparatus to draw any patterns. Using the CPG, it is possible to control beam scan parameters such as scanning speed, *x*–*y* direction, and blanking of the beam, and so 3-D free-space nanowiring can be performed [4.20].

Figure 4.9 illustrates the free-space nanowiring fabrication process using both FIB-induced CVD and CPG. When *phenanthrene* ($C_{14}H_{10}$) gas or tungsten hexacarbonyl [$W(CO)_6$] gas, which is a reactive organic gas, is evaporated from a heated container and injected into the vacuum chamber by a nozzle located 300 µm above the sample surface at an angle of about 45° with respect to the surface, the gas density of the $C_{14}H_{10}$ or $W(CO)_6$ molecules increases on a substrate near the gas nozzle. The nozzle system creates a local high-pressure region over the surface. The base pressure of the sample chamber is 2×10^{-5} Pa and the chamber pressure upon introducing $C_{14}H_{10}$ and $W(CO)_6$ as a source gas was 1×10^{-4} and 1.5×10^{-3} Pa, respectively. If a Ga^+ ion beam is irradiated onto the substrate, $C_{14}H_{10}$ or $W(CO)_6$ molecules adsorbed on the substrate surface are decomposed and deposited onto the surface of the substrate. The direction of deposition growth can be controlled through the scanning direction of the beam. The material deposited using $C_{14}H_{10}$ gas was DLC, as confirmed by Raman spectra and it had a very large Young's modulus of 600 GPa [4.5, 12].

After the two walls shown in Fig. 4.9 were formed, free-space nanowiring was performed by adjusting the

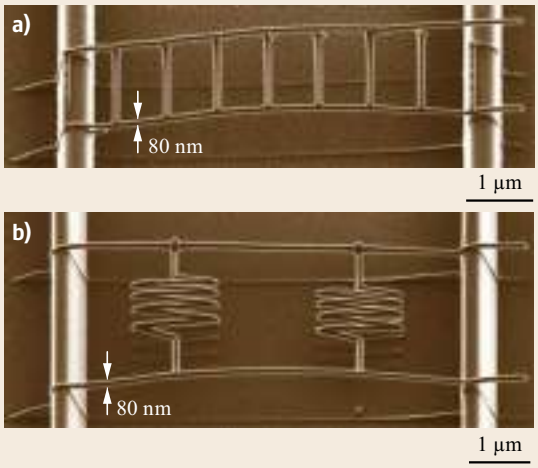


Fig. 4.10 (a) DLC free-space wiring with a bridge shape. (b) DLC free-space wiring with parallel resistances

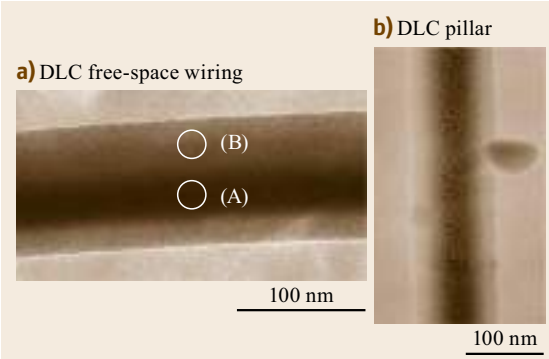


Fig. 4.11a,b TEM images of (a) DLC free-space wiring and (b) DLC pillar

beam scanning speed. The ion beam used was a 30 kV Ga^+ FIB and the irradiation current was 0.8−2.3 pA. The x and y scanning directions and the beam scanning speed were controlled by the CPG. The height in the z-direction was proportional to the irradiation time. Deposition is made to occur horizontally by scanning the beam at a certain fixed speed in the direction of a plane. However, if the beam scanning speed is faster than the nanowiring growth speed, it grows downward or drops. Conversely, if the scanning speed is too slow, the deposition grows slanting upward. Therefore, it is very important to carefully control the beam scanning speed when growing a nanowire horizontally. It turns out that the optimal beam scanning speed to realize a nanowire growing horizontally, using two $C_{14}H_{10}$ gas guns, is about 190 nm/s. The expected pattern resolution for the FIB-induced CVD is around 80 nm, because both the primary Ga ion and secondary-electron scattering occur over distances around 20 nm [4.12, 21].

Figure 4.10 shows examples of free-space nanowiring fabricated by FIB-induced CVD and CPG. All of the structures shown were fabricated using $C_{14}H_{10}$ gas as a precursor gas. Figure 4.10a shows nanobridge free-space wirings. The growth time was 1.8 min and the wiring width was 80 nm. Figure 4.10b shows free-space nanowires with parallel resistances. The growth time was 2.8 min and the wiring width was again 80 nm.

The free-space nanowiring structures were observed using 200 keV TEM. The analyzed area was 20 nm Φ. Figure 4.11a,b show TEM images of DLC free-space wiring and a pillar. It became clear from these EDX (energy dispersive X-ray) measurements that the dark part (A) of Fig. 4.11a corresponds to the Ga core and the outside part (B) of Fig. 4.11a corresponds to amorphous carbon. This free-space nanowiring therefore consists of amorphous carbon with a Ga core. The center position of the Ga core is actually located below the center of the wiring. However, in the case of the DLC pillar, the Ga core is located at the center of the pillar. To investigate the difference between these core positions, the Ga core distribution in free-space nanowiring was observed in detail by TEM. The center position of the Ga core was about 70 nm from the top, which was 20 nm below the center of the free-space nanowiring. We calculated an ion range of 30 kV Ga^+ ions into amorphous carbon, using transport of ions in matter (TRIM), of 20 nm. The calculation indicates that the displacement of the center of the Ga core in the nanowiring corresponds to the ion range.

Kometani et al. proposed a technology to fabricate ultralong horizontal free-space nanowires by using the growth angle dependency of the specimen current and carrying out feedback control of the scanning speed of the Ga^+ FIB [4.22, 23]. The feedback control system monitors the specimen current and controls the FIB scanning speed in real-time while judging the value of the specimen current. In order to verify the effect of real-time feedback control, FIB was scanned laterally 15 μm with a uniform scanning speed (80 nm/s), from the tip of a nanopillar as shown in Fig. 4.12. As Fig. 4.12a shows, the nanowire turned downward after growing approximately 2 μm due to the change in growth characteristics. With the change in nanowire growth angle, the generation of secondary electrons also changed and thus, the specimen current changed during growth, as shown in Fig. 4.12b. Conversely, as shown in Fig. 4.12c, the effect of feedback control was obvious in that the nanowire grew almost horizontally and arrived at its goal length of 15 μm without turning downward. During growth with feedback control, the specimen current remained in the permission range,

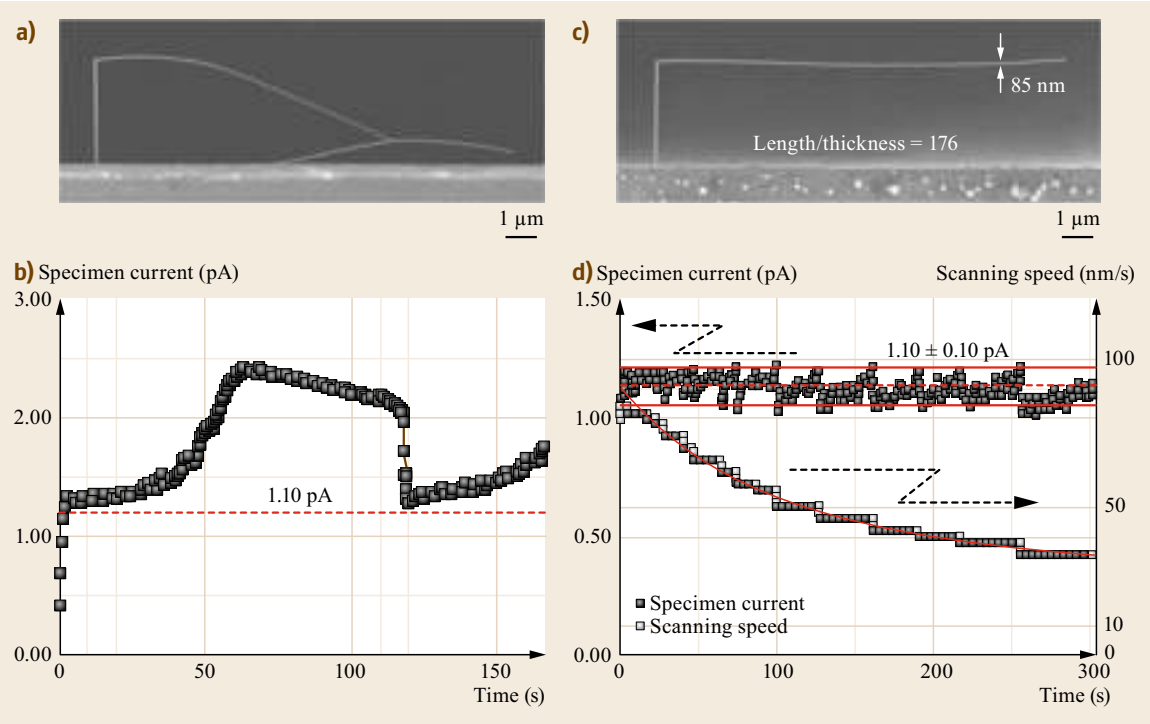


Fig. 4.12a–d Growth results for a 15 μm long free-space nanowire. **(a)** Nanowire grown without feedback control, **(b)** change in specimen current without feedback control, **(c)** nanowire grown with feedback control, and **(d)** changes in specimen current and scanning speed of the Ga^+ FIB with feedback control. (Reprinted with permission from [4.22]. Copyright 2013, American Vacuum Society)

as shown in Fig. 4.12d. Furthermore, an ultralong horizontal free-space nanowire with a length of more than 30 μm and a length-to-thickness ratio as high as 300 was achieved.

Electron holography is a useful technology for direct observation of electrical and magnetic fields at nanoscale and can also provide useful information by detecting the phase shift of the electron wave due to the electrical and magnetic field. The technique requires an electron biprism, which plays the important role of dividing the electron wave into a reference wave and an objective wave. The biprism is composed of one thin filament and two ground electrodes. It is important to fabricate a filament that is as narrow as possible to obtain an interference fringe with a high contrast and good fringe quality. However, fabricating the filament with a diameter below 500 nm is very difficult, because a conventional electron biprism is fabricated by pulling a melted glass rod by hand. To overcome this problem, we introduced a new electron biprism fabrication technique using FIB-induced CVD and evaluated the characteristics of the new biprism. The 80 nm DLC thin wiring works as the filament of the biprism. The diameter and length of the filament are 80 nm and 15 μm, respectively. The results demonstrate the adequacy of the thin filament fabricated by FIB-induced CVD and the new biprism is very useful for accurate observation with a high contrast and good fringe quality in electron holography [4.24].

4.2.5 Electrical Properties

The electrical properties of free-space nanowiring fabricated by FIB-induced CVD using a mixture of $C_{14}H_{10}$ and $W(CO)_6$ were measured [4.20]. Nanowires were fabricated on a Au electrode. These Au electrodes were formed on a 0.2 μm thick SiO_2-on-Si substrate by an EB lithography and lift-off process. The two-terminal electrode method was used to measure the electrical resistivity of the nanowiring. The electrical resistivity (Fig. 4.13a) fabricated using only a $C_{14}H_{10}$ source gas, was 1×10^2 Ωcm. The elemental contents were 90% C and 10% Ga, as measured by an SEM-EDX spot beam. The *I*–*V* curves (b), (c), and (d) correspond to increasing $W(CO)_6$ content in the gas mixture. The *I*–*V* curves (b)–(d) show that as the $W(CO)_6$ content increases, the W content increases. Moreover, the Ga content also increased because nanowiring growth slowed; the irra-

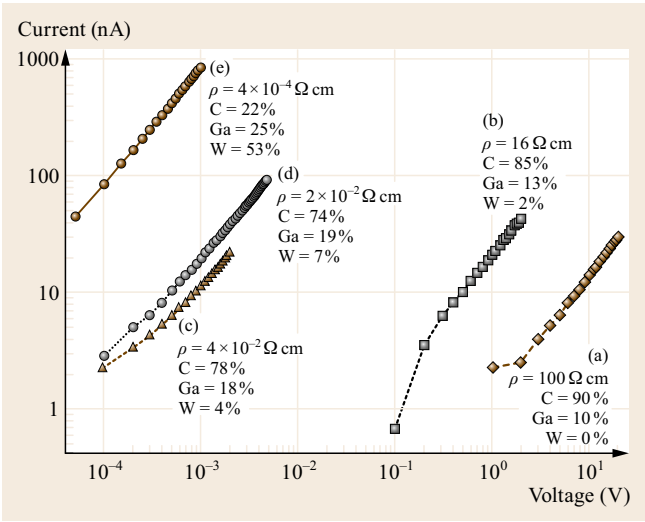


Fig. 4.13 Electrical resistivity measurement for nanowiring. Electrical resistivity ρ was calculated using I–V curves. Elemental contents C, Ga, and W were measured by SEM-EDX

diation time of the Ga FIB became longer. The electrical resistivities of the *I–V* curves (b), (c), and (d) were 16×10^{-2}, 4×10^{-2}, and $2\times10^{-2}\,\Omega$cm, respectively. The electrical resistivity of (e), which was fabricated using only a $W(CO)_6$ source gas, was $4\times10^{-4}\,\Omega$cm. Increasing the Ga and W metallic content decreases the electrical resistivity, as shown by the SEM-EDX measurements in Fig. 4.13. These results indicate that increasing the metallic content results in lower resistivity.

Kometani et al. demonstrated a piezo resistive effect in a DLC nanostructure [4.25] and superconductivity with a T_c as high as 5.8 K in tungsten-carbide nanowires deposited from the precursor mixture of $W(CO)_6$ and $C_{14}H_{10}$ [4.26]. Furthermore, a W−C superconductor–normal metal–superconductor (SNS) Josephson junction was fabricated and a Fraunhofer-like oscillation of critical current in the junction with four periods was observed [4.27].

4.2.6 Brilliant Blue Observation from a *Morpho* Butterfly Scale Quasistructure

The *Morpho* butterfly has brilliant blue wings and the source of this intense color has been an interesting topic of debate for a long time. Because of an intriguing optical phenomenon, the scales reflect an interfered brilliant blue color for any angle of incidence of white light. This color is called a structural color, meaning that it is not caused by pigment reflection [4.28]. When we observed the scales with an SEM (Fig. 4.14a), we found 3-D nanostructures 2 μm in height, 0.7 μm in width, and with a 0.22 μm grating pitch on the scales. These nanostructures cause a similar optical phenomenon to the iridescence produced by a jewel beetle.

We duplicated the *Morpho* butterfly scale quasistructure with a commercially available FIB system (SMI9200: SII Nanotechnology Inc.) using a Ga^+ ion beam operating at 30 kV [4.29]. The beam diameter was about 7 nm at 0.4 pA. The FIB-induced CVD was performed using *phenanthrene* as a precursor. In this experiment, we used a computer-controlled pattern generator, which converted 3-D CAD data into a scanning signal, which was passed to an FIB scanning apparatus in order to fabricate a 3-D nanostructure [4.9]. The scattering range of the Ga^+ primary ions is about 20 nm and the range of the secondary electrons induced by the Ga^+ ion beam is about 20 nm, so the expected pattern resolution of the FIB-induced CVD is about 80 nm. Figure 4.14b shows an SIM image of the *Morpho* butterfly quasistructure fabricated by FIB-induced CVD using the 3-D CAD data. This result demonstrates that

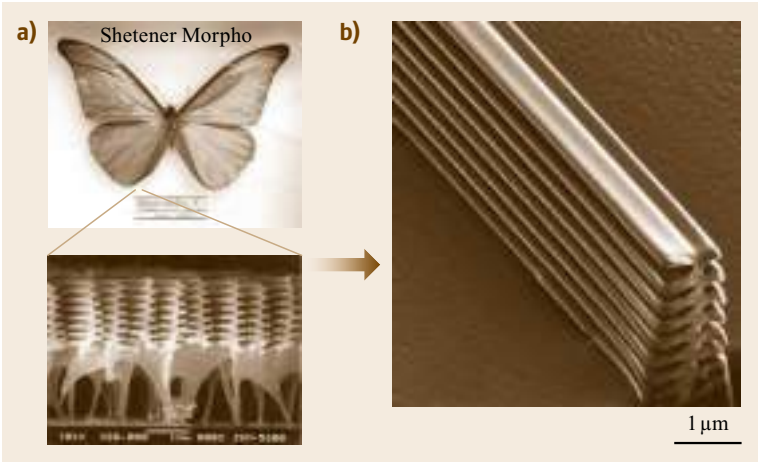


Fig. 4.14a,b *Morpho* butterfly scales. **(a)** Optical microscope image showing top view of *Morpho* butterfly. SEM image showing a cross-sectional view of *Morpho* butterfly scales. **(b)** SIM images showing inclined view of *Morpho* butterfly scale quasistructure fabricated by FIB-induced CVD

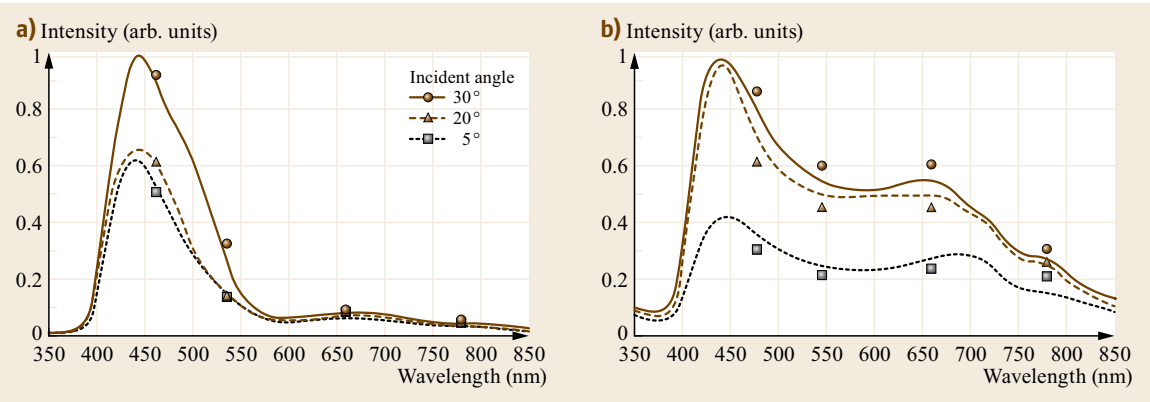


Fig. 4.15a,b Intensity curves of the reflection spectra for **(a)** *Morpho* butterfly scales. **(b)** *Morpho* butterfly scale quasistructure

FIB-induced CVD can be used to fabricate such a quasistructure.

We measured the reflection intensities from *Morpho* butterfly scales and the *Morpho* butterfly scale quasistructure optically; white light from a halogen lamp was directed onto a sample with angles of incidence ranging from 5 to 45°. The reflection was concentrated by an optical microscope and analyzed using a commercially available photonic multichannel spectral analyzer system (PMA-11: Hamamatsu Photonics K.K., Hamamatsu City, Japan). The intensity of light incident from the halogen lamp had a peak at a wavelength close to 630 nm.

The *Morpho* butterfly scale quasistructure was made of DLC. The reflectivity and transmittance of a 200 nm thick DLC film deposited by FIB-induced CVD, measured by the optical measurement system at a wavelength close to 440 nm (the reflection peak wavelength of the *Morpho* butterfly), were 30 and 60%, respectively. Therefore, the measured data indicated that the DLC film had high reflectivity near 440 nm, which is important for the fabrication of an accurate *Morpho* butterfly scale quasistructure. We measured the reflection intensities of the *Morpho* butterfly scales and the quasistructure with an optical measurement system and compared their characteristics. Figure 4.15a,b respectively show the reflection intensities from *Morpho* butterfly scales and the quasistructure. Both gave a peak intensity near 440 nm and showed very similar reflection intensity spectra for various angles of incidence. We have thus successfully demonstrated that a *Morpho* butterfly scale quasistructure fabricated using FIB-induced CVD has almost the same optical characteristics as real *Morpho* butterfly scales.

4.3 3-D Nanostructure Fabrication by Electron Beam

4.3.1 Electron-Beam-Induced Chemical Vapor Deposition

When an EB is irradiated onto a substrate, source gas molecules adsorbed on the substrate are dissociated into nonvolatile and volatile materials by an EB-induced surface reaction. Nonvolatile materials are deposited on the substrate, while volatile materials are evacuated. Nanostructures can be fabricated by EB-induced CVD, because beam diameters, as small as 1 nm, can be formed with an SEM or STEM. Computer-controlled direct writing is also possible for EB-induced CVD. Metals, semiconductors, and inorganic metals can be deposited in this way.

In 1976, *Broers* et al. were able to deposit nanometer-scale carbon contamination structures in a STEM with electrons of 45 keV energy [4.30]. In 1984, *Matsui* and *Mori* obtained the first metal-containing deposits by introducing a metal-organic gas ($Cr(C_6H_6)_2$) vapor into the vacuum chamber of an SEM [4.31]. In the over 30 years since the first experiments, many metal-containing structures have been deposited by EB-induced CVD under a wide range of experimental conditions [4.32–48]. The first metal-containing 3-D nanostructures were deposited in 1986 by *Koops* et al. [4.33], using an SEM.

Several excellent reviews have been published in recent years [4.49–51]. In addition, the principle, spatial resolution, mechanical properties, and magnetic and electric properties of EB-induced CVD have been reported [4.52–55]. Furthermore, various device applications including electrodes for carbon nanotube devices,

plasmonic gold nanostructures, domain wall conduit behavior in cobalt nanowires, optical monopole antenna and metamaterials have been demonstrated [4.56–61]. Its nanometer resolution has led to EB-induced CVD being indispensable for mask repair techniques for the 193 nm node [4.62].

In situ Observation of Electron-Beam-Induced CVD by Auger Electron Spectroscopy

When source gas molecules adsorbed on a substrate are dissociated by EB irradiation, Auger electrons are emitted from deposited materials at the same time. Therefore, the growth process of the deposited material can be observed in situ by Auger electron spectroscopy (AES) [4.63]. Figure 4.16 shows an in situ AES observation system for EB-induced surface reactions. This system was constructed by modifying a conventional AES apparatus. A source chamber and mass flow controller (MFC), which controls the source gas flow rate, were added to the conventional AES apparatus. The system was evacuated by an ion pump (IP) with a 2×10^{-9} Torr base pressure after bake-out. This system also contains an EB gun, a cylindrical mirror analyzer (CMA), and an Ar ion gun. The source gas is introduced through the MFC and the source gas molecules are adsorbed onto the substrate. The source gas molecules adsorbed onto the substrate are dissociated by a primary EB from the AES system and nonvolatile materials are deposited onto the substrate. At the same time, Auger electrons are emitted from the deposited material surface. Therefore, the growth process of the deposited materials can be observed in situ by detecting Auger electrons using the CMA. In this experiment, in situ AES observation has been carried out for W deposition using a WF_6 source gas on a Si(100) substrate. The EB voltage and current were 10 keV and 1 μA, respectively. This experiment was carried out by point beam exposure. The minimum beam diameter of the AES apparatus is ≈ 1 μm. However, a ≈ 30 μm beam diameter, which was obtained by defocusing, was applied to avoid beam profile effects and to obtain a uniformly deposited film. Therefore, the beam current density was $\approx 0.14\,A/cm^2$. The EB was irradiated at room temperature. It was estimated from calculations that the temperature increase on the sample during EB exposure was less than 10 °C.

Figure 4.17 shows Auger intensity variations during EB irradiation at 3.5×10^{-7} Torr WF_6 source gas pressure. W Auger intensity increases exponentially as the EB exposure time increases. W Auger intensity saturates at a 10 min exposure time, because the deposited W film thickness reaches the escape depth. Conversely, Si Auger intensity decreases, because the deposited W film was formed on an Si substrate. The contents (at.%) calculated from the AES spectrum, as shown in Fig. 4.17, were 85% W, 7.5% F, and 7.5% O at 15 min exposure time. It is believed that O originates from residual oxygen in the source and sample chambers. If an ultrahigh vacuum of the order of 10^{-10} Torr is achieved in both chambers, a higher quality film will be deposited. A growth rate for EB-induced CVD has been calculated on the assumption of layer growth. The calculated growth rate at 3.5×10^{-7} Torr is 0.15 nm/min.

In situ Observation of Electron-Beam-Induced CVD by TEM

The growth process of the deposited material can be observed in situ by TEM, if WF_6 source gas molecules are adsorbed on a TEM specimen [4.64]. The EM-002A

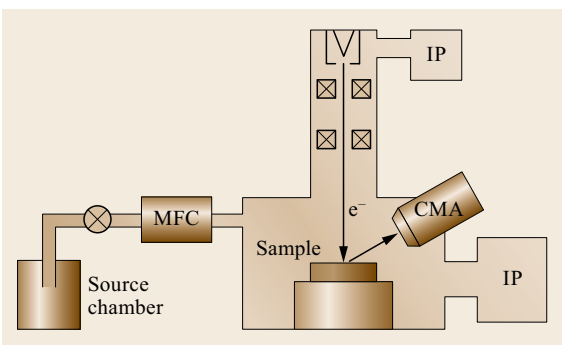


Fig. 4.16 Experimental AES system to perform in situ observation of the EB-induced CVD process. IP: ion pump

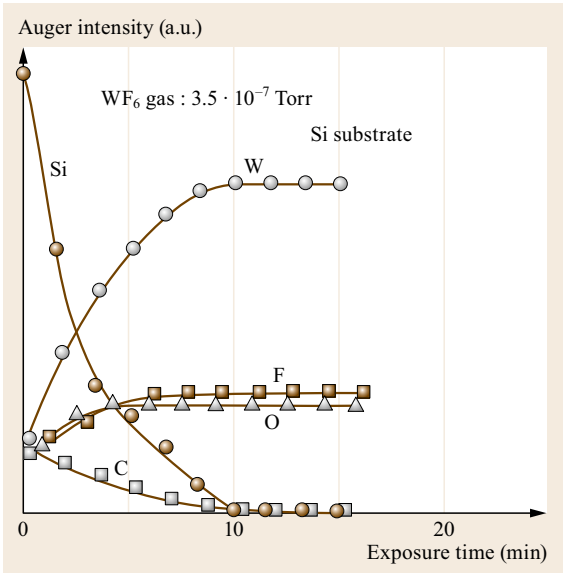


Fig. 4.17 Relations between Auger intensity variations and electron exposure time for W deposition on a Si substrate

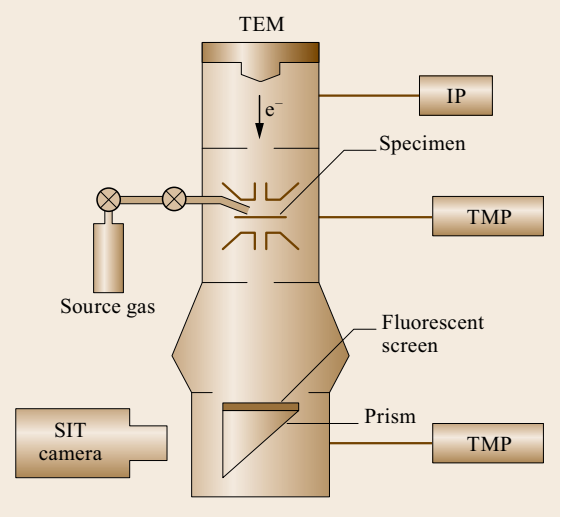


Fig. 4.18 Experimental arrangement diagram of TEM in situ observation. The gas-introducing system and VTR system are included. SIT: Silicon Intensifier Target, TMP: turbo-molecular pumps

(Akashi Beam Technology Corp.) electron microscope was employed, equipped with a real-time TV monitoring system, specially improved for in situ observation as shown in Fig. 4.18. The microscope resolution was 0.23 nm at 120 keV, which allowed the imaging of individual rows of atom columns in W crystals. To reduce problems in regard to specimen contamination, the instrument was operated with an ultrahigh vacuum pumping system with turbo-molecular pumps (TMP) and ion pumps. The gas injection tube had a 3 mm inner diameter and was about 5 mm from the sample surface. Fine and spherical Si particles, about 60 nm in diameter, were used as TEM specimens. The small particles were made by a gas evaporation method in argon under a reduced atmosphere, where the Si vapor was condensed into small particles. They were less than 100 nm in diameter and were usually covered with 1–3 nm thick SiO_2. The gas molecules were adsorbed on the fine Si particles, which were on the TEM specimen grid. The gas molecules were excited by the TEM's EB and dissociated into W metal and F_2 gas. W metal was deposited on the Si surface and growth occurred. The growth process was observed in situ at a single atom resolution level by an electron microscope equipped with a TV monitoring system. All micrographs were reproduced from single frames of the video tape recorder (VTR) tape. Two types of experiments were performed in this experiment. First, the EB was irradiated on a WF_6 adlayer formed on the fine Si particle surface, in order to clarify an initial growth process of EB-induced CVD. Second, the EB was irradiated on the Si fine particle surface while WF_6 was flowing on the surface, to study the layer growth process of EB-induced CVD.

Figure 4.19 shows a typical series of electron micrographs of in situ TEM observation variations during EB irradiation for the WF_6 adlayer. These micrographs were selected from a VTR tape which ran over a 30 min

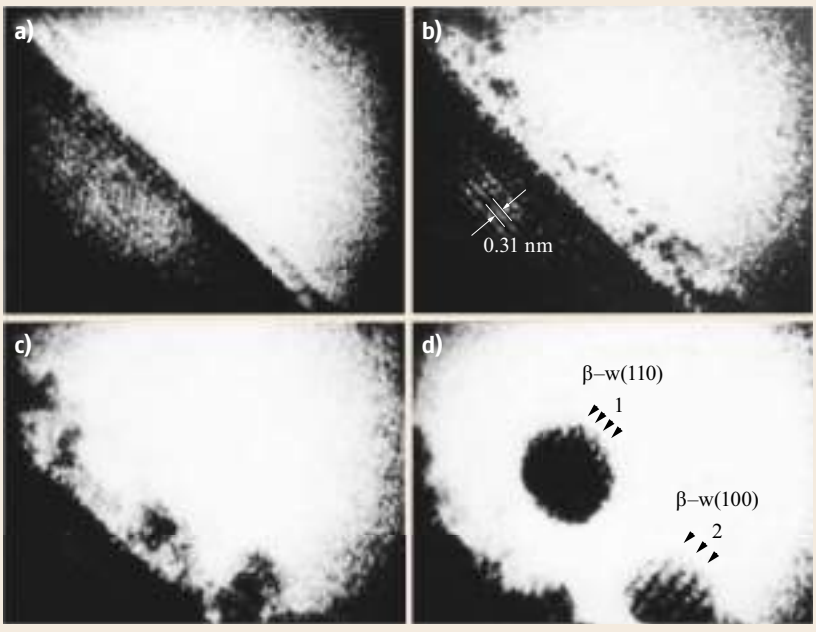


Fig. 4.19a–d EB exposure time dependence of W cluster growth on a Si particle. Exposure times: **(a)** 0 min, **(b)** 3 min, **(c)** 15 min, and **(d)** 30 min

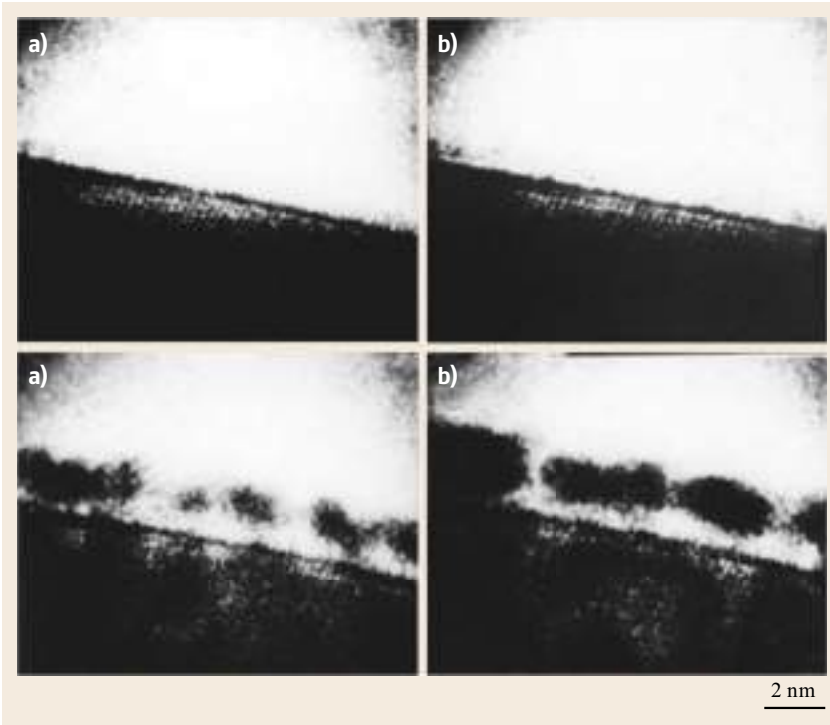


Fig. 4.20a–d EB exposure time dependence of W layer growth on fine Si particles at 5×10^{-7} Torr WF_6 gas pressure. Exposure times: **(a)** 0 s, **(b)** 15 s, **(c)** 25 s, and **(d)** 30 s

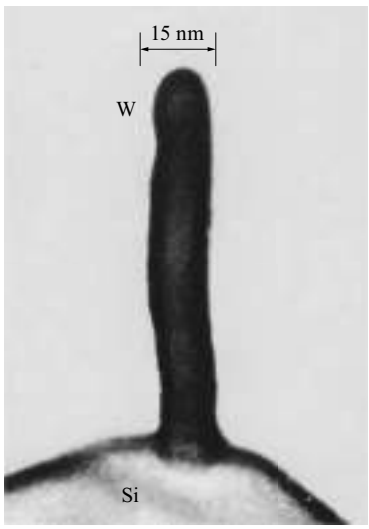


Fig. 4.21 Electron micrograph for a W rod with 15 nm diameter. The W rod was fabricated using a 120 kV EB with 3 nm diameter

span. Electron beam irradiation times for Fig. 4.19a–d were 0, 3, 15, and 30 min, respectively. This result indicates that W atoms, dissociated by EB irradiation from the WF_6 adlayer, coalesced and grew under EB irradiation. According to the real-time observation on a TV screen, moving clusters often collided with each other, causing coalescence. Figure 4.20 shows TEM in situ observation variations during EB irradiation under 5×10^{-7} Torr WF_6 gas pressure. Electron beam irradiation times were 0, 15, 25, and 30 s for Fig. 4.20a–d, respectively. These micrographs indicate that the W clusters were formed as the initial growth state and then the W layer was formed by coalescence of the clusters due to EB irradiation. A nanometer-scale W rod was fabricated by a focused EB (3 nm diameter) moved manually under a 1×10^{-6} Torr WF_6 gas atmosphere. Electron beam energy was 120 keV and its current density was $\approx 100\,A/cm^2$. The moving speed of the beam was 1 nm/s. The EB was parallel to the substrate and its scanning direction was normal to the surface. An electron micrograph of the fabricated W rod is shown in Fig. 4.21. The rod's radius was ≈ 15 nm. The bright inner rod contrast shows that the center of the rod was probably tungsten silicide. The beam conditions for rod formation were not the same as those for taking electron micrographs. The EB was spread to about a 1 μm diameter for taking this photograph.

High Resolution of Electron-Beam-Induced CVD by STEM

Tanaka et al. reported on the resolution of EB-induced CVD by STEM [4.43]. The TEM instrument used in this study was a JEM-2000VF with a field emission gun and the base pressure of the chamber was 1×10^{-7} Pa. The deposition apparatus was attached to a vacuum port of the TEM column. It consisted of a variable leak valve, a gas nozzle ≈ 1 mm in diameter, and a heating system.

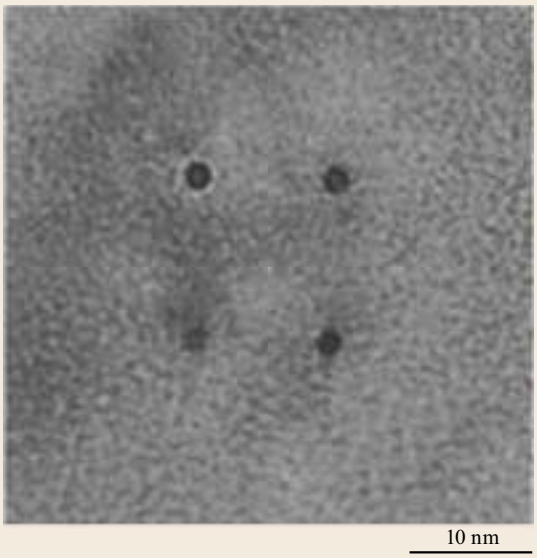


Fig. 4.22 TEM image of W nanodots deposited over deferent times. The *top left* dot was fabricated over 15 s, the ones on the *right* were fabricated over 10 s, and the *bottom left* nanodot was fabricated over 5 s. (Reprinted with permission from [4.43])

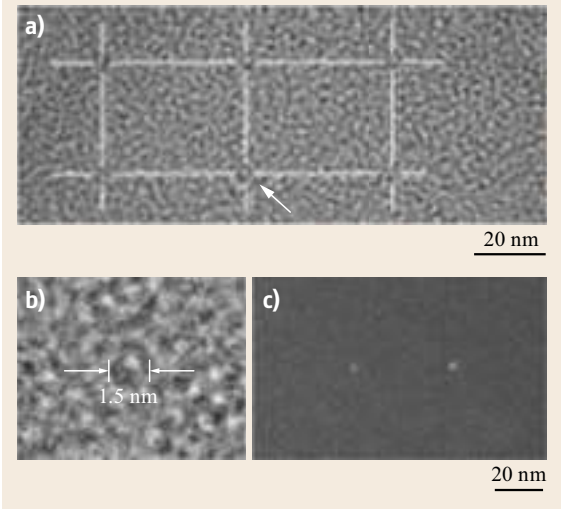


Fig. 4.23a–c An array of the smallest W nanodots ever achieved by EB-induced CVD: **(a)** TEM image of the array; **(b)** magnified image of one of the dots shown by the *arrow* in **(a)**; **(c)** the corresponding HAADF-STEM (High-angle annular dark-field-scanning transmission electron microscope) image. (Reprinted with permission from [4.43])

The substrates were Si(111), conventionally dimpled and ion-milled. Most of the EB-induced CVD experiments were performed on the thinnest parts of the substrate, mostly amorphized during the ion milling. The precursor gas was $W(CO)_6$ and the EB-induced CVD was carried out at room temperature. The microscope was operated at 200 kV. The EB intensity was $\approx 5 \times 10^3 - 5 \times 10^4$ A/cm^{-2} and the prove size was $\approx 1-2$ nm. By using a 2 nm probe, arrays of dots were fabricated successfully from $W(CO)_6$. Figure 4.22 shows a TEM image of one of the arrays. Irradiation times for dot manufacture were 15 s (*top left*), 10 s (*top right* and *bottom right*), and 5 s (*bottom left*), respectively. The pressure during the deposition was kept at 2×10^{-6} Pa. The distance between the dots in the horizontal direction was $\approx$ 12 nm and in the vertical direction was $\approx$ 14 nm. Their structure was shown to be mixed amorphous W-C as they contain carbon from the precursor gas. It was recognized that dot size decreased as irradiation time was shortened. Size measurements showed that the 15 s dot was $\approx$ 3 nm in diameter, the 10 s dots were $\approx$ 2.5 nm, and the 5 s dot was $\approx$ 2 nm, respectively.

Figure 4.23 shows an array of dots formed with a 1 nm probe and a deposition period of 5 s. In Fig. 4.23a, a TEM image of the array is shown. The dots are located at the intersection points of the white lines. They are so small that it is hard to distinguish them from the amorphous substrate. The size of the dot is $\approx$ 1.5 nm. To prove that they really were mixed tungsten nanodots, HAADF-STEM images were taken for the same substrate, which produces Z-contrast images, i. e., the image is atomic number sensitive in that a heavy atom makes a brighter contrast. Figure 4.23c shows one of these images. The distance between the dots and the size of the dots almost match those of the TEM images, thus indicating that they are the nanodots.



Fig. 4.24 This high-resolution TEM image proves a wall roughness of < 2 nm for a deposited rod if a platinum-containing precursor material is used. (Reprinted from [4.39] with permission from Elsevier)

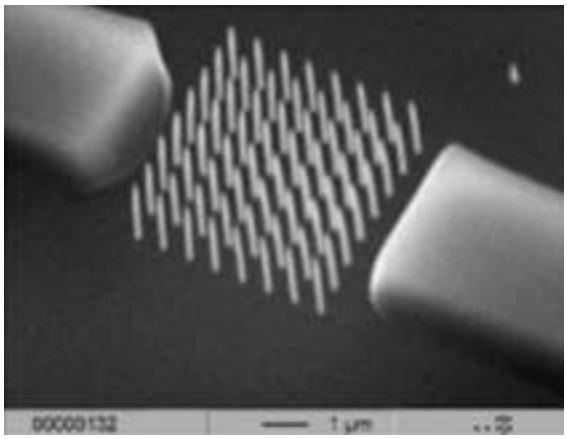


Fig. 4.25 A cubic photonic crystal is constructed in a prefabricated guide in a PMMA (Poly(methyl methacrylate)) waveguide for investigation of the transmission properties of the phonic crystal. (Reprinted from [4.39] with permission from Elsevier)

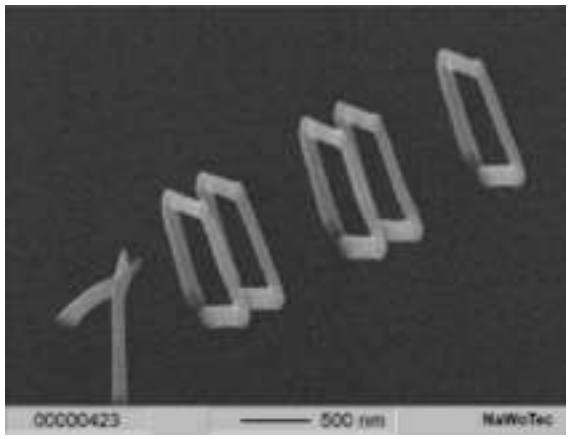


Fig. 4.26 Smith–Purcell electron optics realized with EB-induced CVD. (Reprinted from [4.40] with permission from Elsevier)

4.3.2 3-D Nanofabrication by Electron-Beam-Induced CVD

Koops et al. reported on 3-D nanostructure fabrication using EB-induced CVD. This technique is very well suited for rapid-prototyping of photonic crystals (PhCs) [4.39]. The nanolithography tool in use is based on a SEM (JSM 6300 F, JEOL Ltd.) with the beam control system VIDAS (Kontron). All possible precautions have been taken to compensate for thermal drifts, mechanical vibrations, and magnetic interference by external compensating systems. These measures allow high-precision fabrication of 2-D arrays of dielectric rods with the dimensions required for PhCs. A specially designed environmental chamber is used for supply of precursor material to the waveguide sample [4.37]. The EB-induced CVD process renders 2 nm edge roughness for the deposited rods. This is a surface quality which is unreachable by standard lithography or beam-assisted etching techniques. It allows one to minimize scattering of photons. The high-resolution TEM image shown in Figure 4.24 proves a wall roughness of < 2 nm for a rod grown by EB-induced CVD using cyclopentadienyl-platinum-trimethyl as the precursor material. Electron-beam-induced CVD constructs the rods from nanocrystals of 2 nm diameter embedded into a carbonaceous matrix. This results in an edge roughness of below 2 nm, which is an outstanding feature of this process, as demonstrated in Fig. 4.24.

PhCs with cubic, hexagonal, or honeycomb elementary cells are fabricated under computer control. A VIDAS macro language allows one to quickly adjust the design parameters of the crystals that are to be grown. There are no other process steps required to generate a crystal by this time-consuming growth process. Using 25 keV electrons and a 20 pA current, a crystal of 100 rods is grown in 40 min. Over this period, the sample in the microscope is not allowed to drift more than 20 nm. Figure 4.25 shows a cubic PhC, which is grown inside a prefabricated gap in a PMMA (Poly(methyl methacrylate)) waveguide for investigation of the transmission properties of the PhC. The pitch is 580 μm, the rod diameter is 150 nm, and the height is 2.5 μm. Characterization of PhCs and related devices is a very challenging task. This is because of the small size of the devices, having diameters of 3–6 λ, which is 4.5–9 μm at $\lambda = 1.5$ μm. Monomode fibers have a core diameter of 9 μm. Therefore, a well-designed optical beam converter is required to couple light into PhCs. To measure the polarization-dependent transmission of PhCs, a sophisticated metrology set-up was used. The transmission measurement did prove a band gap filter in the regime from 1250–1650 nm.

Electron-beam-induced CVD is a rapid prototyping technology with a high resolution and accuracy to investigate the activity and characteristics of PhCs and PhC devices. Recently, *Tasco* et al. reported on fabrication of 3-D nanohelices for chiral photonics by FIB- and EB-induced CVD [4.65]. The initial promising results prove the applicability of the technology and the optical action of the devices.

Electron-beam-induced CVD can build similar structures in sub-μm dimensions with wires of 80–200 nm. Figure 4.26 shows a 3-D build-up with conducting material resembling a field-emission electron source, with an emitter tip and an extractor ring around it [4.40]. The structure has a capacitance of 24 aF between the tip and the ring.

4.4 3-D Nanostructure Fabrication by Laser

4.4.1 3-D Nanostructure Fabrication by Femtosecond Laser

The advent of laser micro/nano fabrication can be traced back to the 1960s, soon after the invention of the laser. The observed photomodifications of laser rod materials and the formation of ripples on irradiated surfaces revealed the potential of laser structuring of materials. The development of the ultrafast fs-laser led to the development of the field of laser material processing, and fs-laser-assisted micro/nano fabrication has since been recognized as a promising technique to fabricate 2-D structures on the surface of materials or 3-D structures inside transparent materials [4.66, 67]. The spatial resolution of laser fabrication can reach submicrometer levels and even tens of nanometers owing to the suppression of the involved heat diffusion and nonlinear absorption.

Principles and Methods of Femtosecond Laser Nanofabrication

The mechanisms of fs-laser nanofabrication are based on the photomodification of materials. Photomodification can be categorized into different types such as optical damage or modification of the refractive index in transparent dielectric materials, ablation on the surfaces of solid materials, and multiphoton photopolymerization in photoresists or resins. Currently, there are several 3-D fs-laser nanofabrication methods. The most widely used methods are the direct laser writing (DLW) technique and multiple beam interference lithography, as schematically shown in Fig. 4.27.

Here, we have chosen the DLW technique as an example to explain the most common principles of fs-laser nanofabrication. Typically, DLW is performed by scanning a tightly focused fs-laser beam in accordance with a designed 3-D pattern. The photomodification processes occur in the waist of the focused laser beam, where the optical field intensity is highest. The effective photomodification region can be approximated by an ellipsoid, as illustrated in Fig. 4.27a. The use of tightly focusing optics (i. e., an objective lens with a high numerical aperture) typically allows the lateral dimensions of the ellipsoid to approximate the wavelength λ. For 3-D nanofabrication inside materials, the materials must be transparent at the laser wavelength so that nonlinear absorption is considered for photomodification. Because of the nonlinearity, the effective absorbing region can be even smaller than that of the focal spot, enabling structural resolution beyond the optical diffraction limit. For fs-lasers operating at near-infrared and visible wavelengths, feature sizes of 100–300 nm in 3-D structures are easily achievable. In some special cases, especially for linear structures, a resolution of tens of nanometers has also been demonstrated.

Alternatively, multiple beam interference lithography can be used for laser nanofabrication. This method involves exposing a photosensitive medium to a periodic interference field created by several coherent laser beams that overlap at nonzero mutual angles. Typically, the original beam passes through an optical element (i. e., beam splitter, prism, or grating) that splits the beam into several coherent beams to form the desired interference patterns. In particular, 3-D periodic or quasiperiodic structures can be created with a resolution on the order of several hundred nanometers. Compared to the DLW technique, larger scale structures can be easily implemented. However, this technique is not as flexible as DLW, in which any desired 3-D geometry can be achieved.

Various Nanostructures Fabricated by Femtosecond Laser

A wide range of materials can be used for DLW, including silica or other glasses (by optical damaging), liquid photopolymerizable resins (by photosolidification), or photoresists (by photopolymerization). The

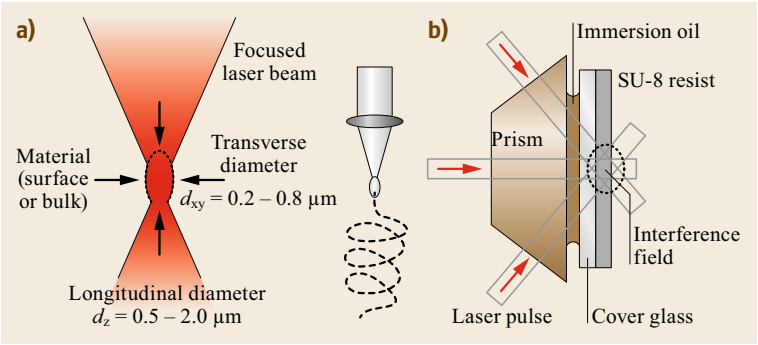


Fig. 4.27 (a) Sketch map of the principle of DLW using a focused fs-laser beam. (b) An example of the configurations for multiple beam interference lithography

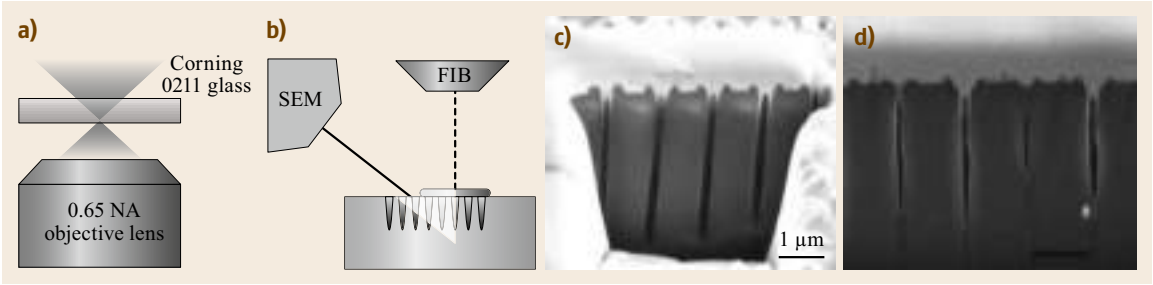


Fig. 4.28 **(a)** An illustration of the machining geometry for the array of nanochannels. **(b)** A sample examined under a dual-beam FIB–SEM to study the cross-sections of nanochannels. **(c)** An example cross-section for back side machining. **(d)** An example cross-section for front side machining (after [4.68])

first example, shown in Fig. 4.28, is high-aspect ratio nanochannel formation in glass by single tightly focused fs-laser pulses [4.68]. The nanochannels can be fabricated from both the back (Fig. 4.28c) and front (Fig. 4.28d) of the glass sample. Here, the width of the nanochannels is in the 150–240 nm range, much smaller than the width (344 nm) of the focused laser beam, providing clear proof that the structural resolution or the feature size of the structures fabricated by fs-lasers can be considerably smaller than the optical diffraction limit.

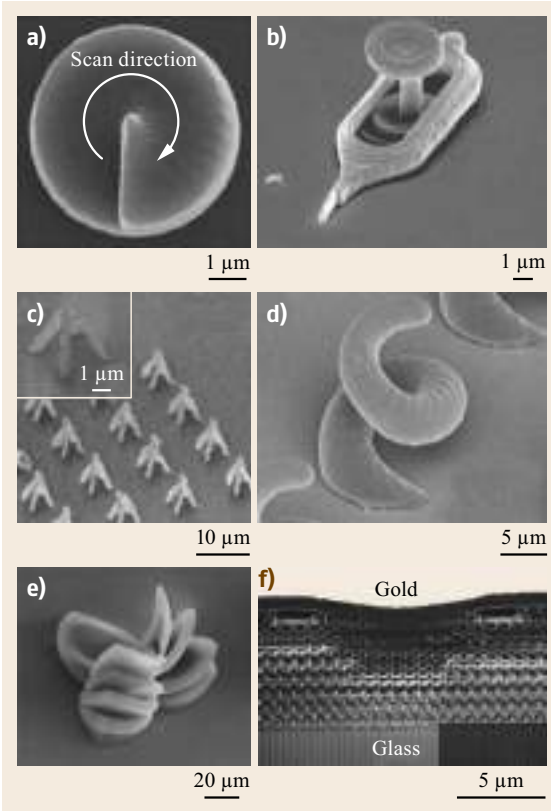


Figure 4.29 shows various 3-D micro/nanostructures fabricated using fs-lasers by different research groups. Direct fs-laser photopolymerization in resists was used to fabricate a high-resolution microscopic spiral phase plate (Fig. 4.29a) that can be used to generate vortex beams [4.69]. Using a similar technique, movable micromanipulators with submicron probe tips were also fabricated and demonstrated the ability to manipulate nanoscale objectives (Fig. 4.29b) [4.70]. Metal micro/nanostructures that may have applications in metamaterials can also be fabricated by fs-lasers. Figure 4.29c shows an array of freestanding 3-D 5 µm high silver pyramids fabricated by fs-laser DLW based on multiphoton photoreduction in a metal-ion solution [4.71]. Coating a polymeric template fabricated by fs-laser DLW with metal using the selective electroless metal plating technique is another approach for obtaining metal structures. Silver spiral structures were obtained using this method, as shown in Fig. 4.29d [4.72]. Biomimetic and biological nanofabrication were also demonstrated using a so-called dynamic laser prototyping technique based on fs-laser DLW. For example, the flower blooming process has been successfully mimicked and a typical bloomed *flower* is shown in Fig. 4.29e [4.73]. Figure 4.29f shows an fs-laser-fabricated 3-D invisibility-cloaking structure operating at optical wavelengths [4.74]. These examples demonstrate the versatility of fs-laser nanofabrication.

Fig. 4.29a–f SEM images of various 3-D structures fabricated based on fs-laser nanofabrication. **(a)** Spiral phase plate (Reprinted from [4.69] with the permission of AIP publishing), **(b)** micromanipulator (Reprinted with permission from [4.70] with the permission of AIP publishing), **(c)** silver pyramid structures (Reprinted from [4.71]), **(d)** silver-coated spirals (From [4.72]), **(e)** flower structure made by biomimetic and bionanofabrication (Reprinted with permission from [4.73]), and **(f)** invisibility-cloaking structure (From [4.74]. Reprinted with permission from AAAS) ◄

4.4.2 3-D Nanodevices: Fabrication of Photonic Crystals

PhCs are periodically structured materials that have attracted much attention owing to their unique properties, such as the manipulation of light propagation at a certain range of frequencies, called photonic band gaps (PBGs) or stop bands. Thus far, a variety of PhC structures have been investigated both theoretically and experimentally. For 1-D (one-dimensional) or 2-D PhCs, periodicity is present only in one or two dimensions, respectively and therefore such PhCs can be formed by many methods. Fabrication of 3-D PhCs is more difficult, especially for those with PBGs at the optical frequencies.

Fabrication of Photonic Crystal in Photoresists by DLW

Fs-laser structuring of materials is a promising technique for successful implementation of 3-D PhCs with spatial resolutions in the hundred-nanometer range. In particular, since 1999, the two-photon polymerization technique in photoresists or resins by DLW has been demonstrated as a powerful tool for the fabrication of 3-D PhCs [4.75, 76]. Two popular PhC geometries are often used in our DLW technique. The first is the woodpiles architecture (Fig. 4.30a,b), and the second is the square spiral architecture (Fig. 4.30c). Regarding the resist materials, SU-8 was previously the most frequently used resist, but organic-inorganic hybrid resists such as SZ2080 have now also become popular.

The optical setup for 3-D DLW used in our experiments is schematically shown in Fig. 4.30d. A chirped-pulsed amplifier (CPA) Ti:sapphire laser (Hurricance, Spectra-Physics) was used as the light source for DLW. It delivered ≈ 120 fs-laser pulses at a central wavelength of 800 nm with a repetition rate of 1 kHz. The laser beam was sent to an optical microscope (IX71, Olympus) and was focused into the sample by an oil-immersion objective lens (magnification 100×, numerical aperture NA = 1.40) along the z-direction. The sample was mounted on a piezoelectric transducer-controlled 3-D translation stage (x-y: P-517.2CL and z: P-518.ZCL, from Physik Instrumente), allowing fabrication within a volume of $100 \times 100 \times 100\,\mu m^3$ with an accuracy of ≈ 2 nm. After exposure to the laser, the samples were developed and rinsed.

The square spiral architecture was proposed to have complete PBGs. The square spiral structures were successfully fabricated from silicon using the glancing angle deposition (GLAD) method which uses the growth of silicon spirals on prepatterned substrates [4.80]. However, the DLW technique enables easier fabrication of both square spirals and circular spirals. Furthermore, variants with greater complexity (e.g., those containing defects or phase-shifted spirals) that are sometimes necessary for a particular application or optimization of the PBG properties can only be fabricated by DLW. Figure 4.31a,b show typical square spiral and circular PhCs, respectively, fabricated by DLW [4.77]. Here, the adjacent spirals are interlaced with each other. The ability to create defects in PhCs by DLW is also illustrated

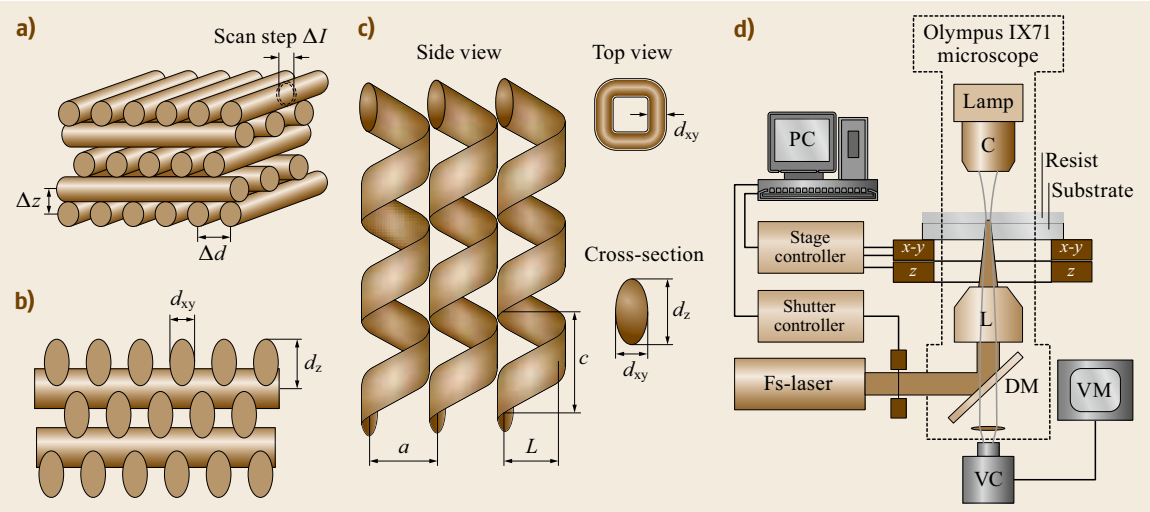


Fig. 4.30a–d Schematic oblique view **(a)** and cross-section **(b)** of the woodpile structure with the main parameters indicated. **(c)** Schematic side and top view of the square spiral structure with main parameters indicated. **(d)** Typical layout of the experimental setup for DLW. C: condenser, L: objective lens, DM: dichroic mirror, VM: video monitor, VC: video camera

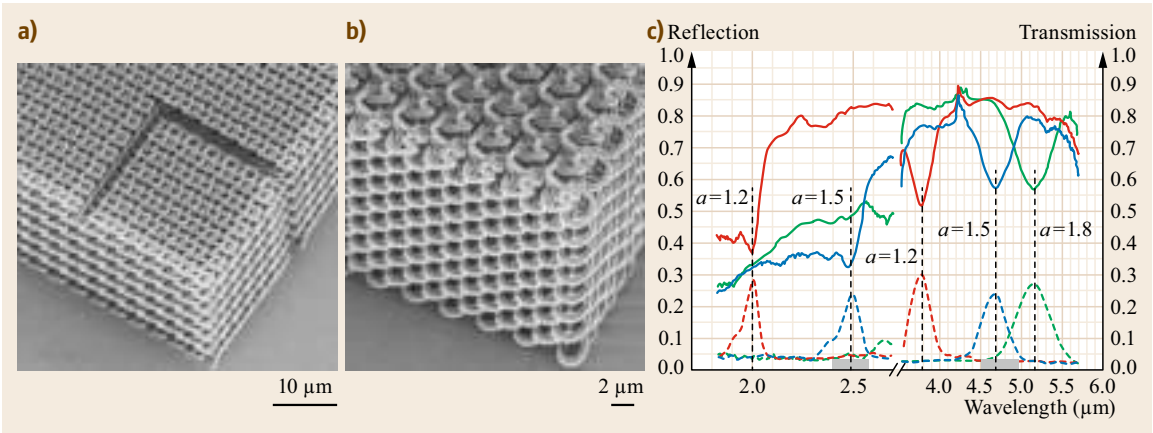


Fig. 4.31a–c SEM images of 3-D PhC structures with spiral architecture in SU-8. **(a)** Square spiral PhC with an *L*-shaped line defect; **(b)** circular spiral PhCs with a π-phase shift between the adjacent spirals. Adjacent spirals are interlaced in both PhCs. **(c)** Infrared reflectivity and transmission spectra of several square spiral PhCs with lattice parameters $L = 1.5a$ and $c = 2.49a$ and different lattice constants $a = 1.2$, 1.5, and 1.8 μm. (Reprinted with permission from [4.77])

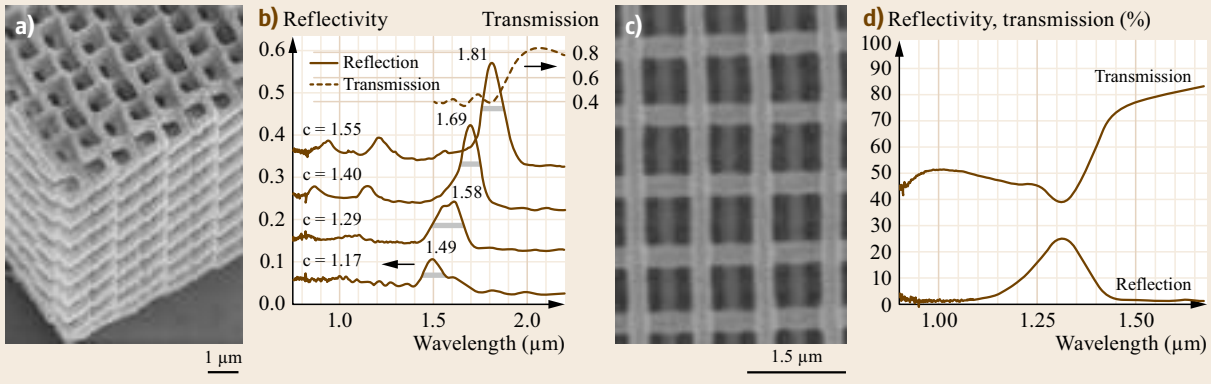


Fig. 4.32 **(a)** SEM images of a 3-D square spiral PhC in SU-8. In this case, the adjacent spirals are not interlaced so that the lattice constant a can be reduced to ≈ 1 μm. **(b)** Transmission and reflection spectra of several square spiral structures with the same parameters $a = 1.0$ μm, $L = 0.7$ μm but different $c = 1.17$, 1.29, 1.40, and 1.55 μm. The *top inset* shows the transmission spectrum of the sample with $c = 1.55$ μm. The central wavelengths of the reflectivity peaks are indicated in the plot in micrometers, demonstrating the stop bands in the telecommunications spectral range. **(c)** SEM image of a woodpile PhC in SZ2080, **(d)** transmission and reflection spectra of the woodpile structure, with the stop band centered at 1.3 μm. ((**a**,**b**) from [4.78], (**c**,**d**) after [4.79])

in Fig. 4.31a, where an L-shaped line defect is formed. Notably, the π-phase shift between the adjacent spirals, as shown in Fig. 4.31b, would not be achievable by the GLAD technique. The reflection and transmission spectra of this type of square spiral PhC are presented in Fig. 4.31c, showing that the photonic stop bands are in the infrared wavelength range of 2–6 μm.

For other types of spiral PhCs, where the spirals are not interlaced, the lattice constant can be reduced to $\approx$ 1 μm as shown in Fig. 4.32a [4.78]. In this case, the stop band can be engineered in the telecommunication wavelengths, as shown in Fig. 4.32b. For woodpile PhCs, it is also possible to tailor the stop band in this wavelength range. For example, Fig. 4.32c,d show an SEM image and reflection spectra, respectively, of a typical woodpile PhC [4.79]. The stop band is located at 1.3 μm. However, engineering the stop band to shift to a shorter wavelength and even into the visible region is very difficult. Such engineering may require the reduction of the voxel size using low-pulse energy close to the polymerization threshold. This will cause the structure to shrink [4.79, 81], as discussed below.

Shrinkage

Shrinkage is usually an undesirable feature for photopolymerized micro/nanostructures fabricated by the

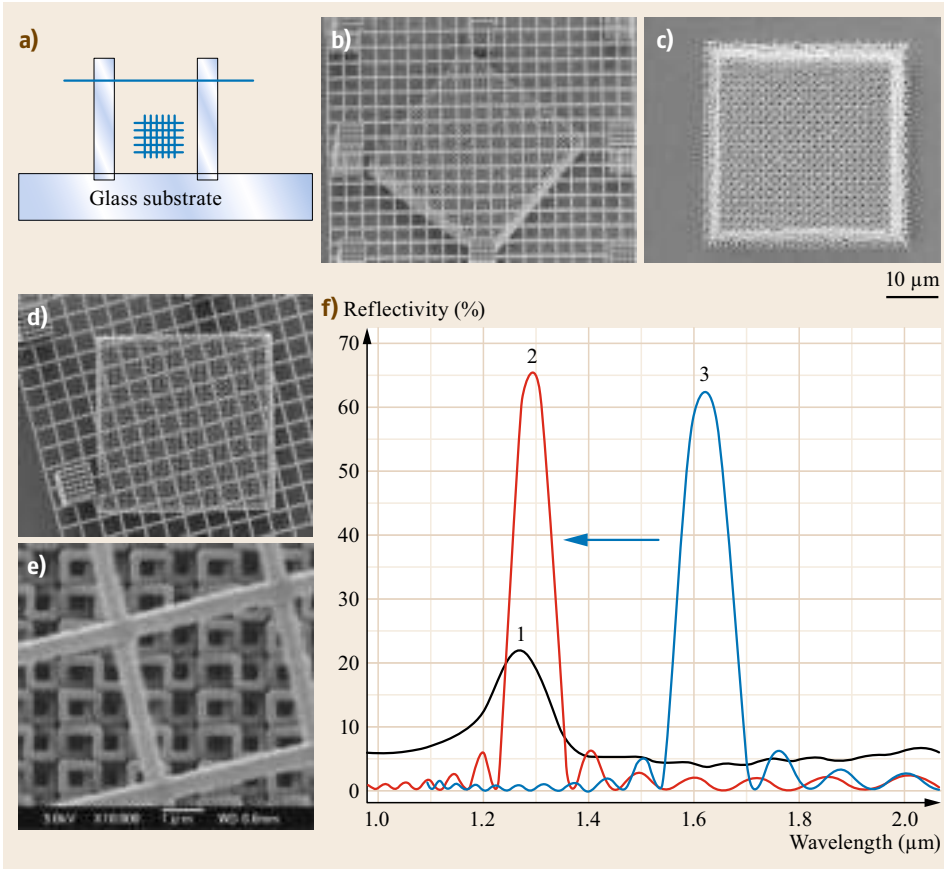


Fig. 4.33 **(a)** Schematic diagram of freestanding architecture fabrication. **(b)** SEM image (oblique view) of a freestanding square spiral PhC in SZ2080. Comparison between a conventional PhC fixed to the substrate **(c)** and a freestanding PhC **(d)**, fabricated with the same laser writing conditions. The scale bar for **(b–d)** is the same, as shown below **(c)**. **(e)** Zoomed top view of the PhC, exhibiting good periodicity. **(f)** FTIR (Fourier transform infrared) measured reflection spectrum of the freestanding woodpile PhC with 20% shrinkage (*curve 1*) and the corresponding finite-difference time-domain (FDTD) simulation (*curve 2*). An FDTD simulation of the designed structure is added for comparison (*curve 3*). (**(b–e)** from [4.81], Copyright from IOP Publishing. Reproduced with permissions. All rights reserved)

DLW technique in resists. The shrinkage is normally nonuniform because the bottom of a conventional structure is attached to the substrate. We proposed a way to fabricate freestanding structures that can enable highly uniform shrinkage [4.81]. Figure 4.33a illustrates the concept of the freestanding architecture. These structures were fabricated while floating on the substrates. However, such structures would escape during the development process. To prevent this, we also fabricated cages to house the freestanding structures. An SEM image (oblique view) illustrating this approach is presented in Fig. 4.33b, which shows that after development the freestanding PhC is moved from its original position (*center* of the cage) as fabricated. The structure still exhibits some shrinkage. However, by comparing conventional and freestanding PhCs (Fig. 4.33c,d), we find that the freestanding PhC shows almost uniform shrinkage, in contrast to the nonuniform shrinkage observed in the conventional PhC. The fabricated freestanding structures allowed us to quantitatively measure the shrinkage, defined as $(W_0 - W)/W_0$, where W and W_0 are the measured structure width and the corresponding design value, respectively. It was found that the amount of shrinkage depends on the pulse energy used for DWL and that the maximum shrinkage can reach 20% with the lowest allowable pulse energy, beyond which the structure collapses. The freestanding PhCs still exhibit high structural quality and good periodicity, as shown in Fig. 4.33e. Therefore, such PhCs exhibit good optical properties and their stop bands can be scaled towards shorter wavelengths, thus benefiting from uniform shrinkage, as illustrated in Fig. 4.33f.

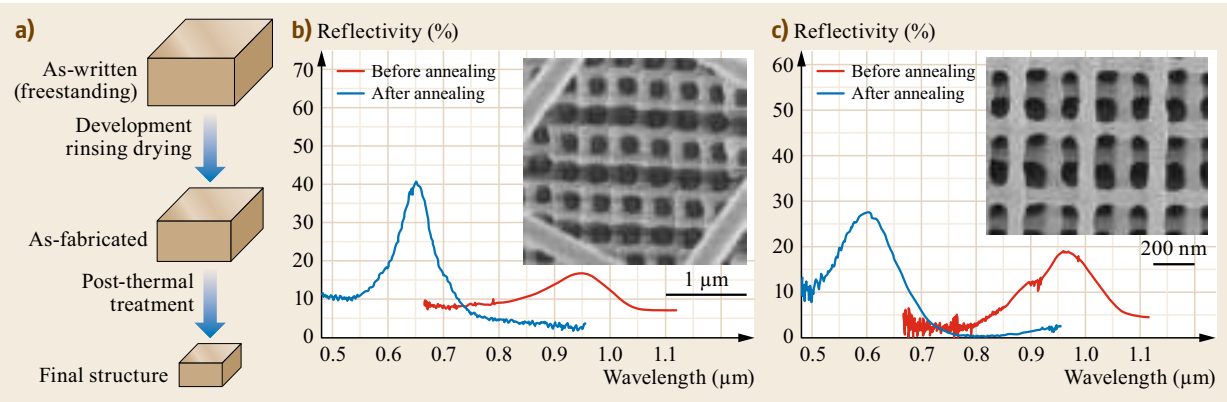


Fig. 4.34 **(a)** Schematic diagram of combined freestanding architecture and post-thermal treatment. Reflection spectra of a square spiral PhC **(b)** and a woodpile PhC **(c)** before and after annealing. The final stop bands for the spiral and woodpile architectures are 650 and 600 nm, respectively. The *insets* in **(b)** and **(c)** are SEM images of the corresponding PhCs

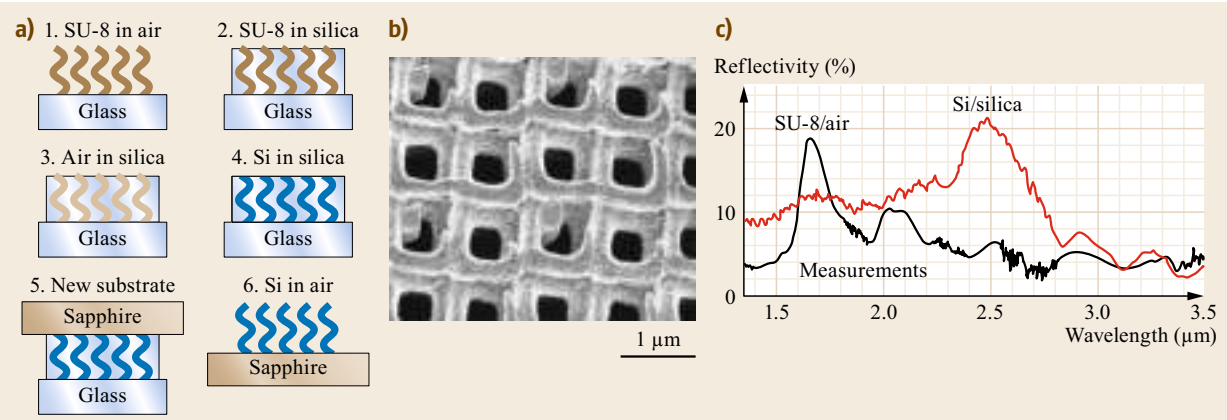


Fig. 4.35 **(a)** Main steps of the double-inversion process of SU-8 templates by silicon. **(b)** SEM image of the silicon PhC fabricated by the double-inversion technique. **(c)** Reflection spectra of the 3-D spiral PhC template of SU-8 before and after its replication by silicon. ((**b**) Copyright 2008, IEEE. Reprinted with permission from [4.82])

We also investigated the effect of the drying process on the shrinkage. By measuring the evolution of the optical properties of fabricated PhCs and recording online optical images during the drying process, we demonstrated that the drying process plays the most important role in the structural shrinkage of photopolymerized structures [4.83]. Furthermore, we found that the structures experienced additional shrinkage when we baked the fabricated structures. This was mostly because our material was an organic-inorganic resist, so the organic part could be degraded during the thermal treatment. We then hypothesized that very large uniform shrinkage could be achieved if we combined the freestanding configuration and such post-thermal treatment, as illustrated in Fig. 4.34a. By carefully choosing the lattice constant of a square spiral PhC and the irradiated pulse energy, we successfully obtained a shrinkage magnitude as large as 46% after baking at 300 °C for 3 h. Owing to such large shrinkage, the stop band of the PhC was successfully shifted into the visible range. As clearly shown in Fig. 4.34b, the stop band was centered at 650 nm, while prior to the thermal treatment it was centered at 950 nm. It is important to note that previously the stop band position of a square spiral PhC fabricated by DLW could hardly reach 1 μm. Additionally, the lattice constant of the PhC was reduced to below 400 nm (*inset* in Fig. 4.34b); this was impossible to achieve with the conventional DLW technique without employing the freestanding configuration and post-thermal treatment. For the woodpile PhCs, we even succeeded in pushing the stop band to 600 nm after annealing, as shown in Fig. 4.34c. The shrinkage can also be used to improve the structural resolution of DLW-fabricated micro/nano structures. As shown in the

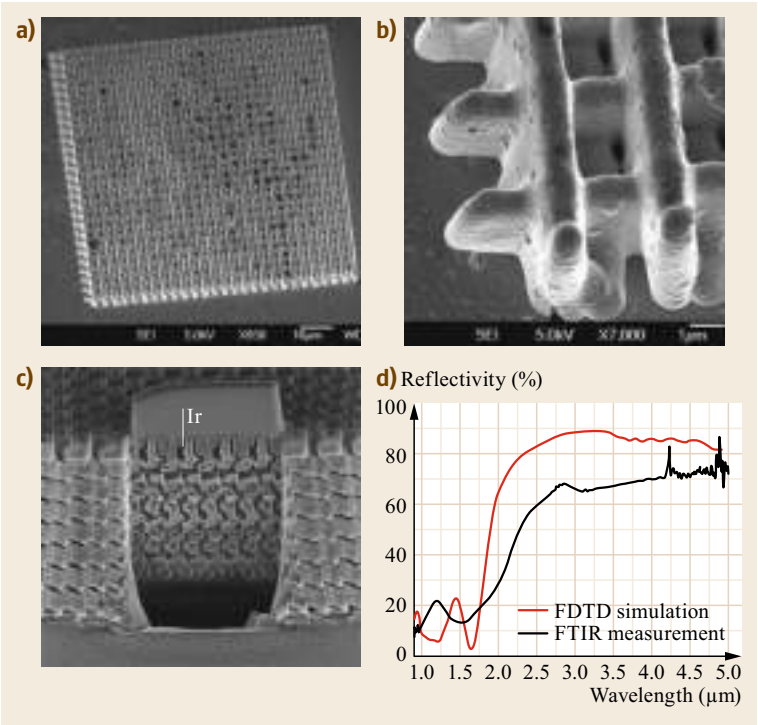


Fig. 4.36 **(a,b)** SEM images of a woodpile PhC coated by Ni film using electroless coating. **(c)** SEM image of a square spiral PhC coated by Ir. The layer of Ir was confirmed by the cross-section cut by FIB. **(d)** FTIR measured and FDTD calculated optical reflection spectra at infrared wavelengths. ((**a,b**) from [4.84])

insets of Fig. 4.34b,c, the linewidth of the structure is only $\approx 100\,\mathrm{nm}$.

Replication of PhCs with Other Materials

The polymeric PhCs fabricated by DLW cannot have omnidirectional PBGs owing to the low index contrast between the photoresist and the surrounding air (1.6–1 for SU-8, 1.5–1 for SZ2080). Nevertheless, the polymeric PhCs are suitable for subsequent infiltration by other dielectric materials with a higher refractive index or a metal material.

As one example, we applied the double-inversion technique to invert the templates to silica structures and then again inverted to the silicon structure [4.82], as schematically shown in Fig. 4.35a. Figure 4.35b shows the SEM image of the replicated PhC after the double-inversion process. Compared to the templates, the PBG band was shifted to a longer wavelength because of the higher refractive index contrast, as shown in Fig. 4.35c.

Polymeric templates can also be coated with metals using metallic coating techniques. Figure 4.36a,b show a woodpile PhC coated by a film of nickel using electroless coating [4.84]. Figure 4.36c shows a square spiral PhC coated by an iridium film using atomic layer deposition. As observed from both the experimental measurement and finite-difference time-domain (FDTD) simulation of the reflection spectra shown in Fig. 4.36d, the iridium PhCs exhibit a very wide photonic band gap with a sharp band edge, showing potential for photonic and energy applications.

References

4.1 S. Matsui: Nanostructure fabrication using electron beam and its application to nanometer devices, Proc. IEEE **85**, 629–642 (1997)

4.2 A. Wargner, J.P. Levin, J.L. Mauer, P.G. Blauner, S.J. Kirch, P. Long: X-ray mask repair with focused ion beams, J. Vac. Sci. Technol. B **8**, 1557–1564 (1990)

4.3 O. Lehmann, M. Stuke: Generation of three-dimensional free-standing metal micro-objects by laser chemical processing, Appl. Phys. A **53**, 343–345 (1991)

4.4 H.W.P. Koops, J. Kretz, M. Rudolph, M. Weber, G. Dahm, K.L. Lee: Characterization and application of materials grown by electron-beam-induced deposition, Jpn. J. Appl. Phys. **33**, 7099–7107 (1994)

4.5 S. Matsui, T. Kaito, J. Fujita, M. Komuro, K. Kanda, Y. Haruyama: Three-dimensional nanostructure fabrication by focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **18**, 3181–3184 (2000)

4.6 H.B. Sun, S. Matsuo, H. Misawa: Three-dimensional photonic crystal structures achieved with two-pho-

ton-absorption photopolymerization of resin, Appl. Phys. Lett. **74**, 786–788 (1999)
4.7 K. Kand, J. Igaki, Y. Kato, R. Kometani, A. Saikubo, S. Matsui: NEXAFA study on carbon-based material formed by focused-ion-beam chemical-vapor-deposition, Radiat. Phys. Chem. **75**, 1850–1854 (2006)
4.8 J. Igaki, A. Saikubo, R. Kometani, K. Kanda, T. Suzuki, K. Niihara, S. Matsui: Elementary analysis of diamond-like carbon film formed by focused-ion-beam chemical vapor deposition, Jpn. J. Appl. Phys. **46**, 8003–8004 (2007)
4.9 T. Hoshino, K. Watanabe, R. Kometani, T. Morita, K. Kanda, Y. Haruyama, T. Kaito, J. Fujita, M. Ishida, Y. Ochiai, S. Matsui: Development of three-dimensional pattern-generating system for focused-ion-beam chemical-vapor-deposition, J. Vac. Sci. Technol. B **21**, 2732–2736 (2003)
4.10 R. Kometani, S. Ishihara, T. Kaito, S. Matsui: In situ observation of the three-dimensional nano-structure growth on focused-ion-beam chemical vapor deposition by scanning electron microscope, Appl. Phys. Express **1**, 055001 (2008)
4.11 E. Buks, M.L. Roukes: Stiction, adhesion energy and the Casimir effect in micromechanical systems, Phys. Rev. B **63**, 033402 (2001)
4.12 J. Fujita, M. Ishida, T. Sakamoto, Y. Ochiai, T. Kaito, S. Matsui: Observation and characteristics of mechanical vibration in three-dimensional nanostructures and pillars grown by focused ion beam chemical vapor deposition, J. Vac. Sci. Technol. B **19**, 2834–2837 (2001)
4.13 M. Ishida, J. Fujita, Y. Ochiai: Density estimation for amorphous carbon nanopillars grown by focused ion beam assisted chemical vapor deposition, J. Vac. Sci. Technol. B **20**, 2784–2787 (2002)
4.14 T. Morita, K. Nakamatsu, K. Kanda, Y. Haruyama, K. Kondo, T. Hoshino, T. Kaito, J. Fujita, T. Ichihashi, M. Ishida, Y. Ochiai, T. Tajima, S. Matsui: Nanomechanical switch fabrication by focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **22**, 3137–3142 (2004)
4.15 R. Kometani, T. Hoshino, K. Kondo, K. Kanda, Y. Haruyama, T. Kaito, J. Fujita, M. Ishida, Y. Ochiai, S. Matsui: Characteristic of nano-electrostatic actuator fabricated by focused ion beam chemical vapor deposition, Jpn. J. Appl. Phys. **43**, 7187–7191 (2004)
4.16 R. Kometani, T. Morita, K. Watanabe, K. Kanda, Y. Haruyama, T. Kaito, J. Fujita, M. Ishida, Y. Ochiai, S. Matsui: Nozzle-nanostructure fabrication on glass capillary by focused-ion-beam chemical vapor deposition and etching, Jpn. J. Appl. Phys. **42**, 4107–4110 (2003)
4.17 R. Kometani, T. Hoshino, K. Kondo, K. Kanda, Y. Haruyama, T. Kaito, J. Fujita, M. Ishida, Y. Ochiai, S. Matsui: Performance of nanomanipulator fabricated on glass capillary by focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **23**, 298–301 (2005)
4.18 R. Kometani, T. Hoshino, K. Kanda, Y. Haruyama, T. Kaito, J. Fujita, M. Ishida, Y. Ochiai, S. Matsui: Three-dimensional high-performance nanotools fabricated using focused-ion-beam chemical vapor deposition, Nucl. Instrum. Methods. Phys. Res. B **232**, 362–366 (2005)
4.19 K. Nakamatsu, M. Nagase, H. Namatsu, S. Matsui: Mechanical characteristics of diamond-like-carbon nanosprings fabricated by focused-ion-beam chemical vapor deposition, Jpn. J. Appl. Phys. **44**, L1228–L1230 (2005)
4.20 T. Morita, R. Kometani, K. Watanabe, K. Kanda, Y. Haruyama, T. Hoshino, K. Kondo, T. Kaito, T. Ichihashi, J. Fujita, M. Ishida, Y. Ochiai, T. Tajima, S. Matsui: Free-space-wiring fabrication in nano-space by focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **21**, 2737–2741 (2003)
4.21 J. Fujita, M. Ishida, T. Ichihashi, Y. Ochiai, T. Kaito, S. Matsui: Graphitization of Fe-doped amorphous carbon pillars grown by focused ion-beam-induced chemical-vapor deposition, J. Vac. Sci. Technol. B **20**, 2686–2689 (2002)
4.22 D. Guo, R. Kometani, S. Warisawa, S. Ishihara: Growth of ultra-long free-space-nanowire by the real-time feedback control of the scanning speed on focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **31**, 061601 (2013)
4.23 R. Kometani, S. Warisawa, S. Ishihara: The 3-D nanostructure growth evaluations by the real-time current monitoring on focused-ion-beam chemical vapor deposition, Microelectron. Eng. **87**, 1044–1048 (2010)
4.24 K. Nakamatsu, K. Yamamoto, T. Hirayama, S. Matsui: Fabrication of fine electron biprism filament by free-space-nanowiring technique of focused-ion-beam + chemical vapor deposition for accurate off-axis electron holography, Appl. Phys. Express **1**, 117004 (2008)
4.25 R. Kometani, K. Yusa, S. Warisawa, S. Ishihara: Piezoresistive effect in the three-dimensional diamondlike carbon nanostructure fabricated by focused-ion-beam chemical vapor deposition, J. Vac. Sci. Technol. B **28**, C6F38–41 (2010)
4.26 J. Dai, K. Onomitsu, R. Kometani, Y. Krockenberger, H. Yamaguchi, S. Ishihara, S. Warisawa: Superconductivity in tungsten-carbide nanowires deposited from the mixtures of $W(CO)_6$ and $C_{14}H_{10}$, Jpn. J. Appl. Phys. **52**, 075001 (2013)
4.27 J. Dai, R. Kometani, K. Onomitsu, Y. Krockenberger, H. Yamaguchi, S. Ishihara, S. Warisawa: Direct fabrication of a W-C SNS Josephson junction using focused-ion-beam chemical vapor deposition, J. Micromech. Microeng. **24**, 055015 (2014)
4.28 P. Vukusic, J.R. Sambles: Photonic structures in biology, Nature **424**, 852–855 (2003)
4.29 K. Watanabe, T. Hoshino, K. Kanda, Y. Haruyama, S. Matsui: Brilliant blue observation from a morpho-butterfly-scale quasi-structure, Jpn. J. Appl. Phys. **44**, L48–L50 (2005)
4.30 A.N. Broers, W.W. Molzen, J.J. Cuomo, N.D. Wittels: Electron-beam fabrication of 80-Å metal structures, Appl. Phys. Lett. **29**, 596–597 (1976)
4.31 S. Matsui, K. Mori: New selective deposition technology by electron beam induced surface reaction, Jpn. J. Appl. Phys. **23**, L706–L708 (1984)

4.32 S. Matsui, K. Mori: New selective deposition technology by electron beam induced surface reaction, J. Vac. Sci. Technol. B **4**, 299–304 (1986)

4.33 H.W.P. Koops, R. Weiel, D.P. Kern, T.H. Baum: High-resolution electron-beam induced deposition, J. Vac. Sci. Technol. B **6**, 477–481 (1988)

4.34 S. Matsui, T. Ichihashi, M. Mito: Electron beam induced selective etching and deposition technology, J. Vac. Sci. Technol. B **7**, 1182–1190 (1989)

4.35 Y. Ochiai, J. Fujita, S. Matsui: Electron-beam-induced deposition of copper compound with low resistivity, J. Vac. Sci. Technol. B **14**, 3887–3891 (1996)

4.36 I. Utke, P. Hoffmann, B. Dwir, K. Leifer, E. Kapon, P. Doppelt: Focused electron beam induced deposition of gold, J. Vac. Sci. Technol. B **18**, 3168–3171 (2000)

4.37 H.W.P. Koops, A. Reinhardt, F. Klabunde, A. Kaya, R. Plontke: Vapor supply manifold for additive nanolithography with electron beam induced deposition, Microcircuit Eng. **57/58**, 909–913 (2001)

4.38 U. Hübner, R. Plontke, M. Blume, A. Reinhardt, H.W.P. Koops: On-line nanolithography using electron beam-induced deposition technique, Microelectron. Eng. **57/58**, 953–958 (2001)

4.39 H.W.P. Koops, O.E. Hoinkis, M.E.W. Honsberg, R. Schmidt, R. Blum, G. Böttger, A. Kuligk, C. Liguda, M. Eich: Two-dimensional photonic crystals produced by additive nanolithography with electron beam-induced deposition act as filters in the infrared, Microelectron. Eng. **57/58**, 995–1001 (2001)

4.40 F. Floreani, H.W.P. Koops, W. Elsäßer: Operation of high power field emitter fabricated with electron beam deposition and concept of a miniaturized free electron laser, Microelectron. Eng. **57/58**, 1009–1016 (2001)

4.41 K. Mitsuishi, M. Shimojo, M. Han, K. Furuya: Electron-beam-induced deposition using a sub-nanometer-sized probe of high-energy electrons, Appl. Phys. Lett. **83**, 2064–2066 (2003)

4.42 M. Shimojo, M. Takeguchi, M. Tanaka, K. Mitsuishi, K. Furuya: Electron beam-induced deposition using iron carbonyl and the effects of heat treatment on nanostructure, Appl. Phys. A **79**, 1869–1872 (2004)

4.43 M. Tanaka, M. Shimojo, M. Han, K. Mitsuishi, K. Furuya: Ultimate sized nano-dots formed by electron beam-induced deposition using an ultrahigh vacuum transmission electron microscope, Surf. Interface Anal. **37**, 261–264 (2005)

4.44 I. Utke, V. Friedli, M. Purrucker, J. Michler: Resolution in focused electron- and ion-beam induced processing, J. Vac. Sci. Technol. B **25**, 2219–2223 (2007)

4.45 J.D. Barry, M. Ervin, J. Molstad, A. Wickenden, T. Brintlinger, P. Hoffman, J. Melngailis: Electron beam induced deposition of low resistivity platinum from $Pt(PF_3)_4$, J. Vac. Sci. Technol. B **24**, 3165–3168 (2006)

4.46 A. Perentes, G. Sinicco, G. Boero, B. Dwir, P. Hoffmann: Focused electron beam induced deposition of nickel, J. Vac. Sci. Technol. B **25**, 2228–2232 (2007)

4.47 A. Botman, D.A.M. de Winter, J.J.L. Muders: Electron-beam-induced deposition of platinum at low landing energies, J. Vac. Sci. Technol. B **26**, 2460–2463 (2008)

4.48 A. Botman, M. Hesselberth, J.J.L. Mulders: Investigation of morphological changes in platinum-containing nanostructures created by electron-beam-induced deposition, J. Vac. Sci. Technol. B **26**, 2464–2467 (2008)

4.49 S.J. Randolph, J.D. Fowlkes, P.D. Rack: Focused, nanoscale electron-beam-induced deposition and etching, Crit. Rev. Solid State Mater. Sci. **31**, 55–89 (2006)

4.50 W.F. van Dorp, C.W. Hagen: A critical literature review of focused electron beam induced deposition, J. Appl. Phys. **104**, 081301 (2008)

4.51 I. Utke, P. Hoffmann, J. Melngailis: Gas-assisted focused electron beam and ion beam processing and fabrication, J. Vac. Sci. Technol. B **26**, 1197–1276 (2008)

4.52 J. Bishop, C.J. Lobo, A. Martin, M. Ford, M. Phillips, M. Toth: Role of activated chemisorption in gas-mediated electron beam induced deposition, Phys. Rev. Lett. **109**, 146103 (2012)

4.53 N. Silvis-Cividjian, C.W. Hagen, L.H.A. Leunissen, P. Kruit: The role of secondary electrons in electron-beam-induced deposition spacial resolution, Microelectron. Eng. **61/62**, 693–699 (2002)

4.54 V. Friedli, I. Utke, K. Mølhave, J. Michler: Dose and energy dependence of mechanical properties of focused electron-beam induced pillar deposits from $Cu(C_5HF_6O_2)_2$, Nanotechnology **20**, 385304 (2009)

4.55 R. Lavrijsen, R. Córdoba, F.J. Schoenaker, T.H. Ellis, B. Barcones, J.T. Kohlhepp, H.J.M. Swagten, B. Koopmans, J.M. De Teresa, C. Magen, M.R. Ibarra, P. Trompenaars, J.J.L. Mulders: Fe:O:C grown by focused-electron-beam-induced deposition: Magnetic and electric properties, Nanotechnology **22**, 025302 (2011)

4.56 T. Brintlinger, M.S. Fuhrer, J. Melngailis, I. Utke, T. Bret, A. Perentes, P. Hoffmann, M. Abourida, P. Doppelt: Electrodes for carbon nanotube devices by focused electron beam induced deposition of gold, J. Vac. Sci. Technol. B **23**, 3174–3177 (2005)

4.57 S. Graells, R. Alcubilla, G. Badenes, R. Quidant: Growth of plasmonic gold nanostructures by electron beam induced deposition, Appl. Phys. Lett. **91**, 121112 (2007)

4.58 A. Fernández-Pacheco, J.M. de Teresa, R. Cordoba, M.R. Ibarra, D. Petit, D.E. Read, L. O'Brien, E.R. Lewis, H.T. Zeng, R.P. Cowburn: Domain wall conduit behavior in cobalt nanowires grown by focused electron beam induced deposition, Appl. Phys. Lett. **94**, 192509 (2009)

4.59 J. Pablo-Navarro, C. Magén, J.M. de Teresa: Three-dimensional core-shell ferromagnetic nanowires grown by focused electron beam induced deposition, Nanotechnology **27**, 285302 (2016)

4.60 H. Acar, T. Coenen, A. Polman, L.K. Kuipers: Dispersive ground plane core-shell type optical monopole antennas fabricated with electron beam induced deposition, ACS Nano **6**, 8226–8232 (2012)

4.61 P. Woźniak, K. Höflich, G. Brönstrup, P. Banzer, S. Christiansen, G. Leuchs: Unveiling the optical properties of a metamaterial synthesized by electron-beam-induced deposition, Nanotechnol-

ogy **27**, 025705 (2016)
4.62 I. Utke, S. Moshkalev, P. Russel (Eds.): *Nanofabrication Using Focused Ion and Electron-Beams* (Oxford Univ. Press, Oxford 2012)
4.63 S. Matsui, K. Mori: In situ observation on electron beam induced chemical vapor deposition by Auger electron spectroscopy, Appl. Phys. Lett. **51**, 646–648 (1987)
4.64 S. Matsui, T. Ichihashi: In situ observation on electron-beam-induced chemical vapor deposition by transmission electron microscopy, Appl. Phys. Lett. **53**, 842–844 (1988)
4.65 V. Tasco, M. Esposito, F. Todisco, A. Benedetti, M. Cuscunà, D. Sanvitto, A. Passaseo: Three-dimensional nanohelices for chiral photonics, Appl. Phys. A **122**, 280 (2016)
4.66 S. Juodkazis, V. Mizeikis, H. Misawa: Three-dimensional microfabrication of materials by femtosecond lasers for photonics applications, J Appl. Phys. **106**, 051101 (2009)
4.67 K. Sugioka, Y. Cheng: Ultrafast lasers-reliable tools for advanced materials processing, Light Sci. Appl. **3**, e149 (2014)
4.68 J.F. Herbstman, A.J. Hunt: High-aspect ratio nanochannel formation by single femtosecond laser pulses, Opt. Express **18**, 16840–16848 (2010)
4.69 E. Brasselet, M. Malinauskas, A. Zukauskas, S. Juodkazis: Photopolymerized microscopic vortex beam generators: Precise delivery of optical orbital angular momentum, Appl. Phys. Lett. **97**, 211108 (2010)
4.70 S. Maruo, K. Ikuta, H. Korogi: Submicron manipulation tools driven by light in a liquid, Appl. Phys. Lett. **82**, 133 (2003)
4.71 Y.Y. Cao, N. Takeyasu, T. Tanaka, X.M. Duan, S. Kawata: 3-D metallic nanostructure fabrication by surfactant-assisted multiphoton-induced reduction, Small **5**, 1144–1148 (2009)
4.72 Y.J. Yan, M.I. Rashad, E.J. Teo, H. Tanoto, J.H. Teng, A.A. Bettiol: Selective electroless silver plating of three dimensional SU-8 microstructures on silicon for metamaterials applications, Opt. Mater. Express **1**, 1548–1554 (2011)
4.73 D.X. Liu, Y.L. Sun, W.F. Dong, R.Z. Yang, Q.D. Chen, H.B. Sun: Dynamic laser prototyping for biomimetic nanofabrication, Laser Photonics Rev. **8**, 882–888 (2014)
4.74 T. Ergin, N. Stenger, P. Brenner, J.B. Pendry, M. Wegener: Three-dimensional invisibility cloak at optical wavelengths, Science **328**, 337–339 (2010)
4.75 H.B. Sun, S. Matsuo, H. Misawa: Three-dimensional photonic crystal structures achieved with two-photon-absorption photopolymerization of resin, Appl. Phys. Lett. **74**, 786–788 (1999)
4.76 B.H. Cumpston, S.P. Ananthavel, S. Barlow, D.L. Dyer, J.E. Ehrlich, L.L. Erskine, A.A. Heikal, S.M. Kuebler, I.Y.S. Lee, D. McCord-Maughon, J.Q. Qin, H. Rockel, M. Rumi, X.L. Wu, S.R. Marder, J.W. Perry: Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication, Nature **398**, 51–54 (1999)
4.77 K.K. Seet, V. Mizeikis, S. Matsuo, S. Juodkazis, H. Misawa: Three-dimensional spiral-architecture photonic crystals obtained by direct laser writing, Adv. Mater. **17**, 541–545 (2005)
4.78 K.K. Seet, V. Mizeikis, S. Juodkazis, H. Misawa: Spiral three-dimensional photonic crystals for telecommunications spectral range, Appl. Phys. A **82**, 683–688 (2006)
4.79 A. Ovsianikov, S.Z. Xiao, M. Farsari, M. Vamvakaki, C. Fotakis, B.N. Chichkov: Shrinkage of microstructures produced by two-photon polymerization of Zr-based hybrid photosensitive materials, Opt. Express **17**, 2143–2148 (2009)
4.80 S.R. Kennedy, M.J. Brett, O. Toader, S. John: Fabrication of tetragonal square spiral photonic crystals, Nano Lett. **2**, 59–62 (2002)
4.81 Q. Sun, S. Juodkazis, N. Murazawa, V. Mizeikis, H. Misawa: Freestanding and movable photonic microstructures fabricated by photopolymerization with femtosecond laser pulses, J. Micromech. Microeng. **20**, 035004 (2010)
4.82 K.K. Seet, V. Mizeikis, K. Kannari, S. Juodkazis, H. Misawa, N. Tetreault, S. John: Templating and replication of spiral photonic crystals for silicon photonics, IEEE J. Sel. Top. Quant. **14**, 1064–1073 (2008)
4.83 Q. Sun, K. Ueno, H. Misawa: In situ investigation of the shrinkage of photopolymerized micro–nanostructures: The effect of the drying process, Opt. Lett. **37**, 710–712 (2012)
4.84 V. Mizeikis, S. Juodkazis, R. Tarozaite, J. Juodkazyte, K. Juodkazis, H. Misawa: Fabrication and properties of metalo-dielectric photonic crystal structures for infrared spectral region, Opt. Express **15**, 8454–8464 (2007)

5. Nanoimprint Lithography

Helmut Schift, Anders Kristensen

Nanoimprint lithography (NIL) is an emerging high-resolution parallel patterning method, mainly aimed towards fields in which electron-beam and high-end photolithography are costly and do not provide sufficient resolution at reasonable throughput. In a top-down approach, a surface pattern of a stamp is replicated in a material by mechanical contact and three-dimensional material displacement. This can be done by shaping a liquid followed by a curing process for hardening, by variation of the thermomechanical properties of a film by heating and cooling, or by any other kind of shaping process using the difference in hardness of a mold and a moldable material. The local thickness contrast of the resulting thin molded film can be used as a means to pattern an underlying substrate at the wafer level by standard pattern transfer methods, but also directly in applications where a bulk modified functional layer is needed. This makes NIL a promising technique for high-volume manufacturing of nanostructured components. At present, structures with feature sizes smaller than 5 nm have been realized, and the resolution is limited by the ability to manufacture the stamp relief. For historical reasons, the term nanoimprint lithography refers to a hot embossing process (thermal NIL). In ultraviolet (UV) NIL, a photopolymerizable resin is used together with a UV-transparent stamp. In both processes thin-film squeeze flow and capillary action play a central role in understanding the NIL process. In this chapter we will give an overview of NIL, with emphasis on general principles and concepts rather than specific process issues and state-of-the-art tools and processes. Material aspects of stamps and resists are discussed. Specific applications are presented, where imprint methods have significant advantages over other structuring methods. We conclude by discussing the areas where further development in this field is required.

5.1 **Emerging Nanopatterning Methods** 115
5.1.1 Next-Generation Lithography..... 116
5.1.2 Nanoprint and Nanoimprint Lithography..... 117

5.2 **Nanoimprint Process** 119
5.2.1 Limits of Molding..... 120
5.2.2 Squeeze Flow of Thin Films..... 120
5.2.3 Residual Layer Thickness Homogeneity . 122
5.2.4 Demolding 122
5.2.5 Curing of Resists and UV-Assisted NIL.... 123
5.2.6 Mix-and-Match Methods and Combination with Directed Self-Assembly..... 124
5.2.7 Multilayer and Multilevel Systems 124
5.2.8 Reversal NIL and Layer Transfer..... 125

5.3 **Tools and Materials for Nanoimprint** ... 125
5.3.1 Resist Materials 126
5.3.2 Stamps 127
5.3.3 Stamp Fabrication and Tooling..... 128
5.3.4 Anti-adhesive Coatings 128
5.3.5 Machines 129

5.4 **Applications**..... 132
5.4.1 Types of Nanoimprint Applications 132
5.4.2 Semiconductor Memory Chips 132
5.4.3 Bit-Patterned (Magnetic) Media (BPM) for Hard Disk Drives..... 133
5.4.4 Subwavelength Metal-Strip Gratings and Transparent Electrodes for Displays 133
5.4.5 High Brightness Light-Emitting Diodes.. 134
5.4.6 Polymer Optics..... 135
5.4.7 Biological Applications 136

5.5 **Conclusion and Outlook**..... 137

References..... 138

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_5

Take a piece of wax between your fingers and imprint your fingerprints into it from both sides. The pressure produced is sufficiently high to replicate the soft surface pattern of your skin into the wax by mechanical deformation. The process is facilitated by the heat resulting from your blood circulation, which softens the wax in order to make it deform until it conforms to the three-dimensional (3-D) pattern of your skin. The pattern resolution of below 1 mm is similar to that of the first records fabricated over 100 yr ago in celluloid. In 1887 *Berliner* applied for a patent on a so-called gramophone, which resembles Edison's phonograph with its wax-coated roll [5.1]. The information is inscribed into wax coated onto a zinc disk. The tracks are cut through the wax down to the solid zinc and are etched before using the zinc disk as a mold to press thermoplastic foils. With a playing time of little more than 1 min, those disks had track widths below 1 mm and resolution in the sub-100 μm range. Over the years the track size was reduced to below 200 μm. The materials changed from shellac to vinyl filled with carbon black. Today's digital versatile discs (DVDs) have pit sizes of below 400 nm and are fabricated in polycarbonate (PC) in a few seconds by injection molding. Disc formats such as blu-ray (BD) with further reduced pit sizes of 100 nm are the current state of the art [5.2, 3].

In this introduction some basic concepts of molding polymers are illustrated, ranging from the shaping of thin films by mechanical pressure to patterning of curable liquid resins by capillary action. The process is closely linked to the choice of stamps, molding materials, and needs to take into account the type of pattern transfer after molding is completed. A softened material can be deformed by pressure and the difference in mechanical properties makes it possible to replicate its surface pattern in a parallel, reproducible way. The squeezing of a thin film of wax leads to a lateral flow of material, but because of the high viscosity the process will slow down quickly and a residual layer, which cannot be thinned down to zero, will remain. Furthermore, the softness of the stamp and the viscosity of the material will determine the completeness of molding and thus the replication fidelity. Similar concepts of molding processes can be observed in daily life, such as imprinting a footprint into snow or clay, or making waffles by casting a liquid dough between two patterned iron plates with subsequent thermocuring, or by replicating a seal into wax (Fig. 5.1). One common important prerequisite of these molding processes is that the mechanical properties of the molded material can be changed by pressure, temperature or chemical processing. The material must be shaped in a viscous state but should keep its form during demolding. The imprint in snow is a hard molding by local densification, while the clay hardens by the squeezing out of water. The waffle is cured due to the thermochemical changes in the dough, and the seal can be demolded with high fidelity because the heat of the wax dissipates into the seal and the wax hardens during cooling. The processes described here are very similar to molding of viscous thermoplastic materials in the nanoimprint lithography (NIL) process [5.4, 5], where a thickness profile in a thin polymer film is generated by pressure, however, with the surprising difference that features below 10 nm can be replicated with unprecedented precision (Fig. 5.2). In contrast to conventional methods based on exposure and development, limitations imposed by the wavelength of exposure or by chemical reactions can be overcome.

The example with the fingerprint may even serve for the illustration of ink transfer in microcontact printing (μCP), as one example of soft lithography (SL) [5.6, 7]. While for NIL a hard stamp would ensure a more complete molding, here the softness of the stamp is essential to make a conformal contact of every protrusion, but at the expense of a possible reduction of the feature resolution due to local deformations in the stamp.

In this chapter we provide an overview of the different processes currently called nanoimprinting, from hot embossing of thermoplastic materials to molding and curing of liquid resins. After this introduction into the basics of molding, Sect. 5.1 places the two main NIL techniques into the context of the emerging nanopatterning methods for lithography and their difference to μCP. Section 5.2 describes the nanoprint and nanoimprint processes, beginning with a discussion about soft and hard elements in mechanical patterning, polymer properties and pattern transfer and their use in thermal and UV-assisted patterning of resists. It discusses the thermal NIL process as an important subset of processes, giving an insight into squeeze flow of thin films. Section 5.3 presents essential fabrication meth-



Fig. 5.1 Printing a seal into viscous wax is a way of replication using hot embossing. The figure shows a seal (stamp), wax tube (candle), and embossed pattern

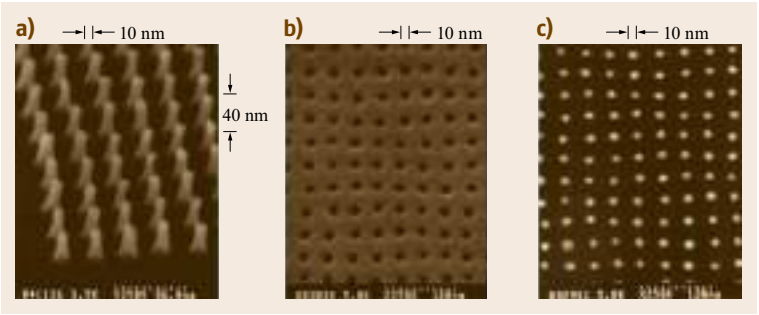


Fig. 5.2a–c Micrographs showing the basic steps of NIL, demonstrated by *Chou* and *Krauss* [5.4]. **(a)** NIL stamp in silicon with a 40 nm-period array of pillars with 40 nm height, **(b)** imprinted 10 nm-diameter holes in a thin polymer film (PMMA), **(c)** 10 nm metal dots after pattern transfer (lift-off), using the thin polymer layer as a mask

ods for stamps and tools for NIL. It discusses current materials and tools for NIL, ranging from stamps and resists to NIL machines. Section 5.4 presents typical applications that are currently envisaged both at an industrial and at a laboratory scale. Although for many people the main driving force behind NIL is its use as next-generation lithography (NGL) for complementary metal-oxide-semiconductor (CMOS) chip fabrication, the reader will be introduced to other applications that do not have the demanding overlay requirements imposed by multilevel processes. We conclude with an outlook in Sect. 5.5, where we discuss the prospects of NIL and aspects of its commercialization. Further information can be found in the references, but not least within this handbook in the chapters about silicon micromachining (Chap. 3) and soft printing (Chap. 6). In this chapter we restrict ourselves to the lithographic patterning of thin films on hard substrates. We present basic concepts rather than state-of-the-art tools and hot scientific issues. As a complement to this chapter, the reader is advised to refer to three publications. A review on NIL [5.8] deals with a range of process issues relevant for research and industry, and a deeper insight into advanced concepts of printing. Specific NIL processes and process flows for a variety of applications are presented in the NaPa (NanoPattering) library of processes (NaPa LoP) [5.9]. After 20 yr of NIL, in which NIL has become a technology much beyond infancy, other reviews have been presented [5.10, 11]. Finally, there are book chapters that refer to manufacturing-related issues of NIL technologies [5.12].

5.1 Emerging Nanopatterning Methods

Nanoimprint Lithography (NIL) is a replication technique that has been proven to provide a resolution unmatched by many other techniques, while at the same time offering parallel and fast fabrication of micro- and nanostructures. On the one hand, this enables the technique to branch out into fields where large areas covered by nanostructures or a number of identical structures for statistical evaluation are needed. This was often impossible due to the low throughput of serial lithographic research tools. On the other hand, the resolution achieved so far by molding is still higher than that used in the industrial manufacturing of processors and memory chips with high-end photolithography (PL). This makes NIL a promising technology for NGL. Apart from these advantages molding offers more: by creating a 3-D resist pattern by mechanical displacement of material, the patterning of a range of specific functional materials and polymers becomes possible, without loss of their chemical properties during molding. Furthermore, this ability can be used to fabricate complex structures e.g., by building up devices with embedded channels or patterning of functional materials e.g., for polymer electronics. These processes are presented in more detail in Sect. 5.2 of this chapter.

In this section we present the basic concepts of NIL and how it can conform to the requirements of state-of-the-art nanofabrication techniques. NIL uses, as do other lithographic techniques, the concept of resist patterning (which can be also found in different chapters in this Handbook). The resist patterns are generated by molding of a viscous material and hardened by cooling and curing, while in PL the resist is patterned by selective local chemical modification of a positive or negative resist by exposure and subsequent wet development. The two main NIL methods are outlined in Fig. 5.3. For lithographic applications, as needed in microelectronics and optics, NIL is in competition with other emerging patterning techniques, which are predominantly probe-based approaches (electrons, photons). Its success will mainly depend on the ability to solve processing issues such as resolution and throughput. It is also important to develop reliable tools with

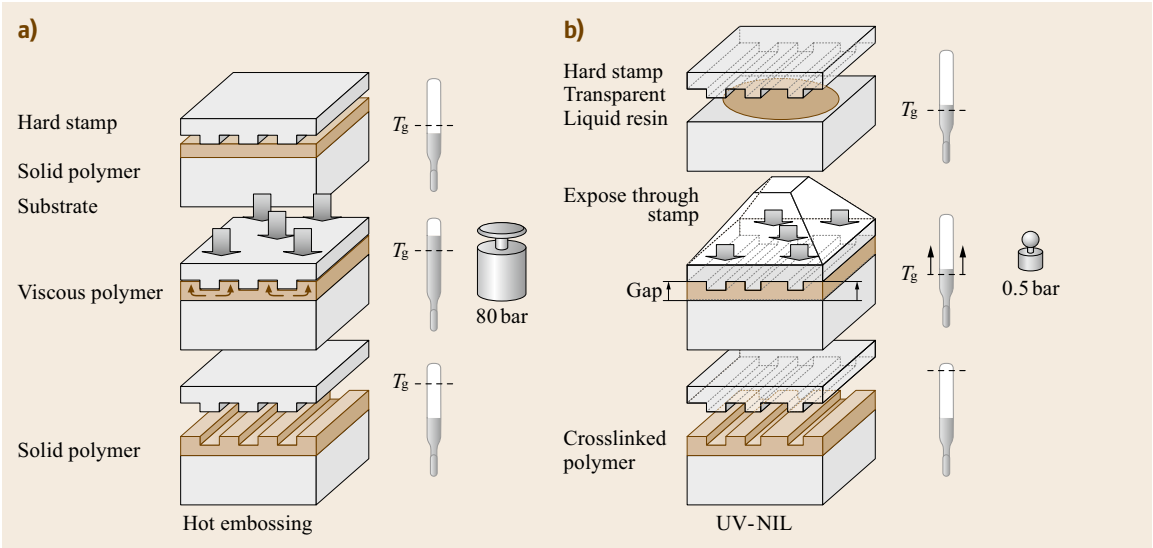


Fig. 5.3a,b Schematic of NIL process: **(a)** thermal NIL (hot embossing) and **(b)** UV-NIL. In both cases a thickness profile is generated in the thin polymer layer. After removing the residual layer, the remaining polymer can serve as a masking layer that can be used as a resist for pattern transfer

a long lifetime, which can be used in combination with other cleanroom process technologies, and to establish standard processes, which can be scaled up to common wafer and substrate sizes.

5.1.1 Next-Generation Lithography

With its integration into the International Technology Roadmap for Semiconductors (ITRS) on NGL in 2003 for the 32 nm node and beyond, NIL has become more than a simple high-resolution method [5.13] (Table 5.1). Since then it has been considered as a candidate for replacing or complementing advanced optical lithographic methods for the fabrication of processors and solid-state memory chips, which over the years have been developed and pushed to higher resolution with a vast investment of resources. Over more than 50 yr, Moore's law has described with amazing accuracy the reduction of feature sizes (and cost per transistor), and therefore serves as a roadmap for the developments needed for future microchips [5.14]. It is driven by economic considerations, and leads to a competition between different candidate manufacturing methods. These do not only have to provide the resolution of smallest feature sizes (node), but also issues such as alignment (overlay of several masking levels), critical dimensions (CD), simple mask fabrication, high throughput (mass manufacturing), and low defectivity, resulting in low cost of ownership (CoO, i. e., no dependence on large machines such as synchrotrons, back-ups and redundancy of tools and masks), which become increasingly difficult to meet if smaller exposure wavelengths have to be used. To date (2017), the semiconductor chip industry has reached 11 nm, which shows the ability of traditional techniques to advance towards smaller nodes, which in the case of 193 nm (DUV) PL was possible by immersion of the lens-resist gap with liquid and self-aligned multiple (mask) patterning techniques. NIL is no longer considered as the only candidate with the ability to reach the requirements of the next few nodes, but still has cost advantages when compared to the huge investments made for extreme ultraviolet (EUV) lithography. As a replacement or complement to existing NGL techniques, it will probably be used in combination with bottom-up techniques for resolution enhancement (mostly for stamps) or functional surfaces. However, as a 1X technique (no projection) the stamp will need to have the same feature sizes as the features to be generated on the device.

As seen in the last years, even huge financial and apparent physical limits have been overcome in developing several techniques towards the next nodes. However, in the case of NIL there is the hope that for chip manufacturing it will be more cost-effective than other techniques. The investments are expected to be lower than for the current front runners, EUV PL or parallel electron-beam exposure. In many other applications, where the introduction of a single lithographic step of 20 cents per masking layer presents a huge investment, alternatives such as NIL are considered as the only way out of this spiral of rising investments.

Table 5.1 ITRS roadmap showing the resolution of different lithographic patterning techniques, and practical and actual resolution limits for different lithography methods. From [5.13], revised and updated

Lithography type	Practical resolution limit (nm)	Ultimate resolution limit (nm)
UV/contact/proximity photolithography	2500	125
Deep UV projection (DUV)	20–40	11 (multiple patterning)
EUV projection (soft x-rays, 13.6) with mask	20–40	7 (multiple patterning)
EUV interference lithography (13.6/6.8)	20	7/4
Ion beam	20–40	Resist: 10–20
Electron beam (low energy beam arrays)	20–40	Resist: 7–20
Thermal nanoimprint (hot embossing)	10–40	0.2–5
UV nanoimprint (hard stamp)	10–40	0.2–5
UV nanoimprint (soft stamp)	50–100	10
Soft lithography (contact printing)	50–500	10
Scanning probe microscopy methods	15	0.5 (atomic resolution)

5.1.2 Nanoprint and Nanoimprint Lithography

Micro-Contact Printing – Ink Transfer by Conformal Contact

NIL is a high-throughput, high-resolution parallel patterning method in which a surface pattern of a stamp is replicated in a material by mechanical contact and 3-D material displacement. Nanoprinting (known as microcontact printing (μCP)), in contrast to nano-*im*printing, transfers an ink from stamp surface protrusions onto a substrate by mechanical contract. Such an ink consists of (linear) chain molecules with the ability to adhere or bind to surface atoms and often create monomolecular films which – due to their density and chemical properties – can act as resists [5.6, 7]. However, such films, in the case of self-assembled monolayers (SAMs) are only 1–2 nm thick, and present chemical contrast against specific etchants. This process has demonstrated sub-100 nm resolution and in principle can pattern much smaller structures, if lateral diffusion of densely printed molecules is inhibited. While NIL is clearly directed towards high-resolution applications, μCP is for research applications mostly in the area of biotechnology. The invention of the technology dates back to 1994 and has similarities to flexographic printing approaches.

The term *soft lithography* (SL) describes a quality of the mold – compliancy and often also elasticity – which is also characteristic to NIL variants such as capillary-assisted molding. SL uses a soft stamp that is able to conform to nonideal substrates i. e., to different length scales of surface height variations. This can be given by surface undulations and deviations in structure height due to roughness, local surface patterns and particles, as well as thickness variations (*waviness*) over the entire substrate area. Due to the compliancy of both its micro- and nanosized stamp protrusions and its backbone (typically a plate several 100 μm to some mm thick) the stamp is able to close local gaps and even *seal* cavities from each other. The compliancy needed for *conformal printing* is however also a matter of pressure and the relative softness of the stamp towards the imprinted material and substrate.

Molding of Thermoplastic Resists by Thermal NIL

NIL was first reported as thermoplastic molding [5.15–17], and is therefore often referred to as thermal NIL (here also named T-NIL) or hot embossing lithography (HEL) [5.5, 18, 19]. The unique advantage of a thermoplastic material is that the viscosity can be changed to a large extent by simply varying the temperature. Figure 5.4 shows viscosity plotted against temperature for various thermoplastic polymers i. e., poly(methyl methacrylate) (PMMA) with different molecular weight, and some commercial resists [5.8, 20]. Switching between solid and a highly viscous state is possible within a range of some tens of degrees Celsius. The first stage of the NIL process is the molding of a thin thermoplastic film using a hard master. During a process cycle the resist material is made viscous by heating and shaped by applying pressure (Fig. 5.5). Here the thermoplastic film is compressed between the stamp and substrate and the viscous polymer is forced to flow into the cavities of the mold, conforming exactly to the surface relief of the stamp. When the cavities of the stamps are filled, the polymer is cooled down, while the pressure is maintained. Thus the molten structure is frozen. After relieving the pressure, the stamp can be retrieved (demolded) without damage, and reused for the next molding cycle.

Molding of UV-Curable Resists by UV-NIL

With the integration of light sources into imprint machines, UV-NIL was developed for curable resists [5.21–24]. The basic difference between UV-NIL and T-NIL is that a resin, which is liquid at room temperature, is shaped at a moderate pressure, and by exposing light through the transparent stamp, the resin is crosslinked and hardened (similar to negative tone resist

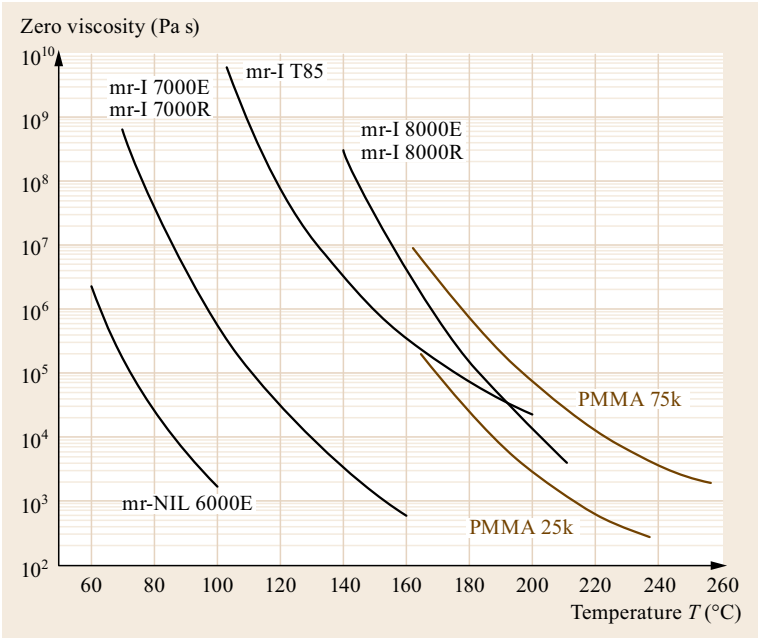


Fig. 5.4 Graph showing zero shear viscosities for some standard resists for thermal NIL for different polymers, taken from different sources: PMMA with M_w of 25–75k and commercial resists (after [5.8, 9]). It shows the potential of rheology and of the large variation of viscosity of thermoplastic polymers with temperature. These curves are presented for the temperature range characterized above the viscous T_g. A process window for imprint is limited by high viscosity where unwanted viscoelastic effects become dominant and molding slow. Lower viscosities than 10^3 Pa s are often not useful because it is often achieved with too low M_w or too high $T_{imprint}$ (after [5.8])

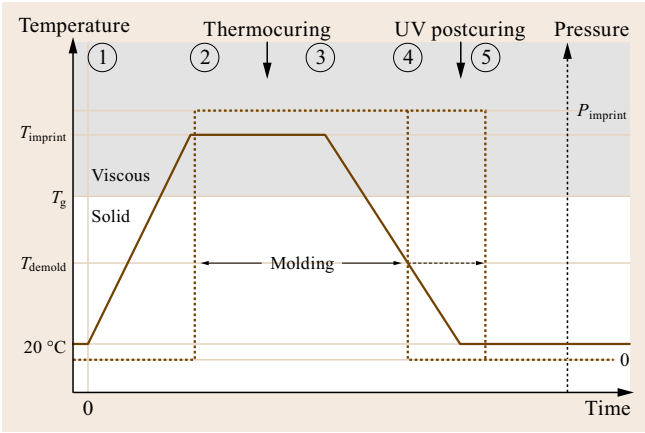


Fig. 5.5 Typical process sequence: schematics of process sequence used for hot embossing (temperature–pressure diagram with time dependence), (1) begin heating, (2) begin embossing, (3) begin cooling, (4) demolding at elevated T, (5) demolding at ambient T

processes). The stamp either sinks down to the substrate or must be kept at constant distance from the substrate during both filling and exposure, due to the low resist viscosities. The mechanical setup has to be able to compensate for wedge errors in a low-imprint-pressure process. Patterning on nonflat substrates or over topography therefore requires a planarization strategy and often small stamps. Because of the small pressures, stamps made of soft materials on a rigid backbone can be used. These soft imprint approaches (*soft lithography*) have become popular for large-area imprint and small and moderate volume manufacturing; e.g., patterned sapphire substrates (PSS). Hard stamps are used in the so-called jet-and-flash imprint lithography (JFIL) approach, favored by Canon Nanotechnologies (CNT, formerly Molecular Imprints Inc., MII) for semiconductor chip manufacturing [5.25].

Pattern Transfer: A Two-Step Process with Window Opening and Conversion

In many cases the lithographic process is only complete when the resist pattern is transferred into another *functional* material. The dimensions (lateral structure size and height) are often similar to the film thickness used. In NIL, in comparison to standard electron (EBL) or photolithography (PL), pattern transfer is a two-step process. First the thin polymer film is transformed into a masking layer. For this, the residual layer remaining in the thin areas of the resist has to be removed, which is done by homogeneously thinning down the entire resist in an (ideally) anisotropic etching process. This way, process windows are opened to the substrate, which allows the substrate to be altered by attack by plasma, etching solutions, deposition of materials and other changes. Because the thinning down of the resist for window opening is also a cleaning step (called descum), it is similar to the wet development of the exposed resist used in PL.

There are an increasing number of process variations, which are mostly variants of the established T-NIL and UV-NIL processes, particularly those using special methods of pattern transfer (e.g., re-

Table 5.2 Comparison of hot embossing (NIL) and UV-imprint (UV-NIL), with typical parameters of current processes

Type of NIL/properties	Thermal NIL (hot embossing (HEL))	UV-NIL (hard stamp)
Basic process sequence (Fig. 5.5)	1) Spin-coat thermoplastic film 2) Place stamp on film 3) Heat until viscous 4) Emboss at high pressure 5) Cool until solid 6) Demold stamp	1) Dispense liquid resin 2) Parallel alignment of stamp with defined gap 3) Imprint at low pressure 4) Expose with UV light through stamp and crosslink 5) Demold stamp
Pressure p (bar)	20–100	0–5
Temperature T_{mold} (°C)	100–200	20 (ambient)
Temperature T_{demold} (°C)	20–80	20 (ambient)
Resist	Solid, thermoplastic $T_g \approx 60$–$100\,°C$	Liquid, UV-curable
Viscosity η (Pa s)	10^3–10^7	10^{-2}–10^{-3}
Stamp material	Si, SiO_2 (opaque)	Glass, SiO_2 (transparent)
Stamp area	Full wafer, > 200 mm diameter	Limited by control of gap Gen2 panel $370 \times 470\,mm^2$
Stamp contact	Facilitated by bending	Planarization layer
Embossing time	From seconds to minutes	< 1 min (per exposure)
Advantage	Low-cost, large-area equipment and stamps	Low viscosity, low pressure, alignment through stamp
Challenge	Process time, thermal expansion due to thermal cycle	Step and repeat needed for large areas
Development needed	Alignment, residual layer homogeneity	Material variety
Step and repeat	Step and stamp with $4 \times 4\,mm^2$ stamps	Jet and flash (JFIL) with $33 \times 26\,mm^2$
Hybrid approaches	Thermoset resists: embossing and curing before demolding	Thermoplastic resists: hot molding and UV-curing before demolding
Advantage	Low temperature variation cycle: demolding at high temperature possible	Solid resist: single-step wafer-scale imprint possible

versal imprint) and hybrid processes (combinations of different processes). All the processes have their specific advantages, e.g., while UV-NIL can be performed at room temperature, hot embossing is often low-cost since nontransparent silicon stamps can be used. The major characteristics of typical processes, along with those of hybrid approaches, are summarized in Table 5.2 and presented in more detail in [5.8].

5.2 Nanoimprint Process

Molding techniques based on imprint processes make use of the differences between the mechanical properties of a structured stamp and a molding material. The viscous molding material is shaped by pressing the hard stamp into it, until the polymer conforms to the stamp surface. In hot embossing processes we mostly deal with thermoplastic materials whose mechanical properties can be repeatedly and reversibly changed from a solid into a viscous state by simply varying the temperature. In order to achieve a reasonable process time and yield, this is normally carried out under high pressure. Thermal NIL deals with a viscosity range that is considered as sufficiently low to enable significant squeeze flow over large distances, but high enough that bending of wafers can be used to equilibrate surface undulations of common substrates and pattern density variations in stamps. While squeeze flow governs high-viscosity molding (where pressure is the driving force to displace the viscous material), in UV-NIL low pressure or even mold filling by mere capillary action (where surface energy controls the wetting and spreading of the viscous material) is possible if very-low-viscosity resins are used. In this section we want to take a closer look at the squeeze flow of thin polymer films as used for thermal NIL, a concept that is quite general and enables us to get an insight into possible parameter variation for process optimization. We will give a brief introduction to the theory of polymers and discuss the implications for NIL [5.26, 27]. This will enable the reader to understand rheology in NIL from a practical point of view. More fundamental questions of squeeze flow are discussed in [5.28] or in [5.29]. In the section on curable resists, we will introduce concepts mainly used in UV-NIL, such as soft UV-NIL.

5.2.1 Limits of Molding

Polymers used as resists in NIL are defined by their chemical composition and physical properties. In the case of molding these are often long-chain molecules with a molecular weight M_w. The polymer M_w is important because it determines many physical properties. Some examples include the temperatures for transitions from liquids over viscoelastic rubbers to solids, and mechanical properties such as stiffness, strength, viscoelasticity, toughness, and viscosity. However, if the M_w is too low, the transition temperatures will be too low and the mechanical properties of the polymer material will be insufficient to be useful as hard resist for pattern transfer, e.g., in reactive ion etching, metal evaporation or electroplating.

It has been known for a long time that polymers can replicate topographies with high fidelity. Up to now, 5 nm resolution of polymer ridges with a pitch of 14 nm has been demonstrated, but even atomic steps in sapphire wafers have been replicated. In comparison to methods like electron-beam lithography (EBL), where a nanoscale chemical contrast can be can be produced by local irradiation-induced chain scission, the polymer chains are only moved and deformed during molding, thus keeping their properties such as M_w. Molding topographic details down to a few nanometers means that single polymer chains have to deform or flow.

A polymer is a large molecule made up of many small, simple chemical units, joined together by chemical reaction. For example, polyethylene $CH_3{-}(CH_2)_n{-}CH_3$ is a long chain-like molecule composed of ethylene molecules $CH_2{=}CH_2$. Most artificially produced polymers are a repetitive sequence of particular atomic groups, and take the form [–A–A–A–]. The basic unit A of this sequence is called a monomer unit, and the number of units n in the sequence is called the degree of polymerization. The molecular weight M_w of a polymer is defined by the weight of a molecule expressed in atomic mass units (amu). It may be calculated from the molecular formula of the substance; it is the sum of the atomic weights of the atoms making up the molecule. For example, poly(methyl methacrylate) (PMMA), a classic resist material, exhibits very good resolution for both EBL and NIL. A high-M_w PMMA, typically above 500 kg/mol (also denoted 500k), is normally used for EBL, since the development contrast between exposed and unexposed areas increases with M_w [5.30]. A lower M_w, of some tens of kg/mol, is patterned in NIL, due to the strong increase in temperature-dependent viscosity with M_w. Apart from their mobility it is expected that shorter chains, which in the case of amorphous polymers are normally present as entangled coils, can move more easily into small mold cavities. A convenient way of expressing the size of a macromolecule present as a statistical coil aggregate is the radius of gyration R_g.

A polymeric liquid, whilst retaining the properties of a liquid, follows a rubber-like elasticity. Substances with both viscous and elastic properties are called viscoelastic. Considering the fact that that molding is achieved by deformation of a polymer network at a molecular level, the question is how the polymer can be permanently shaped and whether the replicated structure will deform back due to internal reordering and relaxation of polymer chains.

The reduced viscosity of polymers at higher temperatures is a result of the increasing ability of the chains to move freely, while entanglements and van der Waals interaction of the chains are reduced. The glass transition of a thermoplastic polymer is related to the thermal energy required to allow changes in the conformation of the molecules at a microscopic level, and above T_g there is sufficient thermal energy for these changes to occur.

Most of our considerations here are valid for a range of practical process parameters, as used in current thermal NIL processes, where linear behavior can be assumed (Newtonian flow regime). This is in particular the case at molding temperatures well above T_g. For thermoplastic molding, however, the T_g is only a rough indication of a temperature for fast molding. More suitable than T_g is the flow temperature T_f, which characterizes a point at which viscosity drops to practical values for molding needed for fast NIL (i. e., $10^3{-}10^7$ Pa s, about 50 °C above T_g for 25k PMMA, Fig. 5.6) [5.8].

5.2.2 Squeeze Flow of Thin Films

During imprinting, linear movement of a stamp is transformed into complex squeeze flow of the viscous material. In the thin polymer films used in NIL, a small vertical displacement of the stamp results in a large lateral flow. The two surfaces of the stamp and the substrate have to come entirely in contact with each other and keep this contact until the desired residual layer thickness is reached. Furthermore, new concepts are possible such as roll embossing (roll-to-roll NIL) and soft imprint using flexible stamps. In Fig. 5.7, the imprint of a stamp with line cavities is schematically shown.

Before imprint, the polymer film has an initial thickness h_0 and the depth of the microrelief is h_r. For a fully inserted stamp, the film thickness under the single stamp protrusions (elevated structures) with width s_i is h_f. We can calculate this specific residual layer height h_f by applying the continuity equation with the

assumption that the polymer melt is incompressible (conservation of polymer volume). It can be directly deduced from the area fill factor ν, i. e., ratio of the area covered by cavities to the total stamp area.

A simple model for the squeezed polymer flow underneath the stamp protrusion is obtained by treating the polymer as an incompressible liquid of constant viscosity, and solving the Navier–Stokes equation with nonslip boundary conditions at the stamp and substrate surfaces. According to this model, given for line-shaped stamp protrusions and cavities in [5.8, 31], we find the following expression, known as the Stefan equation, for the film thickness $h(t)$ underneath the stamp protrusion when a constant imprint force F is applied to the single stamp protrusion

$$\frac{1}{h^2(t)} = \frac{1}{h_0^2} + \frac{2F}{\eta_0 L s^3}\, t\,. \tag{5.1}$$

Inserting the final thickness $h_\mathrm{f} \equiv h(t_\mathrm{f})$ in (5.1) gives the imprint time

$$t_\mathrm{f} = \frac{\eta_0 L s^3}{2F}\left(\frac{1}{h_\mathrm{f}^2} - \frac{1}{h_0^2}\right)\,. \tag{5.2}$$

For many practical cases a constant pressure under each stamp protrusion $p = F/(s\,L)$ can be assumed.

As a direct consequence of the Stefan equation it can be seen that, at identical pressure, small (narrow) stamp protrusions will sink faster than large (wide) ones. The stamp geometry can therefore be optimized by reducing the dimensions of the protrusions. While stamps with nanopillar arrays, as shown in Fig. 5.2, would allow fast imprint of some microseconds, using standard NIL process parameters, already protrusions of some hundreds of micrometers would increase imprint times to some hours. The strong dependence of the imprint time on the pressing area has the consequence that, for a fully inserted stamp relief (full contact over the total stamp area), the flow practically stops (schematically shown in Fig. 5.8). For this case, s becomes large and flow continues only towards the stamp borders. It is also evident that there is only weak influence of the imprint force ($t_\mathrm{f} \propto 1/F$). At first sight there is a similar weak influence for η_0. However, the viscosity can be changed significantly by varying the temperature. For practical use, it is quite important that trade-offs are possible between structure height, resist thickness, pressure, and temperature. For example, within certain limits, a low imprint pressure can be compensated by a longer time or a higher temperature.

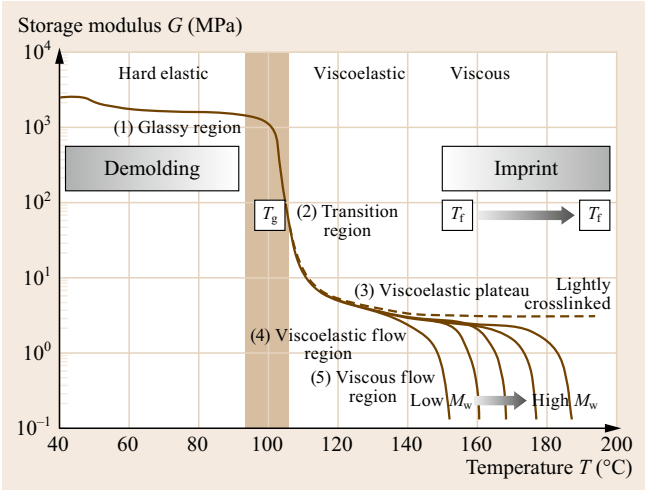


Fig. 5.6 Mechanical properties of polymers dependent on temperature, molecular weight, and crosslinking (after [5.26]). Schematic for a polymer with a T_g around 100 °C for normal process conditions. Particularly important for thermal NIL are the large drops of G at two temperatures, T_g and T_f. At T_g the thermomechanical properties between stamp and polymer become sufficiently different for repeated molding. T_f characterizes a point at which viscosity drops to practical values for molding needed for fast imprint

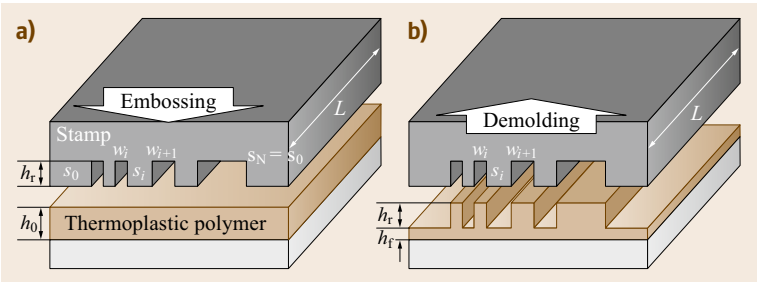


Fig. 5.7a,b Geometrical definitions used for the description of the flow process for a stamp with line cavities and protrusions: **(a)** before molding, and **(b)** after demolding. In the case of viscous molding, where volume conservation can be anticipated, the residual layer thickness can be calculated from geometrical parameters such as the initial film thickness and the size and density of cavities

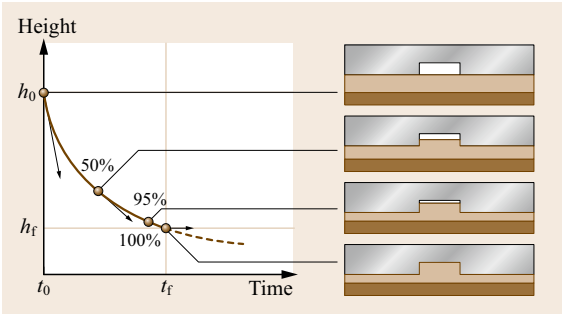


Fig. 5.8 Schematics (*right*) of the squeeze flow of a compressed polymer film into one cavity. Once the cavity is filled the stamp continues to sink but at a much slower rate (*left*), as a direct consequence of the Stefan equation

The implications of squeeze flow are discussed in more detail in [5.19, 32], including rheological issues, bending of stamps in large-area imprint and influence of vacuum and self-assembly [5.33–36].

It is has to be noted that while the Stefan equation governs the sinking of the stamp on the basis of squeeze flow, very small stamp protrusions with high aspect ratio deviate from this behavior. The slowing down is then more governed by the Stokes drag, which does not allow the polymer to proceed upwards in an narrow cavity due to the reduced wetting of the sidewalls [5.36]. In contrast to this, for large stamp protrusions, bending of the entire stamp e.g., in areas of high density of cavities where sinking is enhanced, enables printing faster than predicted by the Stefan law.

5.2.3 Residual Layer Thickness Homogeneity

The main difference between NIL and lithography based on exposure and development is that a residual layer below the stamp protrusions is left after demolding. As seen before, this is a result of the molding process slowing down due to the squeeze flow. For many applications, when pattern transfer has to be achieved after the imprinting, it is important to determine the final residual thickness h_f of this polymer layer (Fig. 5.7) before the next process step. Furthermore it is important to know the thickness variation over the imprinted area. Otherwise, parts of the structure will be lost during pattern transfer. As will be shown in the following, bending of stamps has to be taken into account, as well as effects such as air inclusions, dewetting and self-assembly of resist [5.35]. In most cases a homogeneous residual layer can be achieved by optimizing the pattern design, but also by using adapted processes that create thin residual layers independent from design. In contrast to this, pattern transfer processes that are insensitive to thickness variations have to be used, e.g., by using a resist with high etch resistance or an intermediate layer as a hard mask. The following examples demonstrate how soft and hard elements for equilibration are used to achieve a homogeneous molding.

5.2.4 Demolding

During demolding the rigid stamp is detached from the molded structure, which can be in a parallel way when using small, rigid stamps, or by delamination, if thin, wafer-like substrates are used. If fully molded, the thickness profile in the resist exhibits the inverse polarity of the relief of the stamp surface. The demolding process, also called de-embossing, is normally performed in the frozen state, i. e., when both mold and molded material are considered solid. For thermoplastic materials this happens at a temperature well below T_g, but high enough that frozen stress due to thermal contraction does not lead to damage during demolding. In such cases in which the resist is cured before demolding i. e., crosslinked by exposure or heat, demolding can take place at temperatures similar to the molding temperature. A successful demolding process relies on a controlled balance of forces at the interfaces between stamp, substrate and molded polymer film. Therefore mechanical, physical and chemical mechanisms responsible for adhesion have to be overcome. The following effects have to be avoided or reduced [5.37] (Fig. 5.9):

- Undercuts or negative slopes in the stamp may lead to mechanical interlocking of structures, which in their frozen state are elastically elongated and deformed before ripping. Sidewalls with positive or at best vertical inclination are prerequisite for demolding without distortion.
- Friction due to surface roughness may occur during the sliding of molded structures along vertical cavity walls. The effect of this can only be overcome if the surface of the molded material is elastic and enables gliding of the wall without sticking.
- The enlarged surface area of the patterned stamp leads to an increase of hydrogen bridges and van der Waals forces, or other chemical bonding effects due to ionic, atomic and metallic binding. This effect can only be overcome if the stamp surface can be provided with sufficient anti-adhesive properties.

The most critical point is that demolding forces largely depend on the geometry of the mold, and the overall design of a stamp structure has to be taken into account. Therefore structures with high aspect ratio may be more prone to ripping, and if many neighboring

structures exert high forces to the underlying substrate, whole areas of resist may be detached from the substrate surface. Anti-adhesion layers on the mold can reduce friction forces, but have to be thin and durable. In thermal NIL the expansion coefficient of the substrate $\alpha_{substrate}$ and of the stamp α_{stamp} should be similar, to avoid a distortion due to mechanical stress induced by the cooling. In the case of very thin polymer layers, the lateral thermal expansion of the resist is determined by the substrate. For structures with a higher aspect ratio the demolding temperature T_{demold} should be well below T_g to enable the demolding of a hardened resist without distortion, but as near as possible to T_g, because the stress induced by thermal shrinkage should not be exceeded in critical areas where structures tend to break.

5.2.5 Curing of Resists and UV-Assisted NIL

Curing by UV exposure, by thermal treatment or chemical initiation is a way to crosslink polymers and to make them durable for demolding [5.21, 22, 38–43]. A high reaction speed, as caused by a high exposure dose, high initiator content or curing at high temperatures, leads to fast, but weak crosslinking, whereas a slow reaction leads to highly polymerized, tougher materials because the slow polymerization enables a more complete process. As shown in Sect. 5.1.2, different process strategies have been developed. In most of them the curing step is independent from the molding step, and can be initiated by light or a specific temperature after the molding is complete. Because curing involves a change in the physical conformation of the polymer, it always goes along with a volumetric shrinkage of the polymer. For example, acrylate polymerization is known to be accompanied by volumetric shrinkage that is the result of chemical bond formation. Consequently, the size, shape, and placement of the replicated features may be affected. In the following the main processes that involve curing are presented in more detail:

- In the UV-NIL process, the resist is cured after molding, but before demolding of the stamp. The process relies on the photopolymerization of a low-viscosity, acrylate-based solution. Shrinkage was found to be less than 10% of the total volume in most cases. The liquid is a multicomponent solution. The silylated monomer provides etch resistance in the O_2 transfer etch, and is therefore called an etch barrier. Crosslinking monomers provide thermal stability to the cured etch barrier and also improve the cohesive strength of the etch barrier. Organic monomers serve as mass-persistent components and lower the viscosity of the etch barrier formulation. The photoinitiators dissociate to form radicals upon UV irradiation, and these radicals initiate polymerization.
- Polymer chemistries with oxygen inhibition do not cure within ambient atmosphere. Therefore stamps with inherent permeability to gas (e.g., PDMS) are not usable. When using hard stamps, resin between stamp and substrate will cure upon exposure, while the excess material on the stamp borders will stay soft. Many UV-NIL resists are noxygen non-inhibiting.
- If a solid curable resist exhibits thermoplastic behavior, it can be molded at an elevated temperature and then crosslinked, either before or after demolding. The advantage of the process is that low-M_w resists with a low T_g can be provided, which can be processed at moderate temperatures. However, before pattern transfer, a hardening is often necessary. They can also be used for mix-and-match with PL or for polymeric stamp copies.
- Thermoset resists can be crosslinked by heat. Here it is of advantage that the temperature for molding is lower than the curing temperature. Then the structure is first molded and then heated to its crosslinking temperature to induce crosslinking, before the stamp is demolded from the hardened surface relief.

More information about curing and multilayer resists can be found in Sect. 5.3 of this chapter.

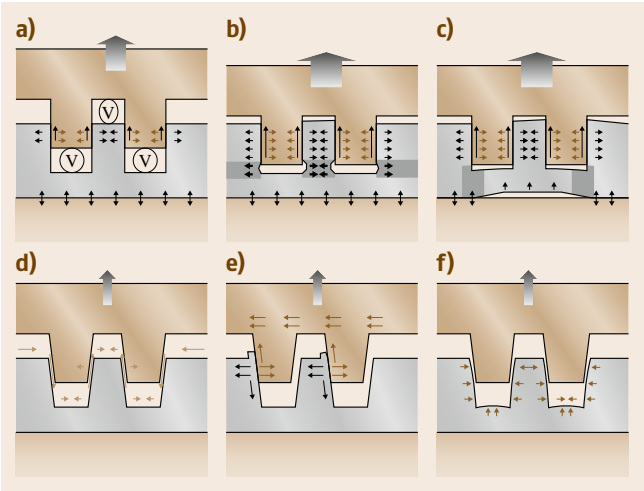


Fig. 5.9a–f Demolding issues: **(a)** generation of vacuum voids (V), **(b)** elongation and ripping of single structures, **(c)** ripping of resist from substrate, **(d)** penetration of air into voids (inclined sidewalls), **(e)** shrinkage and generation of rims, **(f)** relaxation of frozen-in-strain (after [5.8])

5.2.6 Mix-and-Match Methods and Combination with Directed Self-Assembly

Mix-and-match approaches are used to combine the advantages of two or more lithographic processes or simply to avoid their mutual disadvantages [5.44–46]. It is also a way to improve throughput and reliability, e.g., since the fabrication of large-area nanostructures is often costly, the definition of microstructures can be done with PL, while the nanopatterning of critical structures in small areas is done by NIL. In many cases NIL would be used as the first process step and by adding alignment structures along with the nanopatterns, the less critical structures can be added after the NIL step using PL with an accuracy given by the mask aligner (in the range of 1 μm).

5.2.7 Multilayer and Multilevel Systems

Multilayer resists systems are used if the etching selectivity of a masking layer has to be enhanced, e.g., for the fabrication of high-aspect-ratio structures, or undercuts to be generated, e.g., when lift-off or a planarization layer has to be employed for printing over topography [5.47, 48].

The most important application of double resists is for low-pressure processes such as UV molding (Fig. 5.10). For prestructured substrates with topography, a planarization layer is needed, because the low pressure of below 1 bar is often not sufficient to achieve a conformal contact of the transparent mask with the nonflat substrate surface. Otherwise parts of the resist stay unmolded.

Often top layers with high etching resistance, e.g., silicon-containing resists (similar to hardening by silanization) are chosen for UV-NIL. After molding the top layer, the pattern is transferred into the underlying planarization (transfer) layer. The top layer can be kept thin, while the etching depth can be further generated by choosing a thick bottom layer. Normally the tone of a stamp pattern is inversed when etching is used for pattern transfer. The tone can be preserved if another tone reversal process is used. This can be achieved by imprinting a pattern into the thick transfer layer, and by spin coating a silicon-containing resin on top of it. If the top residual layer of the planarized film is etched away, the high etch resistance of the silicon remaining in the trenches of the bottom layer will enable patterning of the transfer layer with reversed tone. This strategy has the advantage that stamp contamination containing silicon residues is avoided [5.49].

The 3-D patterning capability of NIL makes it possible to reduce the number of process steps in contact layer fabrication of microchips by using innovative pattern transfer. The connection of the transistors is done using several levels of lateral wires, each contacted vertically by through holes. This contact layer of a chip is fabricated using lithography and copper electroplating. For the wiring scheme of a chip, as shown in Fig. 5.11, eight levels of wiring layers are needed, each of which is done in a so-called dual hard damascene process. A two-tiered stamp with three height levels makes it possible to pattern the through holes as well as the wires in one step [5.50–52]. In this way, several exposure steps can be replaced by a single imprint with patterns of different residual polymer layer thickness. Similarly, a so-called SAIL (SAIL) process was developed that allows the patterning of transistors on flexible electronics, with several aligned metal layers. This is achieved by imprinting a multilevel pattern into a resist, and by subsequent etching of each resist level followed by a selective etching step where the metal layers are etched. Figure 5.11 shows the pattering scheme for one level of the contact layer of an IBM power computer microprocessor. Obducat has used a similar process for the generation of micrometer-sized contact holes in printed circuit boards and Hewlett

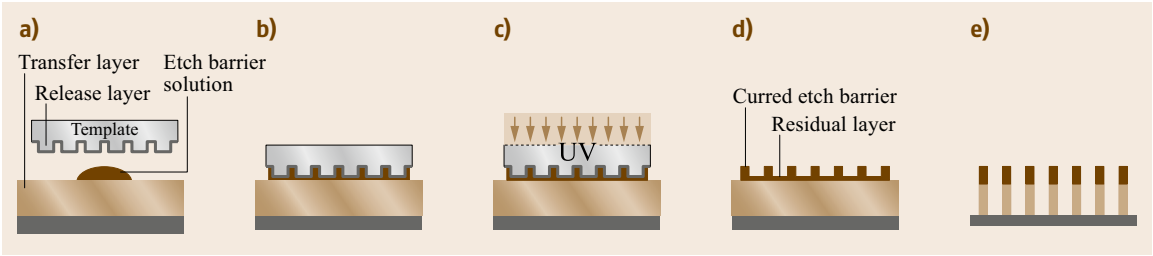


Fig. 5.10a–e Process scheme of UV imprint and pattern transfer, using a double layer (also called direct JFIL). The molded top layer, also called the etch barrier, is coated on a transfer layer, which serves as a planarization layer. It has also antireflective properties for the UV exposure through the stamp. **(a)** Dispensing of viscous resist droplet, **(b)** imprint, **(c)** UV exposure and curing, **(d)** demolding of hardened resist, **(e)** residual layer etch and transfer into bottom layer (breakthrough etch/window opening)

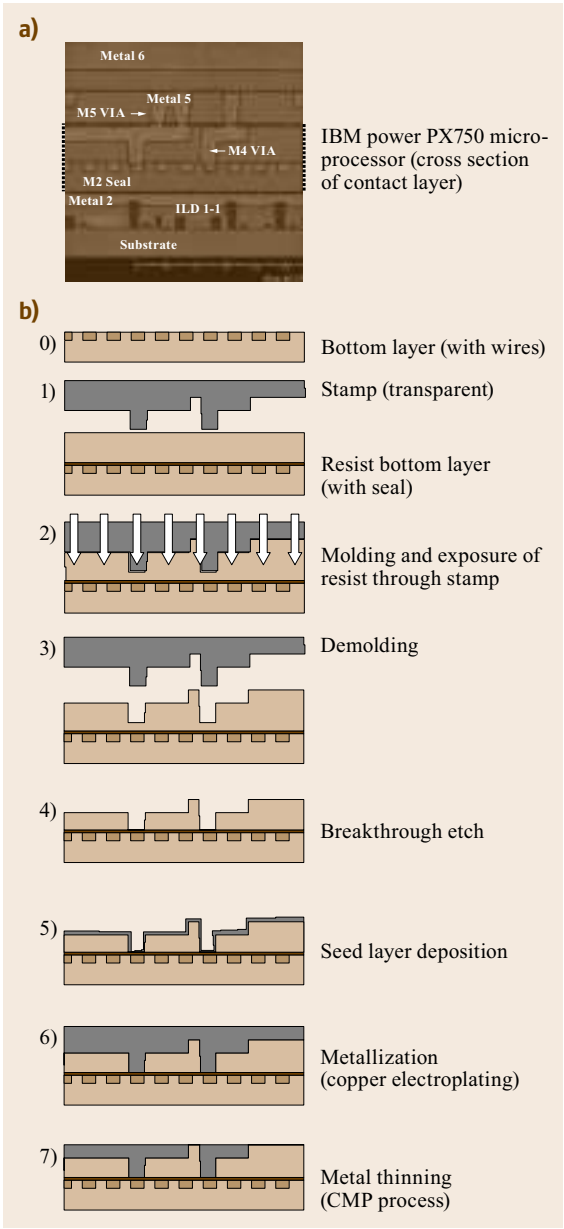


Fig. 5.11a,b Modified JFIL process proposed by Sematech to replace a dual top hard damascene process for copper contact plating by a two-tiered stamp (after [5.51]). **(a)** Scanning electron microscope (SEM) of a contact layer of a microchip (cross section) with interconnecting copper layers, **(b)** process scheme (after [5.50]) ◄

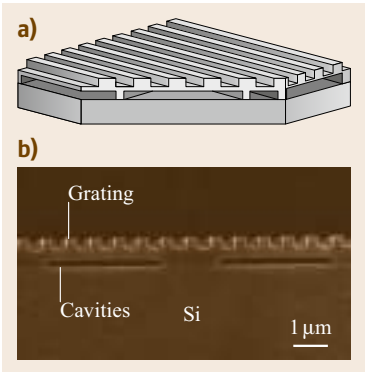


Fig. 5.12a,b Reverse microfluidic channels fabricated by double-sided imprinting: **(a)** 3-D schematic of a resist with a top grating and embedded channels. **(b)** SEM micrographs of cross-sections of imprinted nanofluidic channels: 3000 nm (width) × 200 nm (height) channels, with a 700 nm pitch grating on top (after [5.53])

Packard employed the SAIL process for flexible electronics.

5.2.8 Reversal NIL and Layer Transfer

In contrast to NIL, in reversal NIL the resist is patterned either directly on the stamp or on an auxiliary substrate, e.g., by spin coating, casting or imprint, and then transferred from the mold to a different substrate by bonding. Thus patterned resist structures are obtained as in direct NIL, and even embedded channels can be created. The concept is well presented in [5.53–55]. In reversal NIL it is possible to transfer patterns onto substrates that are not suitable for spin coating or have surface topographies. However, complete transfer does not only depend on a good balance of the surface energies, but also on the pattern density and roughness of the structures. As an example, embedded channels generated by reversal NIL are shown in Fig. 5.12 [5.53].

5.3 Tools and Materials for Nanoimprint

Mechanical nanofabrication techniques based on molding require tools and materials with matched mechanical properties. The mold has to be made from a material that is sufficiently hard to sustain at least one processing cycle. From the viewpoint of mass fabrication, a mold is considered as a tool that survives the molding process unaltered and uncontaminated, and thus can be reused many times after each molding step. In this way many identical replicas can be drawn (copied) from one mold. Due to the conformal molding, the surface of these

copies is the negative structure of the original (inverted polarity). Therefore a true replica of the mold is generated, when a negative is again molded into a positive structure. Here, we use the terms replica and copy in the more general sense that also negatives are considered as true copies of an original.

For thermal NIL as well as UV-NIL, single-step wafer-scale processing and step-and-repeat approaches have been developed. The aim is to make the reader familiar with concepts rather than presenting machines and materials sold on the market.

5.3.1 Resist Materials

Resists used for NIL are either used as an intermediate masking layer for the substrate or as a functional layer for a specific application. Both the processing properties and those for the final application purpose have to be considered. Many resists, as those used for PL and EBL, exhibit thermoplastic behavior. A typical example is PMMA, a regular linear homopolymer, with a short side-chain. It is used as a high-resolution standard material for EBL and also as bulk material for hot embossing and injection molding. When used as an etching mask, e.g., for Si, it exhibits a sufficient, but not high etching resistance. The glass transition temperature T_g of PMMA (100–120 °C) is low enough to enable molding at temperatures below 200 °C, but high enough to ensure a sufficient thermal stability in etching processes. Acrylate-based polymers can also be used with crosslinking agents. Apart from PMMA, a number of resists have been developed and characterized; they exhibit different T_g values, and were optimized for greater etching resistance, better flow at lower temperatures or equipped with intrinsic release properties. In Table 5.3 we give an overview of NIL resists with references to further information on these materials. Further information can be found in [5.9].

UV-curable NIL materials are composed of a mixture of monomers (or prepolymers) and a suitable photoinitiator, and often chemicals are added that decrease the effect of radical scavengers on photopolymerisation [5.56–59]. Immediately during contact of the stamp with the liquid mixture, filling of the mold starts by capillary forces, which pulls the stamp towards the substrate. Therefore, the general strategy is as follows: low viscosities are needed both for rapid dispensing and filling of mold cavities. Thin resin layers on top of a thicker transfer layer are used to achieve a homogeneous film thickness. Crosslinking and photopolymer conversion is adapted to achieve high curing speed and high etch resistance in the following breakthrough plasma etching process.

In UV-NIL a chemical reaction between stamp and the resist cannot be excluded. Small feature sizes along with high silicon content and a large degree of crosslinking make any residual imprint polymer left on the mold almost impossible to remove from the template without damaging the expensive quartz template. It was shown that a fluorosilane release layer applied to an UV-NIL stamp undergoes attack by acrylate, methacrylate and vinyl ether UV-curable resist systems, pointing to its degradation being intrinsic to the chemistries involved. Future resist chemistries have

Table 5.3 Thermoplastic or thermocurable materials used as resist layers for thermal and combined thermal and UV-NIL processes, with values for glass transition temperature T_g, recommended imprint temperature range $T_{imprint}$, viscosity η and Young's modulus E (after [5.9])

Material	T_g and $T_{imprint}$	$\eta(T)$ and resulting E	Comments
Thermoplastic (for variothermal molding)			
Poly(methyl metacrylate) (PMMA)	T_g 100–120 °C, $T_{imprint}$ 140–190 °C	10^5–10^7 Pa s (170 °C); 380–540 MPa	The *classic* NIL resist, also suitable for grayscale electron beam lithography
mr-I 7000 R[a]	T_g 55 °C $T_{imprint}$ 120–140 °C	$< 3 \times 10^4$ Pa s (120 °C)	Low T_g NIL polymer with very good flow ability and high plasma etch resistance, effective release force reduction due to fluorinated additives, 12 nm resolution proved
mr-I 8000 R[a]	T_g 155 °C $T_{imprint}$ 150–180° *C*	$< 10 \times 10^4$ Pa s (175 °C)	NIL polymer with very good flow ability, good thermal stability and high plasma etch resistance, effective release force reduction due to fluorinated additives
Thermoplastic and UV-curable (crosslinking during photoradiation for combined NIL and UV-photolithography)			
mr-NIL 6000 E[a]	T_g 1 °C (before imprint, no T_g after imprint) $T_{imprint}$ 65–70 °C	$< 2 \times 10^3$ Pa s (100 °C)	Low T_g NIL resist for combined thermal and UV-NIL (STU from Obducat), mix-and-match, multilevel patterning, reverse UV-NIL

[a] Commercially available from *micro resist technology* GmbH, Berlin, Germany

to satisfy the criterion of low reactivity toward antiadhesive coatings and stamp materials [5.60, 61].

5.3.2 Stamps

Stamps are characterized by a 3-D surface profile with protruding and recessed areas that are placed on a backbone conforming to the molding tool and the substrates to be used. The formed cavities, once the protruding areas are in contact with the moldable material, are filled during the process by displacement. Not only the mechanical, but also optical and chemical properties are important when choosing a stamp material for NIL. Critical mechanical parameters and their implications for NIL are hardness and thermal stability (lifetime and wear), thermal expansion coefficients and Poisson's ratio (dimension mismatch leading to distortions during demolding), roughness (higher demolding force and damage), Young's modulus (bending), and notch resistance (lifetime and handling). Issues related to fabrication are processability (etching processes, selectivity, cleanroom environment) and surface quality (resolution). The use in a NIL process is determined by additional properties such transparency, conductivity, wetting and antisticking properties (with/without antiadhesive coating, e.g., by covalent coating), availability and cost (standard materials and sizes, tolerances, processing equipment and time), and how easy it is to employ in NIL (e.g., fixing by clamping, thermobonding, gluing). In Table 5.4 we give a brief overview of mechanical and thermal properties of materials used for stamps. Further information can be found in [5.9, 62].

UV-NIL and Soft Lithography with Conformable Stamps in Low Pressure Imprint

The forces on the stamp protrusion with liquid resists are induced by capillary action rather than by squeeze flow and are therefore low. Therefore in UV-NIL, compliant stamps made from elastomeric materials e.g., PDMS (poly-dimethyl siloxane), a UV transparent rubber, can also be applied. The concept of layered stamps – a thin PDMS relief coated on a harder substrate – is particularly useful in full wafer concepts. It combines the complementary mechanical properties of a soft surface relief for the achievement of local conformal contact and a rigid, but bendable backbone, which can be used for mounting and alignment. A process

Table 5.4 Comparison of different materials for stamps (after [5.9])

Material for stamp originals and copies						
Material	**Young's modulus (GPa)**	**Poisson's ratio**	**Thermal expansion (10^{-6} K^{-1})**	**Knoop microhardness (kg mm^{-2})**	**Thermal conductivity (W m^{-1}K^{-1})**	**Specific heat (J kg^{-1} K^{-1})**
Hard stamp materials						
Silicon	131	0.28	2.6	1150	170	705
Fused silica (SiO_2, bulk)	73	0.17	0.6	500	1–6	700
Quartz (SiO_2, fused)	70–75	0.17	0.6	> 600 (8 GPa)	1.4	670
Silicon nitride (Si_3N_4)	170–290	0.27	3	1450	15	710
Diamond	1050	0.104	1.5	8000–8500	630	502
Nickel (Ni)	200	0.31	13.4	700–1000	90	444
Titane nitride (TiN)	600	0.25	9.4	2000	19	600
Material for stamp copying						
Material	**Characteristic**		**η and E**		**Comments**	
Thermocurable (thermally initiated crosslinking during heating)						
s-PDMS	Precursor + initiator, fully crosslinked after curing		4.575 Pa s; 1.8 MPa		Elastomeric, thermocurable silicone (Sylgard 184[b])	
h-PDMS	Four parts + solvent		Tunable; 8.2 MPa		Improved mechanical stability for sub-1 µm features	
UV-curable precursor (UV-initiated crosslinking after spin coating or casting)						
OrmoStamp[a]	UV-curable prepolymer, O_2 insensitive		0.4 ± 0.2 Pa s; 0.650 GPa		Organic-inorganic hybrid polymer for stamp copies used in T-NIL and UV-NIL, sub-10 nm resolution, high mechanical stability for aspect ratio ≥ 5	

[a] Commercially available from *micro resist technology* GmbH, Berlin, Germany
[b] From *Dow Corning* Inc., Midland, MI, USA

working with moderate resist viscosities (with $\eta_0 =$ 50 mPa s and below) for providing liquid films by spin coating was developed and can be applied at reduced environmental pressure in [5.39].

5.3.3 Stamp Fabrication and Tooling

Any kind of process generating a surface profile in a hard material can be used to fabricate stamps for NIL. The most common lithographic processes are based on resist patterning with subsequent pattern transfer. Therefore the requirements for these processes such as resolution, aspect ratio, depth homogeneity, sidewall roughness, and sidewall inclination are similar to the processes needed to do pattern transfer after NIL. For highest resolution down to sub-10 nm, both serial and parallel fabrication methods are available, however, with different area, throughput, and freedom of design. The processes are standard processes for nanolithography, such as electron beam lithography (EBL), even focused ion beam (FIB) lithography, but also interference lithography (IL), which also can be used directly for patterning. For large-area patterning with relaxed requirements on resolution, or for mix-and-match with high resolution techniques, photolithography (PL) has become a standard process. For 3-D topographies, apart from multiple patterning techniques, grayscale lithography and 2-photon-polymerization are more and more used. Using stamp copies instead of the original (as described for PDMS in the section before) is a way of enhancing the lifetime of a stamp, simply because the original is reserved for the copying process. There are different methods to generate copies from hard masters with proved resolutions below 100 nm: materials are now available that offer not only high replication fidelity down to 10 nm, but are also suitable for copying e.g., the fabrication of working stamps or of bendable molds that are used for roll-to-roll processes. A specific advantage of stamp manufacturing by copying is the possibility of use of resolution enhancement. Apart from combination of lithography with directed self-assembly, atomic layer deposition followed by spacer etching has been become popular because it enables the use of the original structures as guiding structures for further topography or chemical contrast selective processing. As an alternative to electroplating, imprinting into and casting of polymers, ormocers (hybrid organic and inorganic materials) and silicones (rubber-like *soft* stamps) have become popular ways of preparing stamp copies for prototyping and fabrication of a moderate number of replicas. However, the lifetime of polymeric molds is still low in comparison to nickel and silicon molds and good solutions for anti-adhesive coatings have to be found. Sylgard 184 from Dow Corning enables the fabrication of a number of transparent templates with sub-μm details. The liquid is cured by heat, and the PDMS solid is peeled away from the stamp original. Newer formulations with higher Young's moduli enable resolutions down to 50 nm. Hybrid molds use different materials for the surface relief and the support. They consist of a substrate plate as a mechanical support covered with a thin polymer layer with a nanostructured relief. In the case of NIL they have the advantage that a substrate material can be chosen with thermomechanical properties adapted to the substrate to be patterned. Furthermore it is useful if thin flexible substrates are needed.

The methods differ mostly in the properties of the materials used for the stamps (mechanical robustness, thermal expansion coefficient, transparency, fabrication tolerance) and the surface properties of the patterned relief (anti-adhesive coating possibility).

5.3.4 Anti-adhesive Coatings

One of the most important tasks for NIL is to provide stamps with good antisticking surface properties [5.63, 64]. The stamp surface should allow the molded surfaces to detach easily from the mold, and once released, provide a low friction resulting in a continuous vertical sliding movement without sticking. Nanoscopic interlocking of structures caused by sidewall roughness should be elastically absorbed by the molded material, while the surface maintains its antisticking properties. Because the molded polymer film is squeezed between the two surfaces of stamp and substrate, they need to exhibit opposite surface properties. The adhesion at both interfaces must be different to an extent that while the polymer film adheres perfectly at the substrate surface, the stamp can be separated from the structures without any damage at any location of the stamp. If the stamp material does not exhibit good anti-adhesive properties to the molded material, the stamp has to be coated with a thin antistiction layer (ASL). A low-surface-energy release layer on stamp surfaces not only helps to improve imprint qualities, but it also increases the stamp lifetime significantly by preventing surface contamination. An ASL has to be chemically inert and hydrophobic, but at the same time allow a filling of the mold cavities when the polymer is in its viscous state.

One of the major advantages of using Si or SiO_2 stamps for NIL is that they can be coated using silane chemistry. Fluorinated trichlorosilanes with different carbon chain lengths are commonly used due to their low surface energy, high surface reactivity, and high resistance against temperature and pressure. They support multiple long imprint sequences with repeated temper-

ature cycles higher than 200 °C. Currently it seems that as long as mechanical abrasion can be avoided, the silanes match the *normal use lifetime* of a Si stamp, which is some tens of cycles for NIL in a laboratory environment or thousands if automated step-and-repeat imprint processes or injection molding processes are used. For mass manufacturing it is advisable to use resists with intrinsic antisticking properties e.g., fluorinated components that are able to migrate towards the stamp's surface while being imprinted. The silane coating can be performed by immersion in a solution of iso-octane, or by chemical vapor deposition (CVD), either at ambient pressure by heating the silane on a hot plate or by applying a moderate vacuum of some mbar. One of the most prominent advantages of the vapor deposition method is that it is not affected by the wetting ability of a surface, and that it is suitable for stamp surfaces with extremely small nanostructures.

A commercially available silane used is shown in Fig. 5.13. F_{13}-OTCS = (tridecafluoro-1,1,2,2-tetrahydrooctyl)-trichlorosilane is the standard material for anti-adhesive coatings on silicon.

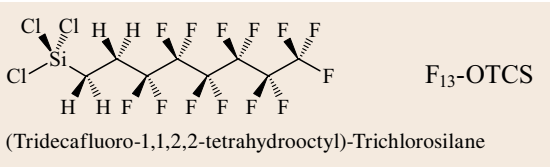


Fig. 5.13 Molecular structures of a fluorinated silane with a reactive trichlorosilane head group and a long alkyl chain with fluorine substituents (length about 2 nm). The silane binds covalently to the silicon oxide of the stamp surface and is used as the standard coating of silicon stamps in NIL

5.3.5 Machines

NIL can be carried out using three different types of machines:

1. Single step (Fig. 5.14a,b)
2. Step-and-repeat (Fig. 5.14c)
3. Roller imprinting.

An imprinting machine needs a precise pressing mechanism with high requirements on uniformity and homogeneity over large areas [5.8, 65–67]. At the same time it should adapt to local variations of pressure and temperature, due to imperfections and tolerances in stamps and substrates, and simply because the stamp protrusions are inhomogeneously distributed.

Imprint machines generate a desired pressure pattern over the total area of the stamp. High throughput for manufacturing devices at the full wafer scale can be achieved either by parallel patterning of large areas or by fast repeated patterning using a semiserial stepping process. The pressure field can also be applied sequentially by using a rigid but stepped imprint mechanism or continuously scanned pressure field, as used in roll embossing (or roller imprint) (Fig. 5.15). In all cases a defined area of the molding material is sandwiched between the solid stamp and substrate, which are backed up by a pressing mechanism.

Single-Step Wafer-Scale NIL (Large Area)

Single-step NIL machines pattern the surface on an entire wafer (current state 300 mm) or extended substrate (e.g., Gen2 panel $370 \times 470\,mm^2$) in one step. Thus the stamp must have the same size as the substrate to be patterned. The simplest mechanism for full-wafer imprint is a parallel plate imprint system. A linear movement

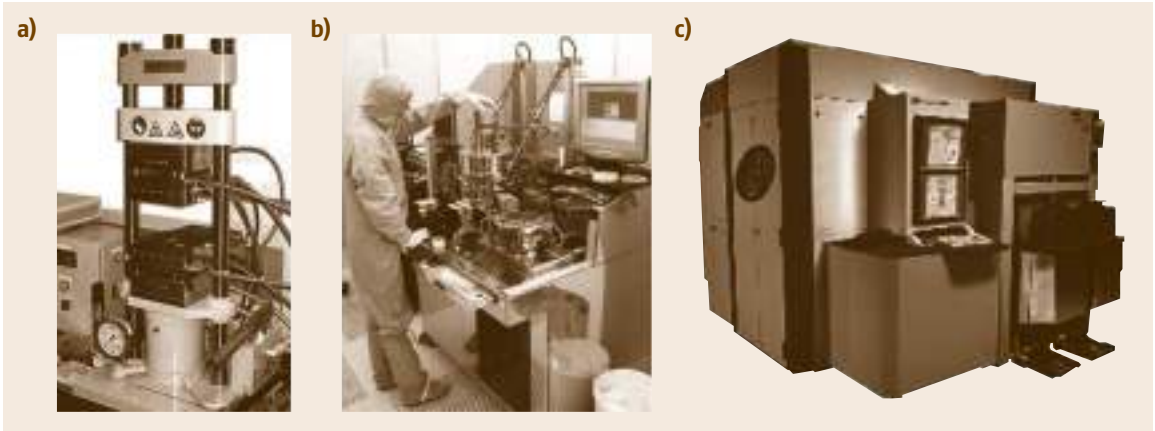


Fig. 5.14a–c Three examples of NIL presses. **(a)** Simple hydraulic press, with temperature-controlled pressing plates. **(b)** Semiautomated, hydraulic full-wafer NIL press, based on an anodic bonder. **(c)** Automated jet-and-flash UV-NIL production tool

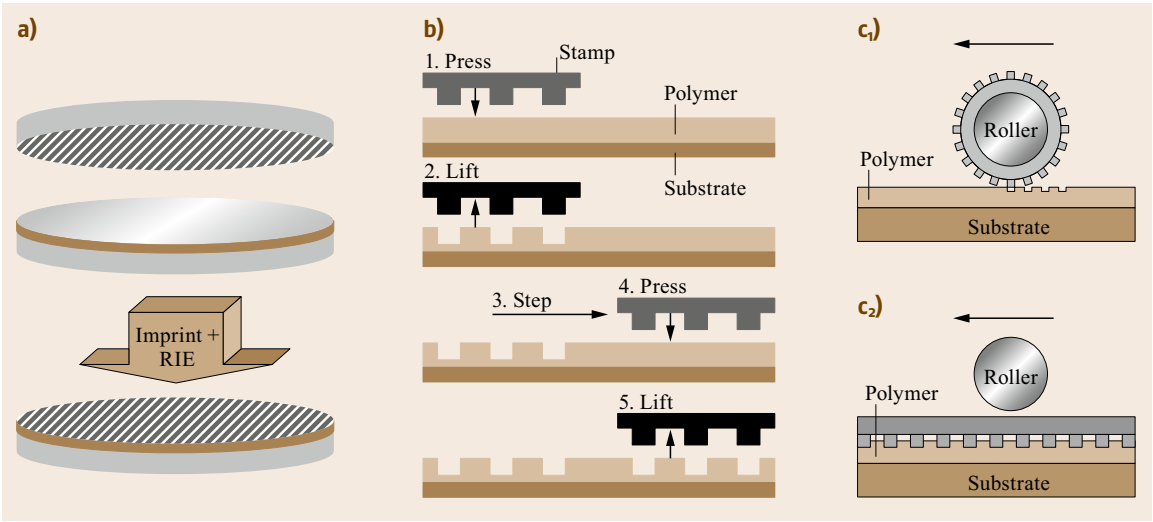


Fig. 5.15a–c Outline of the three most common types of NIL machines **(a)** full-wafer parallel press, **(b)** step-and-repeat press, and **(c)** two roll-embossing setups

of the piston behind the stamp leads to a local thinning of the polymer under the stamp protrusions, which is possible because the polymer is moved from squeezed areas into voids in the stamp. This movement can be generated using pneumatic, hydraulic, or motor-driven pistons. The pressure must be maintained during the whole molding process, until the voids are filled, and the molded structures are *fixed* during the cooling or curing step, depending on the method used. However, under normal process conditions, an imprint with a hard master does not work without a cushioning mechanism. This cushion balances both thickness variations due to tolerances of the setup and due to the nature of the molding process. The latter is caused by the fact that the size and shape of the stamp surface relief leads to a local pressure variation during the squeeze flow and, if the stamp can bend, to a local difference in sinking velocity. When using thick polymer plates, where the molding leads to a surface modulation of a bulky material, the cushion is formed by the viscous material itself. However, in NIL, a thickness profile has to be generated in a resist of which the thickness is often lower than the thickness tolerances of the substrates and mechanical setup used. Furthermore, height defects in the range up to some μm, such as dust particles, have to be equilibrated. Therefore the cushioning has to be achieved by the pressing mechanism and its ability to compensate has to be larger than the defects and tolerances of stamps and substrates. A lateral spreading and dispersion of the applied pressure can be achieved by using a spring mechanism, which can consist of an additional plastic or elastic layer. For example, a mattress made of rubber, Teflon or elastic graphite can be used. Due to the high pressure used in NIL, a compensation of small wedges, i. e., nonparallel alignment, is not needed. An even better pressure homogeneity can be obtained, when the cushion effect is generated by compressed air or liquid. This can be realized by constituting a pressure chamber, which is sealed against the backside of the stamp. In practice this is realized by placing a metallic or polymeric membrane between the pressure chamber and the stamp, which deforms around the stamp and substrate, and which is sealed with the counterforce of the stamper. Such tools are particularly well suited for soft stamps and UV-curable liquid resists, which can be imprinted at low pressure. The advantage of this soft stamping method is that a very gentle contact between stamp and substrate can be achieved by adjusting the air pressure, so that the surface can assume parallel alignment before the molding starts, and even local progression and retraction of the stamp can be programmed (e.g., in the so-called substrate conformal imprint lithography, SCIL [5.68]. During molding the pressure is equilibrated without delay, which assures a constant press force in all areas of the stamp only limited by the bending of the stamp.

All press concepts can be realized with heating elements for NIL, or with a UV exposure tool that enables exposure of the resist during molding. Furthermore, combinations of thermoplastic molding and UV exposure are possible. The main difference between thermoplastic molding and UV imprint are the pressures needed for imprint. Pressures from 1 to 100 bar are used in NIL, while less than 1 bar (0.1 MPa, 14.5 psi) is sufficient in UV-NIL.

Step-and-Repeat NIL

In PL, stepping was needed because of the limitation of the maximum field size to be exposed, and because the continuous reduction of structure sizes and diffraction effects is only possible by optical reduction of the masking structures into the resist by high-resolution optics. In contrast to this, step-and-repeat (S&R) NIL machines are physically identical to single-step NIL machines. They pattern a smaller area of a wafer at a time, and then move to an unpatterned area, where the process is repeated. The process is continued until the whole wafer is patterned. Many applications, e.g., with chip-based designs, enable the enlargement of the imprinted area by repeated printing with a smaller stamp, as long as the following imprints do not affect adjacent patterned areas. While this setup enables the use of smaller and more cost-effective molds, with which higher alignment accuracy can be achieved, higher process times and stitching errors at the borders of the patterned fields have to be taken into account. In the case of NIL heating and cooling times can be reduced because of the lower thermal mass, and in the case of UV-NIL illumination schemes can be simplified for smaller exposure fields; see Fig. 5.16.

In thermal NIL the thermal mass of the parts being thermally cycled should be minimized, in order to reduce the obtainable process time. This problem is readily addressed in S&R and in roll-to-roll NIL approaches, but has also found a solution in the concept of heatable stamps [5.69] or by surface heating by means of pulsed laser light [5.70]. The most advanced system is JFIL, which is directed towards high-volume manufacturing using 1 : 1 stepping of $26 \times 33\,\mathrm{mm}^2$ stamps. Apart from this, stamp fabrication and area enlargement, similar to recombination in counterfeit hologram production, is the most attractive field for S and R processes.

Roll-to-Roll NIL

Roll-to-roll (R2R) processing, also known as web processing, or reel-to-reel processing, involves the patterning of flexible materials such as plastics or metal foils (*web*), which are transferred from a roll (or reel) to another roll. If surface topographies are transferred to the foil, the term *roll embossing* is often used, which is here – in accordance with the scope of the chapter – named R2R-NIL [5.67, 71, 72]. To ensure continuous patterning of a resist on the foil, it is fed into the gap between two cylinders (*rolls*) from which one is patterned and the pattern is transferred. The area of highest pressure in which displacement of material is possible, is – due to the flexibility of either cylinders or web – in a *widened line*, or length of the web, which is called *nip*. This nip, from a few millimeters to centimeters, defines, combined with the speed of feeding, the process time available for pattering. For this, often bendable stamps are needed, which are wrapped around the cylinder.

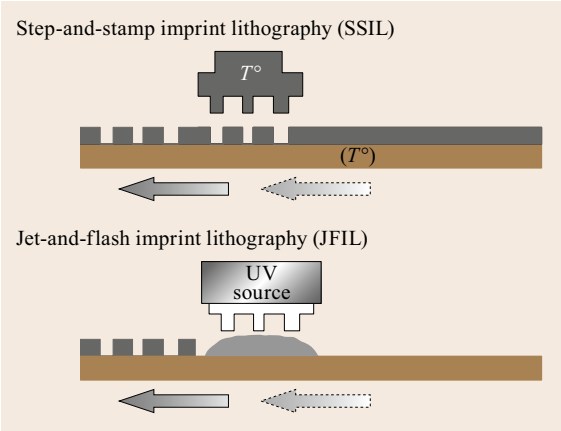


Fig. 5.16a,b Step-and-repeat processes. **(a)** In NIL: step-and-stamp imprinting lithography (SSIL), and **(b)** in UV-NIL: step and flash, aka jet and flash, imprint lithography (SFIL/JFIL). While in SFIL the liquid resin is cured locally by exposure through the stamp, in SSIL the resist is locally heated above its glass transition temperature by the hot stamp ($T°$ denotes a temperature often set above room temperature)

Continuous processing techniques are considered to have cost and processing advantages and are typically employed if speed and throughput have to be increased, but also if a constant quality over a large quantity of devices has to be achieved. However, they are physically identical to noncontinuous (sequential and stationary) processing techniques and require similar processing times for pattern formation and transfer. Therefore, apart from R2R, hybrid processes such as roll-to-plate (R2P) and plate-to-roll (P2R) sometimes have specific advantages over R2R. In R2P a pattern from a bent stamp can be imprinted or a resist transferred from a foil onto a flat substrate, and in P2R a flat stamp can be printed on a continuous foil in S&R semi-continuous mode.

In general, the advantage of roll processing is that once the process is stabilized, it can be run as long as contamination does not accumulate, antistiction layers (ASL) are preserved and processes are kept within a narrow process window. The disadvantage is that sometimes long optimization runs are needed before stability is achieved, causing waste of material, and the difficulty to interrupt processes for process control.

5.4 Applications

5.4.1 Types of Nanoimprint Applications

NIL applications can be divided into two main categories. In the first category, pattern transfer applications, the imprinted resist structure is used as a temporary masking layer for a subsequent pattern-transfer step. In the second category, polymer devices, the imprinted pattern adds functionality to the polymer film, which is the end product.

In many pattern-transfer applications, the main issue is high throughput at nanoscale resolution. Disregarding this issue, it is of minor importance whether the resist film is patterned by means of electromagnetic radiation, electrons, or by mechanical deformation. Only a few steps in the process flow are different, for example the dry etch step to remove the 10–100 nm-thick residual polymer layer after the imprint. Both additive and subtractive pattern transfer processes have been demonstrated, such as lift-off, electroplating and etching [5.73–76].

The advantages of NIL come into play if high resolution is needed over a large area. For such applications, NIL is a cost-effective alternative to current cutting-edge lithography techniques such as deep ultraviolet (DUV) lithography, dedicated to CMOS chip manufacturing [5.12, 77]. Examples of such products are bit-patterned media (BPM) for hard-disk drives (HDDs) [5.78, 79], patterned sapphire substrates (PSS) for enhanced light extraction from light-emitting diodes (LEDs [5.80] and subwavelength wire grid polarizers for high-definition television (HDTV) projectors [5.81].

The NIL process offers new possibilities to form polymer devices with microscale to nanoscale features. Nanoscale patterned polymer films find a wide range of applications within optics, electronics and nanobiotechnology. The capability to form 3-D polymer structures, with curved surfaces and high aspect ratio, paves the way for new classes of polymer-based passive optical devices, such as lenses and zone plates [5.46], photonic crystals (PhCs) [5.82] and integrated polymer optics [5.83]. The NIL technique allows for the choosing of a wide range of polymers with optimized optical properties, and allows for the patterning of thin films of organic light-emitting materials and polymers doped with laser dyes to create organic light-emitting devices (OLEDs) [5.84, 85] and lasers [5.86]. NIL is also suitable for nanoscale patterning of conducting organic films for cost-effective organic electronics [5.87].

Also within the rapidly growing field of lab-on-a-chip applications [5.88], NIL offers an attractive, cost-effective method for molding of complex structures, integrating micro- and nanofluidics, optics, mechanics and electronics on a single chip [5.89]. For example, the micro- to nanoscale fabrication capabilities are used to create single-use polymer devices containing nanopillar arrays [5.90] and nanofluidic channels [5.91] for DNA separation and sequencing.

In this section we will give an overview of different fields of applications. We start with two examples of pattern-transfer applications, which are close to production: bit-patterned media for HDD, and subwavelength metal wire gratings for HDTV projectors. We then discuss a few examples of laboratory-scale potential high-impact applications of NIL. These examples were selected from a large number of NIL applications. The number of laboratory-scale NIL applications is rapidly growing, reflecting a wealth of new possible device architectures becoming feasible by NIL. Some of the applications are directly relevant for industrial production, and others are directed towards research. Even in research the nanostructuring capability of replication processes are needed. Further insight into this field is given in Sect. 5.1.2 about pattern transfer and in Sect. 5.5 about commercialization aspects of NIL.

5.4.2 Semiconductor Memory Chips

The semiconductor industry is currently (2016) considering NIL as possible NGL to deliver the 14 nm node and beyond for the manufacturing of advanced devices such as NAND (negative AND logic) flash memory [5.13, 92]. The requirements for single-substrate patterning with moderate defect density and for chip manufacturing with 50 aligned masking levels are quite different from single-layer devices. One-step patterning at 15 nm half-pitch was achieved with JFIL, eliminating the need for complicated and expensive self-aligned multiple patterning. In addition, patterns are not limited to repeating structures such as lines and spaces. For chip manufacturing not only are the resolution issue and cost of masks important, but more and more the defectivity rate, which was until now considered as prohibitive because NIL – as a mechanical contact-based technology – is prone to contamination and wear. As an example, for a 16 nm NAND flash device, a 3σ overlay of approximately 5 nm, and less than 5 defect/cm^2 are required at a throughput of 20 wph (wafers per hour). While NIL is here mostly considered as the replacement of individual lithographic steps (pattern transfer applications), its 3-D capability, e.g., for the dual-damascene patterning of the interconnect (metalization) layer on top of the CMOS transistors, is still an option for future devices (Sect. 5.2.7). Here a functional material

is patterned, which after exposure remains as dielectric between the electroplated metal interconnects.

5.4.3 Bit-Patterned (Magnetic) Media (BPM) for Hard Disk Drives

Since the first demonstration of NIL, patterned magnetic media for HDD has been a key application for NIL technology [5.93]. After invention of the HDD in 1957, the storage capacity, quantified in areal density of bits, has increased to the current (2017) level exceeding 1 Tb/in^2. The size and density of the individual bits, defined by local magnetization of a homogeneous, or unpatterned thin magnetic film, was reduced by application of multilayer magnetic films as recording media, the sensitivity of the read-head was increased by exploiting the giant-magnetoresistance effect in multilayer thin-film conductors [5.94], and the magnetization was applied perpendicular to the surface of the recording media, while MEMS technology for the mechanical parts has been developed to a rather extreme level. In current HDDs the read-write head flies at a height of 2–3 nm above the surface of the disk plate. An overview of HDD technology is given in [5.95].

The possibilities to increase the bit density with the current technology, where bits are written by local magnetization of an unpatterned thin magnetic film, are mainly limited by the read-write width, the positioning of the magnetic head, and by thermal instability, induced by superparamagnetism in the grains of the magnetic film. These challenges are addressed by patterning the magnetic film, first implemented by discrete track recording (DTR) media [5.96], where the magnetic film is patterned with a spiral land, and groove tracks have been developed to overcome the problems associated with the read-write width and positioning of the magnetic head.

The magnetic film consists of small, weakly coupled magnetic grains, which behave as single-domain magnetic particles. Each bit consists of the order of 100 grains to obtain a reasonable signal-to-noise ratio. In order to keep this ratio of grains per bit, the grain size must be reduced with bit size. Now, with BPM, it is even increased to cover a single bit. The magnetic energy of a single grain scales with the volume of the grain. This implies that the bit can be erased thermally, when the grain size becomes sufficiently small and weakly coupled to neighboring grains. This is referred to as the superparamagnetic limit.

The superparamagnetic limit can be overcome by lithographically defining each bit as a magnetic nanoparticle or *nanomagnet*. In such a *quantized magnetic disk* [5.97, 98] each magnetic nanoparticle is a single magnetic domain with a well-defined shape, with a uniaxial magnetic anisotropy, so the magnetization only has two possible stable states, equal in magnitude but opposite in direction, as illustrated in Fig. 5.17. Such defined bits can be thermally stable for sizes down below 10 nm.

The feasibility of NIL for fabrication of patterns of magnetic nanostructures for quantized magnetic disks has been investigated by several research groups as reviewed in [5.78, 79]; see Fig. 5.18. While extensive research has been performed by enterprises such as Western Digital, Seagate and Toshiba during the last years, introduction of BPM has been postponed.

5.4.4 Subwavelength Metal-Strip Gratings and Transparent Electrodes for Displays

Metallic wire gratings with a period below 200 nm can be used to create polarizers, polarization beam splitters and optical isolators in the visible range. Such devices have many applications in compact and integrated optics, and not the least for liquid-crystal flat panel displays.

The polarizing functionality of subwavelength wire gratings is based on form-birefringence, an optical anisotropy that appears when isotropic material is structured on a length scale much smaller than the wavelength of light, λ. In this limit, the description of light propagation based on the laws of diffraction, refraction and reflection is not valid, and a rigorous solution of Maxwell's equations with the relevant boundary conditions must be applied. For a review on subwavelength optics, see [5.99]. The subwavelength linear grating of period $d < \lambda/2$, line width a, and height h, as illustrated in Fig. 5.19, will behave as a film of bire-

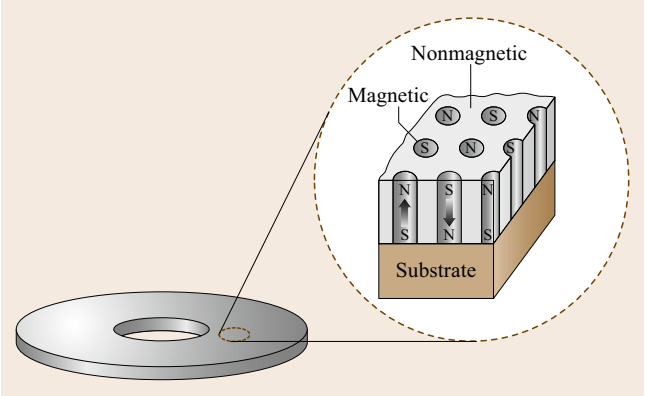


Fig. 5.17 Outline of a patterned magnetic disk for high-density data storage. Each bit is a lithographically defined, single-domain magnetic nanostructure, embedded in a nonmagnetic matrix (after [5.79])

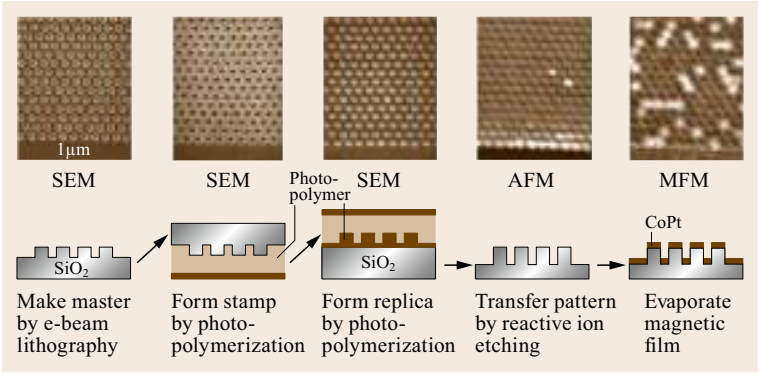


Fig. 5.18 Outline of the process flow for fabrication of 55 nm-diameter magnetic islands by UV-NIL. The *top panel* shows SEM, atomic-force microscope (AFM) and magnetic-force microscope (MFM) micrographs at the different stages of the process. The MFM micrograph shows quantized up and down magnetization of isolated domains (after [5.79])

fringent material with the refractive indices n_s and n_p for the s-polarized (E-field parallel to the grating) and p-polarized (E-field perpendicular to the grating) light

$$n_p^2 = \frac{d}{a} n_1^2 + \left(1 - \frac{d}{a}\right) n_2^2 ,$$

$$n_s^2 = \frac{n_1^2 n_2^2}{\frac{d}{a} n_2^2 + \left(1 - \frac{d}{a}\right) n_1^2} ,$$

where n_1 and n_2 are the refractive indices of the isotropic grating and fill materials respectively.

For applications in liquid-crystal display (LCD) and liquid crystal on silicon (LCoS) projection devices, it is a key challenge to obtain a sufficiently high extinction ratio, larger than 2000, at the shorter wavelengths, i. e., for blue light ($\lambda \approx 450\,\text{nm}$). This requires a pitch d of 100 nm or smaller, which is not practical for production with conventional photolithography. *Yu* et al. [5.100], demonstrated a large-area ($100 \times 100\,\text{mm}^2$), $d = 100\,\text{nm}$ by NIL. The stamp gratings were formed by interference lithography using an Ar-ion laser ($\lambda = 351.1\,\text{nm}$) to achieve a pitch around 200 nm, which was transferred into a SiO_2 film by reactive-ion etching (RIE). The pitch was subsequently halved by *spatial frequency doubling*: conformal CVD deposition of Si_3N_4, and anisotropic trifluormethan/oxygen (CHF_3/O_2) RIE. Researchers at LG electronics have realized $d = 100\,\text{nm}$ Al wire grating polarizers by thermal NIL. The process is outlined in Fig. 5.20. The large-area grating stamp is fabricated by laser interference lithography in photoresist, and transferred into the underlying 200 nm-thick SiO_2 film using CF_4 and O_2 RIE [5.81]. The $50 \times 50\,\text{mm}^2$ devices (Fig. 5.21) have an extinction ratio over 2000 and a transmittance above 85% in the blue, at $\lambda = 450\,\text{nm}$. Large-area wire grid polarizers with extinction ratios larger than 1000 were fabricated by a roll-based JFIL process [5.101].

Metal films patterned as wire grids by nanoimprint lithography also have potential for transparent electrodes [5.102], for plasmonic devices [5.103] and the generation of plasmonic colors and color filters [5.104].

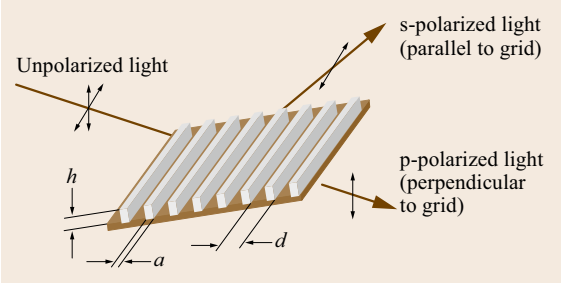


Fig. 5.19 Subwavelength wire grid polarizer. By application of subwavelength gratings, with a pitch below 100 nm for visible light, first-order diffraction with a high acceptance angle and low dispersion birefringence is obtained (after [5.81])

5.4.5 High Brightness Light-Emitting Diodes

GaN-based light-emitting diodes (LEDs) have large potential as energy efficient, long lifetime, environmentally friendly and stable light sources, and are currently entering a range of applications, such as full-color displays and projectors, traffic lights, and automotive and architectural lighting. Due to a high refractive index of the semiconductor material, the emitted light is easily trapped in waveguide modes inside the device, which strongly reduces external efficiency of the light source. The light extraction from the device can be significantly enhanced by nanopatterning the surface [5.105]. *Kim* et al. [5.106] applied a 2-D photonic crystal – an array of holes – with a photonic bandgap that prohibits propagation of photons of frequencies within the bandgap, leading to an enhanced extraction of photons through the surface of the device, and demonstrated a ninefold enhancement of the photoluminescence intensity of GaN-based green LEDs by means of 2-D PhC struc-

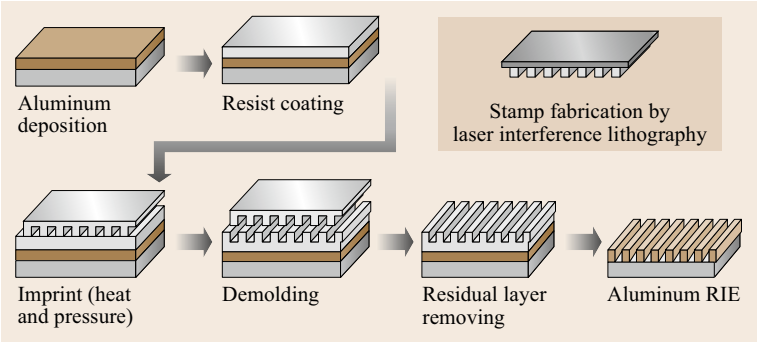


Fig. 5.20 Outline of the NIL process to fabricate $d = 100\,\text{nm}$ pitch aluminum wire grating polarizers

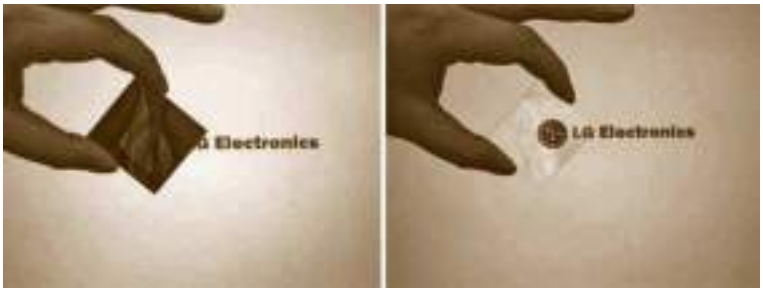


Fig. 5.21 Large-area 100 nm pitch wire grid polarizer with 85% transmission and extinction ratio larger than 2000 at wavelength $\lambda = 450\,\text{nm}$ (blue light)

tures of 180 nm-diameter, 100 nm-deep holes arranged in a square lattice with a period of 295 nm. The PhC pattern was defined by thermal NIL and RIE through a Cr mask. The NIL stamp was patterned by laser interference lithography.

Apart from such PhC surface patterns, which have also have the potential to enhance the efficiency of solar cells, NIL is used to create prepatterned substrates for semiconductor film growth, e.g., patterned sapphire (PSS) [5.107], which has been widely adopted in the industry because the epitaxial quality and light extraction efficiency can be improved simultaneously. This is done by patterning a resist and etching a regular microtopography (e.g., an orthogonal pattern with about 3 μm period) into a sapphire substrate. This pattern can be transferred into sapphire by (wet or dry) etching. After etching, the rectangular shape of the resist results in square pyramids, cones or domes, with sizes of about 3 μm and 1.0–2.0 μm height, due to the crystalline orientation of sapphire. The same principle can be employed by patterning to top of vertical-cavity surface-emitting lasers (VCSEL) [5.68] and entire photovoltaic cells [5.108].

5.4.6 Polymer Optics

NIL is ideally suited for the fabrication of polymer nanophotonics and waveguide devices with submicron critical dimensions, defined over large areas. It is also compatible with many polymer materials, giving large freedom to choose a material with specific optical properties [5.109–111].

In Fig. 5.22 we show a polymer microring resonator fabricated by NIL [5.83]. This type of device has been realized in PMMA, PC (polycarbonate), and PS (polystyrene) on SiO_2 substrates. The resonator consists of a planar waveguide and an adjacent microring waveguide. The waveguide and microring are coupled through the evanescent field in the coupling region. Resonant dips in the transmission through the waveguide occur when the phase pick-up in a trip round the microring is equal to $2\pi m$, where m is an integer. The device works as a narrow bandwidth filter, and finds applications within integrated optics and for biosensing [5.112].

The evanescent coupling coefficient between the waveguide and microring depend exponentially on the size of the gap. The devices are realized with 1.5 μm-high waveguides, and a coupling air-gap of 100–200 nm. The process flow is outlined in Fig. 5.22c. A thin initial polymer layer is spin cast onto a SiO_2 substrate layer. The stamp has a very large fill factor and large protrusion areas, implying that a large polymer flow is needed to fill the stamp cavities. A thin residual polymer layer is obtained by combining a high imprint pressure, a high process temperature and a long imprint time. The mode confinement in the PS waveguides is enhanced by etching the substrate oxide layer isotropically in hydrofluoric acid (HF), to create a pedestal structure. The Q-factor of

the resonator device depends critically on the surface scattering losses in the waveguides. The surface roughness of the polymer waveguides can be reduced by a controlled thermal reflow. The device is heated to 10–20 °C below the glass transition, and the surface reflows under the action of surface tension. A loss reduction of more than 70 dB/cm was achieved by this approach [5.113].

Apart from integrated micro- and nano-optics, antireflection gratings and subwavelength polarizers, NIL is particularly interesting for the fabrication of polymer devices such as light guides used for backlight illumination of displays and organic solar cells. For those applications, NIL is not only an easy way to patterns surfaces, but because it provides the ability to use R2R processes for upscaling and low-cost fabrication.

5.4.7 Biological Applications

Micro- and nanofabrication technology has enabled methods to manipulate and probe individual molecules and cells on a chip [5.114–117]. This type of application often requires a large area covered with nanostructures. Sometimes a large number of identical devices are needed for statistical evaluation, or to give redundancy, e.g., against clogging of nanofluidic channels. With these requirements, NIL is advantageous, or sometimes the only viable lithography method, even for laboratory-scale experiments and prototyping. Another example is devices for investigation of cell response to nanostructured surface topography, which requires nanometer-scale patterned surface areas in the mm^2 to cm^2 range.

Nanofluidic channels can be used to stretch DNA [5.118–120] for high-throughput, linear analysis. It enables measuring the length L of individual DNA molecules, or possibly sequencing by detection of fluorescent labels attached to the specific DNA sequences [5.121]; see Fig. 5.23. The linear analysis relies on a uniform stretching of the DNA molecules without coiling, as they are driven through the narrow channel. This implies that the nanofluidic channel should have cross-sectional dimensions D close to or smaller than the persistence length of the DNA, $L_p \approx 50$ nm [5.121]. The assumption of uniform stretching of the molecule also puts strong requirements to the channel sidewall smoothness.

Cell growth and adhesion can be strongly influenced by surface topography on the micrometer to nanometer length scale [5.122], and antibacterial nanotextured surfaces can be applied to minimize biofilm formation [5.123].

Gadegaard et al. [5.124] created a three-dimensional tubular scaffold for tissue engineering of blood vessels, which reproduces the basic structure of natural blood vessels: a layer of smooth muscle cells – fibroblasts – coaxially embedded between an outer collagen mesh and an inner linen of endothelial cells. Such artificially grown blood vessels with tight control of cellular attachment, migration and growth are expected to reduce problems with cellular debris and inflammation.

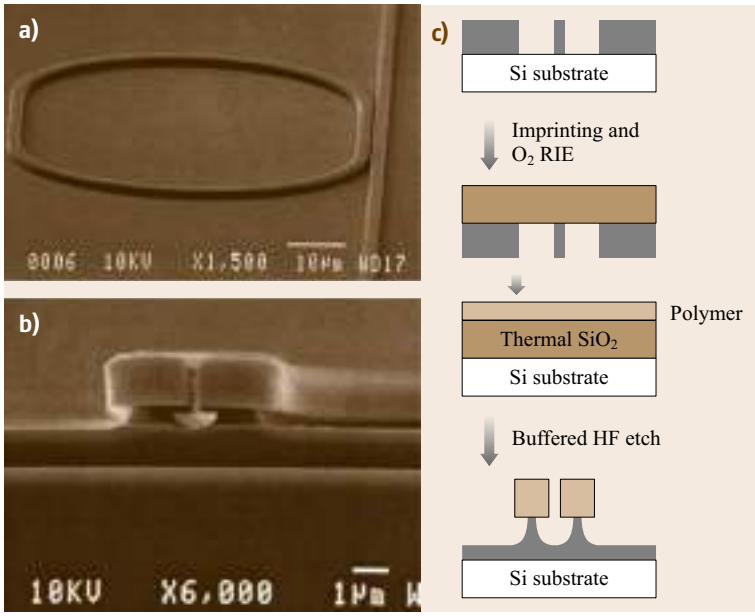


Fig. 5.22a–c Nanoimprinted polymer microring resonator. **(a)** SEM picture of the imprinted device. **(b)** Cross-sectional SEM picture of the polymer waveguides in the coupling region of the microring device. **(c)** Outline of the process flow (after [5.83])

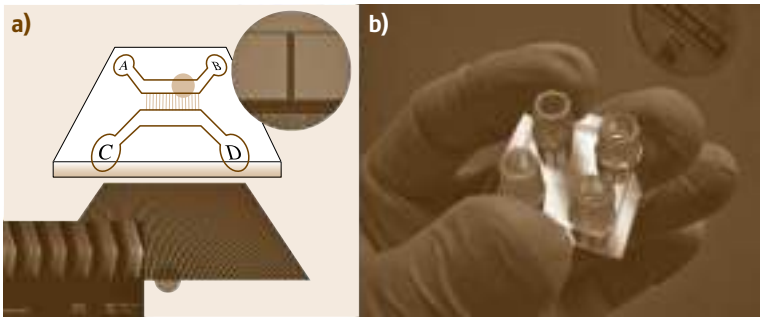


Fig. 5.23a,b Nanofluidic device for high-throughput linear DNA analysis. **(a)** Microfluidic channels A–B and C–D are connected via an array of 100 nm-wide nanofluidic channels. The $5 \times 1\,mm^2$ nanofluidic channel array is defined by NIL. **(b)** Shows the finished device package (after [5.120])

5.5 Conclusion and Outlook

Technological development is much based on so-called enabling techniques. For example, Gutenberg's book printing with movable metal letters was based on a combination of different existing techniques (large wine presses and metallurgy for letter casting), solving throughput and flexibility problems, and was developed at a time of globalization when information needed to be multiplied and spread (e.g., books, leaflets and maps) (1455 AD) [5.125]. In a similar way, a new lithographic technique with micro- and nanopatterning capability, such as NIL, is not entirely new, but based on patterning techniques coming from silicon micromachining and compact disc molding. In a time of technological dynamics it will lead to advances in different fields:

- In research, as long as machines are affordable and reliable enough that they can replace or complement standard lithographic techniques. In the case of thermal NIL it is advantageous that nanostructures can be replicated with simple molding tools, e.g., hot presses without alignment, thus making it possible to integrate nanopatterning by NIL into device manufacturing. Standard mask aligners can be upgraded to perform UV-NIL with moderate pressures.
- In industry, if they help to cross technological barriers, they can reduce cost and enable expansion into fields that until now have been reserved to high-volume manufacturing (HVM). Success will also depend on whether they fit into the process chain already established in a semiconductor cleanroom environment. Furthermore, substrate sizes, throughput and yield have to correspond to production requirements. As in research, many of the machines already available can be used for moderate-scale production. They can be scaled up to substrate sizes of 300 mm and higher in combination with batch-mode operation (wafer track systems) and by combining imprint tools into clusters.

In semiconductor manufacturing, NIL is currently seen as an alternative to established high-end photolithographies, enabling the replace of single or multiple mask patterning steps by a 1X imprint step – this is mostly owing to the efforts in reducing defectivity and increasing throughput. Introduction has been postponed several times, but should be – at least for NAND memory chips – in 2017 [5.13, 92]. However, NIL has other capabilities, as demonstrated in Sect. 5.4, even if not all requirements of the ITRS roadmap are met at once. Most of the other applications are based on single layers covered with nanostructures, mostly regular high-resolution gratings and dot arrays, and need single-step wafer-scale replication tools for large areas. However, CoO has become the driving force for integration of new techniques, since both 193 nm DUV- and 13.6 nm EUV-lithography have been pushing the limits of photolithography towards single-digit resolution.

NIL has now passed a barrier from a laboratory scale to industrial preproduction and is now used in patterning of PSS and for optical microdevices. Although it seems that room-temperature processes based on UV exposure have an advantage over processes based on thermocycles, even after 20 yr of development it is difficult to say which process will become a standard process and make it to the production line [5.126]. For example, isothermal processes at elevated temperatures using hybrid processes have been established that use both thermal NIL and hardening by UV curing [5.127]. With state-of-the-art UV-NIL equipment [5.125], more than 20 wph with a diameter of 300 mm can now be achieved in a step-and-repeat mode (using a stamp area of $26 \times 33\,mm^2$). Single-step wafer-scale hot imprint

has similar capabilities, and can even push throughput further if heatable stamps with low thermal mass are used. However, NIL is currently such a fast-moving field that preclusion about the final success of one technique is not possible and advisable. Innovative solutions are still needed to solve process and stamp lifetime issues for many different applications. Probably not just a single NIL process will be successfully implemented, but many variants of NIL. This includes hybrid approaches, e.g., NIL in combination with other lithographic processes, or the fabrication and copying of stamps using NIL.

References

5.1 E. Berliner: Gramophone, US Patent 372786 (1887)
5.2 K.C. Pohlmann: *The Compact Disc Handbook, Computer Music and Digital Audio Series*, Vol. 5, 2nd edn. (A-R Editions, Madison 1992)
5.3 H. Schift, C. David, M. Gabriel, J. Gobrecht, L.J. Heyderman, W. Kaiser, S. Köppel, L. Scandella: Nanoreplication in polymers using hot embossing and injection molding, Microelectron. Eng. **53**, 171–174 (2000)
5.4 S.Y. Chou, P.R. Krauss: Imprint lithography with sub-10 nm feature size and high throughput, Microelectron. Eng. **35**, 237–240 (1997)
5.5 R.W. Jaszewski, H. Schift, J. Gobrecht, P. Smith: Hot embossing in polymers as a direct way to pattern resist, Microelectron. Eng. **41/42**, 575–578 (1998)
5.6 Y. Xia, G.M. Whitesides: Soft lithography, Angew. Chem. Int. **37**, 550–575 (1998)
5.7 B. Michel, A. Bernard, A. Bietsch, E. Delamarche, M. Geissler, D. Juncker, H. Kind, J.-P. Renault, H. Rothuizen, H. Schmid, P. Schmidt-Winkel, R. Stutz, H. Wolf: Printing meets lithography: Soft approaches to high-resolution, IBM J. Res. Dev. **45**(5), 697–719 (2001)
5.8 H. Schift: Nanoimprint lithography: An old story in modern times? A review, J. Vac. Sci. Technol. B **26**(2), 458–480 (2008)
5.9 H. Schift (Ed.): *NaPa Library of Processes*, http://www.psi.ch/lmn/helmut-schift (2014), 3rd edn.
5.10 H. Schift: Nanoimprint lithography: 2-D or not 2-D? A review, Appl. Phys. A **121**(2), 415–435 (2015)
5.11 H. Schift, P. Urwyler, P.M. Kristiansen, J. Gobrecht: Nanoimprint lithography process chains for the fabrication of micro- and nanodevices, J. Micro/Nanolith. MEMS MOEMS **13**(3), 031303 (2014)
5.12 B.W. Smith, K. Suzuki (Eds.): *Science and Technology, Microlithography*, 2nd edn. (CRC, Boca Raton 2007)
5.13 ITRS: International Technology Roadmap for Semiconductors, http://itrs2.net
5.14 H. Moore: Cramming more components onto integrated circuits with unit cost falling as the number of components per circuit rises, Electronics **38**(8), 114–117 (1965)
5.15 S.Y. Chou, P.R. Krauss, P.J. Renstrom: Imprint of sub-25 nm vias and trenches in polymers, Appl. Phys. Lett. **67**(21), 3114–3116 (1995)
5.16 S. Fujimori: Fine pattern fabrication by the molded mask method (nanoimprint lithography) in the 1970s, Jpn. J. Appl. Phys. **48**, 06FH01 (2009)
5.17 S.Y. Chou, P.R. Krauss, P.J. Renstrom: Nanoimprint lithography, J. Vac. Sci. Technol. B **14**(6), 4129–4133 (1996)
5.18 L. Baraldi, R. Kunz, J. Meissner: High-precision molding of integrated optical structures, Proc. SPIE **1992**, 21–29 (1993)
5.19 L.J. Heyderman, H. Schift, C. David, J. Gobrecht, T. Schweizer: Flow behaviour of thin polymer films used for hot embossing lithography, Microelectron. Eng. **54**, 229–245 (2000)
5.20 H. Schulz, M. Wissen, N. Bogdanski, H.-C. Scheer, K. Mattes, C. Friedrich: Impact of molecular weight of polymers and shear rate effects for nanoimprint lithography, Microelectron. Eng. **83**, 259–280 (2006)
5.21 J. Haisma, M. Verheijen, K. van den Heuvel, J. van den Berg: Mold-assisted lithography: A process for reliable pattern replication, J. Vac. Sci. Technol. B **14**, 4124–4128 (1996)
5.22 M. Colburn, S. Johnson, M. Stewart, S. Damle, T. Bailey, B. Choi, M. Wedlake, T. Michealson, S.V. Sreenivasan, J. Ekerdt, C.G. Willson: Step and flash imprint lithography: A new approach to high-resolution patterning, Proc. SPIE **3676**, 379–389 (1999)
5.23 D.J. Resnick, W.J. Dauksher, D. Mancini, K.J. Nordquist, T.C. Bailey, S. Johnson, N. Stacey, J.G. Ekerdt, C.G. Willson, S.V. Sreenivasan, N. Schumaker: Imprint lithography: Lab curiosity or the real NGL?, Proc. SPIE **5037**, 12–23 (2003)
5.24 D.J. Resnick, S.V. Sreenivasan, C.G. Willson: Step and flash imprint lithography, Mater. Today **8**, 34–42 (2005)
5.25 Z. Ye, K. Luo, X. Lu, B. Fletcher, W. Liu, F. Xu, D. LaBrake, D.J. Resnick, S.V. Sreenivasan: Defect reduction for semiconductor memory applications using jet and flash imprint lithography, J. Micro/Nanolith. MEMS MOEMS **11**(3), 031404 (2012)
5.26 M. Doi: *Introduction to Polymer Physics* (Claredon, Oxford 1996)
5.27 D.W. van Krevelen: *Properties of Polymers* (Elsevier, Amsterdam 1990)
5.28 H. Schift, L.J. Heyderman: Nanorheology-squeezed flow in hot embossing of thin films. In: *Nanostructure Science and Technology, Alternative Lithography*, ed. by D.J. Lockwood, C. Sotomayor Torres (Kluwer Academic, New York 2003) pp. 46–76
5.29 H.-C. Scheer, H. Schulz, T. Hoffmann, C.M. Sotomayor Torres: Nanoimprint techniques. In: *Handbook of Thin Film Materials*, Vol. 5, ed. by H.S. Nalva (Academic, New York 2002) pp. 1–60

5.30 E.A. Dobisz, S.L. Brandow, R. Bass, J. Mitterender: Effects of molecular properties on nanolithography in polymethyl methacrylate, J. Vac. Sci. Technol. B **18**, 107–111 (2000)

5.31 R.B. Bird, C.F. Curtis, R.C. Armstrong, O. Hassager: *Dynamics of Polymeric Liquids, Fluid Mechanics*, Vol. 1 (Wiley, New York 1987)

5.32 H.-C. Scheer, H. Schulz: A contribution to the flow behaviour of thin polymer films during hot embossing lithography, Microelectron. Eng. **56**, 311–332 (2001)

5.33 C. Gourgon, C. Perret, G. Micouin, F. Lazzarino, J.H. Tortai, O. Joubert, J.-P.E. Grolier: Influence of pattern density in nanoimprint lithography, J. Vac. Sci. Technol. B **21**(1), 98–105 (2003)

5.34 H.D. Rowland, W.P. King: Polymer deformation and filling modes during microembossing, J. Micromech. Microeng. **14**, 1625–1632 (2004)

5.35 H. Schift, L.J. Heyderman: M. Auf der Maur, J. Gobrecht: Pattern formation in hot embossing of thin polymer films, Nanotechnology **12**, 173–177 (2001)

5.36 T.J. Ui, R.G. Hussey, R.P. Roger: Stokes drag on a cylinder in axial motion, Phys. Fluids **27**, 787 (1984)

5.37 Y. Hirai, S. Yoshida, N. Takagi: Defect analysis in thermal nanoimprint lithography, J. Vac. Sci. Technol. B **21**(6), 2765–2770 (2003)

5.38 M. Colburn, B.J. Choi, S.V. Sreenivasan, R.T. Bonnecaze, C.G. Willson: Ramifications of lubrication theory on imprint lithography, Microelectron. Eng. **75**, 321–329 (2004)

5.39 A. Fuchs, M. Bender, U. Plachetka, U. Hermanns, H. Kurz: Ultraviolet-based nanoimprint at reduced environmental pressure, J. Vac. Sci. Technol. B **23**(6), 2925–2928 (2005)

5.40 M. Colburn, I. Suez, B.J. Choi, M. Meissl, T. Bailey, S.V. Sreenivasan, J.G. Ekerdt, C.G. Willson: Characterization and modelling of volumetric and mechanical properties for step and flash imprint lithography photopolymers, J. Vac. Sci. Technol. B **19**(6), 2685–2689 (2001)

5.41 D.J. Resnick, W.J. Dauksher, D. Mancini, K.J. Nordquist, T.C. Bailey, S. Johnson, N. Stacey, J.G. Ekerdt, C.G. Willson, S.V. Sreenivasan, N. Schumaker: Imprint lithography for integrated circuit fabrication, J. Vac. Sci. Technol. B **21**(6), 2624–2631 (2003)

5.42 M. Komuro, J. Taniguchi, S. Inoue, N. Kimura, Y. Tokano, H. Hiroshima, S. Matsui: Imprint characteristics by photo-induced solidification of liquid polymer, Jpn. J. Appl. Phys. **39**, 7075–7079 (2000)

5.43 H. Schulz, H.-C. Scheer, T. Hoffmann, C.M. Sotomayor Torres, K. Pfeiffer, G. Bleidiessel, G. Grützner, C. Cardinaud, F. Gaboriau, M.-C. Peignon, J. Ahopelto, B. Heidari: New polymer materials for nanoimprinting, J. Vac. Sci. Technol. B **18**(4), 1861–1865 (2000)

5.44 X. Cheng, L.J. Guo: A combined-nanoimprint-and-photolithography patterning technique, Microelectron. Eng. **3/4**, 277–282 (2004)

5.45 X. Cheng, L.J. Guo: One-step lithography for various size patterns with a hybrid mask-mold, Microelectron. Eng. **3/4**, 288–293 (2004)

5.46 M. Tormen, L. Businaro, M. Altissimo, F. Romanato, S. Cabrini, F. Perennes, R. Proietti, H.-B. Sun, S. Kawata, E. Di Fabrizio: 3-D patterning by means of nanoimprinting, x-ray and two-photon lithography, Microelectron. Eng. **73/74**, 535–541 (2004)

5.47 X. Sun, L. Zhuang, W. Zhang, S.Y. Chou: Multilayer resist methods for nanoimprint lithography on nonflat surfaces, J. Vac. Sci. Technol. B **16**(6), 3922–3925 (1998)

5.48 A. Lebib, M. Natali, S.P. Li, E. Cambril, L. Manin, Y. Chen, H.M. Janssen, R.P. Sijbesma: Control of the critical dimension with a trilayer nanoimprint lithography procedure, Microelectron. Eng. **57/58**, 411–416 (2001)

5.49 M.W. Lin, H.-L. Chao, J. Hao, E.K. Kim, F. Palmieri, W.C. Kim, M. Dickey, P.S. Ho, C.G. Willson: Planarization for reverse-tone step and flash imprint lithography, Proc. SPIE **6151**, 688–699 (2006)

5.50 W. Trybula: Sematech, AMRC, and Nano. In: *Oral Presentation at Nanoprint and Nanoimprint Tech. (NNT) Conf., Vienna* (2004)

5.51 S. Johnson, D.J. Resnick, D. Mancini, K.J. Nordquist, W.J. Dauksher, K. Gehoski, J.H. Baker, L. Dues, A. Hooper, T.C. Bailey, S.V. Sreenivasan, J.G. Ekerdt, C.G. Willson: Fabrication of multi-tiered structures on step and flash imprint lithography templates, Microelectron. Eng. **67/68**, 221–228 (2003)

5.52 H.-J. Kim, M. Almanza-Workman, B. Garcia, O. Kwon, F. Jeffrey, S. Braymen, J. Hauschildt, K. Junge, D. Larson, D. Stieler, A. Chaiken, B. Cobene, R. Elder, W. Jackson, M. Jam, A. Jeans, H. Luo, P. Mei, C. Perlov, C. Taussig: Roll-to-roll manufacturing of electronics on flexible substrates using self-aligned imprint lithography (SAIL), J. Soc. Inf. Disp. **17**(11), 963–970 (2009)

5.53 Y.P. Kong, H.Y. Lowa, S.W. Pang, A.F. Yee: Duo-mold imprinting of three-dimensional polymeric structures, J. Vac. Sci. Technol. B **22**(6), 3251–3265 (2004)

5.54 T. Borzenko, M. Tormen, G. Schmidt, L.W. Molenkamp: Polymer bonding process for nanolithography, Appl. Phys. Lett. **79**(14), 2246–2248 (2001)

5.55 N. Kehagias, V. Reboud, G. Chansin, M. Zelsmann, C. Jeppesen, C. Schuster, M. Kubenz, F. Reuther, G. Gruetzner, C.M. Sotomayor Torres: Reverse-contact UV nanoimprint lithography for multilayered structure fabrication, Nanotechnology **18**, 175303 (2007)

5.56 C.G. Willson, R.A. Dammel, A. Reiser: Photoresist materials: A historical perspective, Proc. SPIE **3049**, 28–41 (1997)

5.57 M.D. Stewart, C.G. Willson: Photoresists. In: *Encyclopedia of Materials: Science and Technology*, ed. by K.H.J. Buschow, R.W. Cahn, M.C. Flemings, B. Ilschner (Elsevier, Amsterdam 2001) pp. 6973–6978

5.58 B.K. Long, B.K. Keitz, C.G. Willson: Materials for step and flash imprint lithography (S-FIL), J. Mater. Chem. **17**, 3575–3580 (2007)

5.59 M. Vogler, S. Wiedenberg, M. Mühlberger, I. Bergmair, T. Glinsner, H. Schmidt, E.-B. Kley,

G. Grützner: Development of a novel, low-viscosity UV-curable polymer system for UV-nanoimprint lithography, Microelectron. Eng. **84**, 984–988 (2007)
5.60 F.A. Houle, C.T. Rettner, D.C. Miller, R. Sooriyakumaran: Antiadhesion considerations for UV nanoimprint lithography, Appl. Phys. Lett. **90**, 213103 (2007)
5.61 F.A. Houle, E. Guyer, D.C. Miller, R. Dauskardt: Adhesion between template materials and UV-cured nanoimprint resists, J. Vac. Sci. Technol. B **25**(4), 1179–1185 (2007)
5.62 H. Schift, S. Park, J. Gobrecht, S. Saxer, F. Meier, W. Raupach, K. Vogelsang: Hybrid bendable stamp copies for molding fabricated by nanoimprint, Microelectron. Eng. **78/79**, 605–611 (2005)
5.63 R.W. Jaszewski, H. Schift, B. Schnyder, A. Schneuwly, P. Gröning: The deposition on anti-adhesive ultra-thin teflon-like films and their interaction with polymers during hot embossing, Appl. Surf. Sci. **143**, 301–308 (1999)
5.64 H. Schift, S. Saxer, S. Park, C. Padeste, U. Pieles, J. Gobrecht: Controlled co-evaporation of silanes for nanoimprint stamps, Nanotechnology **16**, S171–175 (2005)
5.65 N. Roos, T. Luxbacher, T. Glinsner, K. Pfeiffer, H. Schulz, H.-C. Scheer: Nanoimprint lithography with a commercial 4 inch bond system for hot embossing, Proc. SPIE **4343**, 427–436 (2001)
5.66 T. Haatainen, J. Ahopelto, G. Grützner, M. Fink, K. Pfeiffer: Step and stamp imprint lithography using a commercial flip chip bonder, Proc. SPIE **3997**, 874–879 (2000)
5.67 H. Tan, A. Gilbertson, S.Y. Chou: Roller nanoimprint lithography, J. Vac. Sci. Technol. B **16**(6), 3926–3928 (1998)
5.68 R. Ji, M. Hornung, M.A. Verschuuren, R. van de Laar, J. van Eekelen, U. Plachetka, M. Moeller, C. Moormann: UV enhanced substrate conformal imprint lithography (UV-SCIL) technique for photonic crystals patterning in LED manufacturing, Microelectron. Eng. **87**(5–8), 963–967 (2010)
5.69 M. Tormen, E. Sovernigo, A. Pozzato, M. Pianigiani, M. Tormen: Sub-100 μs nanoimprint lithography at wafer scale, Microelectron. Eng. **141**, 21–26 (2015)
5.70 S.Y. Chou, C. Keimel, J. Gu: Ultrafast and direct imprint of nanostructures in silicon, Nature **417**, 835–837 (2002)
5.71 H. Schift: Roll embossing and roller imprint. In: *Science and New Technol. Nanoimprint Adv. Technol. Appl. Nanoimprint*, ed. by Y. Hirai (Frontier, Tokyo 2006) pp. 74–89
5.72 N. Kooy, K. Mohamed, L.T. Pin, O.S. Guan: A review of roll-to-roll nanoimprint lithography, Nanoscale Res. Lett. **9**, 320 (2014)
5.73 H. Schift, R.W. Jaszewski, C. David, J. Gobrecht: Nanostructuring of polymers and fabrication of interdigitated electrodes by hot embossing lithography, Microelectron. Eng. **46**, 121–124 (1999)
5.74 L.J. Heyderman, H. Schift, C. David, B. Ketterer, M. Auf der Maur, J. Gobrecht: Nanofabrication using hot embossing lithography and electroforming, Microelectron. Eng. **57/58**, 375–380 (2001)
5.75 L.J. Heyderman, B. Ketterer, D. Bächle, F. Glaus, B. Haas, H. Schift, K. Vogelsang, J. Gobrecht, L. Tiefenauer, O. Dubochet, P. Surbled, T. Hessler: High volume fabrication of customised nanopore membrane chips, Microelectron. Eng. **67/68**, 208–213 (2003)
5.76 H. Schift, P. Urwyler, P.M. Kristiansen, J. Gobrecht: Nanoimprint lithography process chains for the fabrication of micro- and nanodevices, J. Micro/Nanolith. MEMS MOEMS **10**, 031303 (2014)
5.77 T. Higashiki, T. Nakasugi, I. Yoneda: Nanoimprint lithography and future patterning for semiconductor devices, J. Micro/Nanolith. MEMS MOEMS **10**(4), 043008 (2011)
5.78 T.R. Albrecht, H. Arora, V. Ayanoor-Vitikkate, J.-M. Beaujour, D. Bedau, D. Berman, A.L. Bogdanov, Y.-A. Chapuis, J. Cushen, E.E. Dobisz, G. Doerk, H. Gao, M. Grobis, B. Gurney, W. Hanson, O. Hellwig, T. Hirano, P.-O. Jubert, D. Kercher, J. Lille, Z. Liu, C.M. Mate, Y. Obukhov, K.C. Patel, K. Rubin, R. Ruiz, M. Schabes, L. Wan, D. Weller, T.-W. Wu, E. Yan: Bit patterned magnetic recording: theory, media fabrication, and recording performance, IEEE Trans. Magn. **51**(5), 44 (2015)
5.79 G.M. McClelland, M.W. Hart, C.T. Rettner, M.E. Best, K.R. Carter, B.D. Terris: Nanoscale patterning of magnetic islands by imprint lithography using a flexible mold, Appl. Phys. Lett. **81**, 1483–1485 (2002)
5.80 Y.-C. Lee, S.-C. Yeh, Y.-Y. Chou, P.-J. Tsai, J.-W. Pan, H.-M. Chou, C.-H. Hou, Y.-Y. Chang, M.-S. Chu, C.-H. Wu, C.-H. Ho: High-efficiency InGaN-based LEDs grown on patterned sapphire substrates using nanoimprinting technology, Microelectron. Eng. **105**, 86–90 (2013)
5.81 S.-W. Ahn, K.-D. Lee, J.-S. Kim, S.H. Kim, S.H. Lee, J.-D. Park, P.-W. Yoon: Fabrication of subwavelength aluminum wire grating using nanoimprint lithography and reactive ion etching, Microelectron. Eng. **78/79**, 314–318 (2005)
5.82 H. Schift, S. Park, C.-G. Choi, C.-S. Kee, S.-P. Han, K.-B. Yoon, J. Gobrecht: Fabrication process for polymer photonic crystals using nanoimprint lithography, Nanotechnology **16**, S261–S265 (2005)
5.83 C.-Y. Chao, L.J. Guo: Polymer microring resonators fabricated by nanoimprint technique, J. Vac. Sci. Technol. B **20**, 2862–2866 (2002)
5.84 J. Wang, X. Sun, L. Chen, S.Y. Chou: Direct nanoimprint of submicron organic light-emitting structures, Appl. Phys. Lett. **75**, 2767–2769 (1999)
5.85 X. Cheng, Y. Hong, J. Kanicki, L.J. Guo: High-resolution organic polymer light-emitting pixels fabricated by imprinting technique, J. Vac. Sci. Technol. B **20**, 2877–2880 (2002)
5.86 D. Nilsson, S. Balslev, A. Kristensen: A microfluidic dye laser fabricated by nanoimprint lithography in a highly transparent and chemically resistant cyclo-olefin copolymer (COC), J. Micromech. Microeng. **15**, 296–300 (2005)
5.87 C. Clavijo Cedeno, J. Seekamp, A.P. Kam, T. Hoffmann, S. Zankovych, C.M. Sotomayor Torres,

C. Menozzi, M. Cavallini, M. Murgia, G. Ruani, F. Biscarini, M. Behl, R. Zentel, J. Ahopelto: Nanoimprint lithography for organic electronics, Microelectron. Eng. **61/62**, 25–31 (2002)
5.88 L.I. Segerink, J.C.T. Eijkel: Nanofluidics in point of care applications, Lab Chip **14**, 3201–3205 (2014)
5.89 A. Manz, N. Graber, H.M. Widmer: Miniaturized total chemical analysis systems: A novel concept for chemical sensing, Sens. Actuators B **1**, 244–248 (1990)
5.90 E. Verpoorte, N.F. De Rooij: Microfluidics meets MEMS, Proc. IEEE **91**, 930–953 (2003)
5.91 A. Pepin, P. Youinou, V. Studer, A. Lebib, Y. Chen: Nanoimprint lithography for the fabrication of DNA electrophoresis chips, Microelectron. Eng. **61/62**, 927–932 (2002)
5.92 K. Emoto, F. Sakai, C. Sato, Y. Takabayashi, H. Nakano, T. Takabayashi, K. Yamamoto, T. Hattori, M. Hiura, T. Ando, Y. Kawanobe, H. Azuma, T. Iwanaga, J. Choi, A. Aghili, C. Jones, J.W. Irving, B. Fletcher, Z. Ye: Defectivity and particle reduction for mask life extension, and imprint mask replication for high-volume semiconductor manufacturing, Proc. SPIE **9777**, 97770C (2016), Alternative Lithographic Technol. **VIII**
5.93 S.Y. Chou: Patterned magnetic nanostructures and quantized magnetic disks, Proc. IEEE **85**, 652–671 (1997)
5.94 M.N. Baibich, J.M. Broto, A. Fert, F.N. Van Dau, F. Petroff, P. Eitenne, G. Creuzet, A. Friederich, J. Chazelas: Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices, Phys. Rev. Lett. **61**, 2472–2475 (1988)
5.95 Y. Li, A.K. Menon: Magnetic recording technologies: Overview. In: *Encyclopedia of Materials: Science and Technology*, ed. by K.H.J. Buschow, R.W. Cahn, M.C. Flemings, B. Ilschner (Elsevier, Amsterdam 2001) pp. 4948–4957
5.96 D. Wachenschwanz, W. Jiang, E. Roddick, A. Homola, P. Dorsey, B. Harper, D. Treves, C. Bajorek: Design of a manufacturable discrete track recording medium, IEEE Trans. Magn. **41**, 670–675 (2005)
5.97 S.Y. Chou, M. Wei, P.R. Krauss, P.B. Fisher: Study of nanoscale magnetic structures fabricated using electron beam lithography and quantum magnetic disk, J. Vac. Sci. Technol. B **12**, 3695–3698 (1994)
5.98 S.Y. Chou: Patterned magnetic nanostructures and quantized magnetic disks, Proc. IEEE **85**, 652–671 (1997)
5.99 P. Lalanne, M. Hutley: Artificial media optical properties – Subwavelength scale. In: *Enclopedia of Optical Engineering*, ed. by R.G. Driggers, C. Hoffman (Marcel Dekker, New York 2003) pp. 62–71
5.100 Z. Yu, W. Wu, L. Chen, S. Chou: Fabrication of large area 100 nm pitch grating by spatial frequency doubling an nanoimprint lithography for subwavelength optical applications, J. Vac. Sci. Technol. B **19**, 2816–2819 (2001)
5.101 S. Ahn, J. Yang, M. Miller, M. Ganapathisubramanian, M. Menezes, J. Choi, F. Xu, D.J. Resnick, S.V. Sreenivasan: High performance wire grid polarizers using jet and flash imprint lithography, Proc. SPIE **8680**, 86800W (2013), Alternative Lithographic Technol. **V**
5.102 M.G. Kang, L.J. Guo: Nanoimprinted semitransparent metal electrodes and their application in organic light-emitting diodes, Adv. Mater. **19**(10), 1391 (2007)
5.103 A. Boltasseva: Plasmonic components fabrication via nanoimprint, J. Opt. A: Pure Appl. Opt. **11**, 114001 (2009)
5.104 A. Kristensen, J.K.W. Yang, S.I. Bozhevolnyi, S. Link, P. Nordlander, N.J. Halas, N.A. Mortensen: Plasmonic color generation, Nat. Rev. Mater. **2**, 1–14 (2016)
5.105 G. Kang, J. Yoo, J. Ahn, K. Kim: Transparent dielectric nanostructures for efficient light management in optoelectronic applications, Nano Today **10**(1), 22–47 (2015)
5.106 S.H. Kim, K.-D. Lee, J.-Y. Kim, M.-K. Kwon, S.-J. Park: Fabrication of photonic crystal structures on light emitting diodes by nanoimprint lithography, Nanotechnology **18**, 055306 (2007)
5.107 H. Chen, Q. Zhang, S.Y. Chou: Patterning of light-extraction nanostructures on sapphire substrates using nanoimprint and ICP etching with different masking materials, Nanotechnology **26**(8), 085302 (2015)
5.108 A. Mellor, H. Hauser, C. Wellens, J. Benick, J. Eisenlohr, M. Peters, A. Guttowski, I. Tobías, A. Martí, A. Luque, B. Bläsi: Nanoimprinted diffraction gratings for crystalline silicon solar cells: Implementation, characterization and simulation, Opt. Express **21**(S2), A295 (2013)
5.109 D. Nilsson, T. Nielsen, A. Kristensen: Solid state micro-cavity dye lasers fabricated by nanoimprint lithography, Rev. Sci. Instrum. **75**, 4481–4486 (2004)
5.110 C. Clavijo Cedeno, J. Seekamp, A.P. Kam, T. Hoffmann, S. Zankovych, C.M. Sotomayor Torres, C. Menozzi, M. Cavallini, M. Murgia, G. Ruani, F. Biscarini, M. Behl, R. Zentel, J. Ahopelto: Nanoimprint lithography for organic electronics, Microelectron. Eng. **61/62**, 25–31 (2002)
5.111 L.J. Guo, X. Cheng, C.Y. Chao: Fabrication of photonic nanostructures in nonlinear optical polymers, J. Modern Opt. **49**, 663–673 (2002)
5.112 L.J. Guo: Recent progress in nanoimprint technology and its applications, J. Phys. D: Appl. Phys. **37**, R123–R141 (2004)
5.113 C.-Y. Chao, L.J. Guo: Reduction of surface scattering loss in polymer microrings using thermal-reflow technique, IEEE Photonics Technol. Lett. **16**, 1498–1500 (2004)
5.114 H.C. Hoch, L.W. Jelinski, H.C. Craighead (Eds.): *Nanofabrication and Biosystems: Integrating Materials Science, Engineering and Biology* (Cambridge Univ. Press, Cambridge 1996)
5.115 H.G. Craighead: Nanoelectromechanical systems, Science **290**, 1532–1535 (2000)
5.116 L.R. Huang, J.O. Tegenfeldt, J.J. Kraeft, J.C. Sturm, R.H. Austin, E.C. Cox: A DNA prism for high-speed continous frationation of large DNA molecules, Nat. Biotechnol. **20**, 1048–1051 (2002)

5.117 H.G. Craighead: Nanostructure science and technology: Impact and prospects for biology, J. Vac. Sci. Technol. A **21**, S216–S221 (2003)
5.118 J.O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W.W. Reisner, R. Riehn, Y.M. Wang, E.C. Cox, J.C. Sturm, P. Silberzan, R.H. Austin: The dynamics of genomic-length DNA molecules in 100-nm channels, Proc. Natl. Acad. Sci. USA **101**, 10979–10983 (2004)
5.119 L.J. Guo, X. Cheng, C.-F. Chou: Fabrication of size-controllable nanofluidic channels by nanoimprinting and its application for DNA stretching, Nano Lett. **4**, 69–73 (2004)
5.120 L.H. Thamdrup, A. Klukowska, A. Kristensen: Stretching DNA in polymer nanochannels fabricated by thermal imprint in PMMA, Nanotechnology **19**, 125301 (2008)
5.121 J.O. Tegenfeldt, C. Prinz, H. Cao, R.L. Huang, R.H. Austin, S.Y. Chou, E.C. Cox, J.C. Sturm: Micro- and nanofluidics for DNA analysis, Anal. Bioanal. Chem. **378**, 1678–1692 (2004)
5.122 M.J. Dalby, N. Gadegaard, R. Tare, A. Andar, M.O. Riehle, P. Herzyk, C.D.W. Wilkinson, R.O.C. Oreff: The control of human mesenchymal cell differentiation using nanoscale symmetry and disorder, Nat. Mater. **6**, 997–1003 (2007)
5.123 J. Hasan, K. Chatterjee: Recent advances in engineering topography mediated antibacterial surfaces, Nanoscale **7**, 15568 (2015)
5.124 K. Seunarine, D.O. Meredith, M.O. Riehle, C.D.W. Wilkinson, N. Gadegaard: Biodegradable polymer tubes with lithographically controlled 3-D micro- and nanotopography, Microelectron. Eng. **85**(5/6), 1350–1354 (2008)
5.125 A. Kapr: Johann Gutenberg: The man and his invention, http://www.gutenberg.de/publ.htm (1996)
5.126 J.A. Liddle, G.M. Gallatin: Nanomanufacturing: A Perspective, ACS Nano (2016) doi:10.1021/acsnano.5b03299
5.127 M. Beck, B. Heidari: Nanoimprint lithography for high volume HDI manufacturing, OnBoard Technol. **Sept.**, 52–55 (2006), http://www.Onboard-Technology.com

6. Stamping Techniques for Micro- and Nanofabrication

John A. Rogers, Etienne Menard

Soft-lithographic techniques that use rubber stamps and molds provide simple means to generate patterns with lateral dimensions that can be much smaller than 1 μm and can even extend into the single nanometer regime. These methods rely on the use of soft elastomeric elements typically made out of the polymer poly(dimethylsiloxane). The first section of this chapter presents the fabrication techniques for these elements together with data and experiments that provide insights into the fundamental resolution limits. Next, several representative soft-lithography techniques based on the use of these elements are presented:

1. Microcontact printing, which uses molecular *inks* that form self-assembled monolayers.
2. Near- and proximity-field photolithography for producing two- and three-dimensional structures with subwavelength resolution features.
3. Nanotransfer printing, where soft or hard stamps print single or multiple layers of solid inks with feature sizes down to 100 nm.

The chapter concludes with descriptions of some device-level applications that highlight the patterning capabilities and potential commercial uses of these techniques.

6.1 **High-Resolution Stamps** 144
6.2 **Microcontact Printing** 146
6.3 **Nanotransfer Printing** 148
6.4 **Applications** 152
6.4.1 Unconventional Electronic Systems 152
6.4.2 Lasers and Waveguide Structures 157
6.5 **Conclusions** 159
References 159

There is considerable interest in methods that can be used to build structures that have micron or nanometer dimensions. Historically, research and development in this area has been driven mainly by the needs of the microelectronics industry. The spectacularly successful techniques that have emerged from those efforts – such as photolithography and electron beam lithography – are extremely well suited to the tasks for which they were principally designed: forming structures of radiation-sensitive materials (including photoresists or electron beam resists) on ultraflat glass or semiconductor surfaces. Significant challenges exist in adapting these methods for new emerging applications and areas of research that require patterning of unusual systems and materials, (including those in biotechnology and plastic electronics), structures with nanometer dimensions (below 50–100 nm), large areas in a single step (larger than a few square centimeters), or nonplanar (rough or curved) surfaces. These established techniques also have the disadvantage of high capital and operational costs. As a result, some of the oldest and conceptually simplest forms of lithography – embossing, molding, stamping, writing, and so on – are now being reexamined for their potential to serve as the basis for nanofabrication techniques that can avoid these limitations [6.1]. Considerable progress has been made in the last few years, mainly by combining these approaches or variants of them with new materials, chemistries, and processing techniques. This chapter highlights some recent advances in high-resolution printing methods, in which a *stamp* forms a pattern of *ink* on a surface that it contacts. It focuses on approaches whose capabilities, level of development, and demonstrated applications indicate a strong potential for widespread use, especially in areas where conventional methods are unsuitable.

Contact printing involves the use of an element with surface relief (the *stamp*) to transfer material applied to its surface (the *ink*) to locations on a substrate that it contacts. The printing press, one of the earliest manufacturable implementations of this approach, was introduced by Gutenberg in the fifteenth century. Since

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_6

then, this general approach has been used almost exclusively for producing printed text or images with features that are 100 μm or larger in their smallest dimension. The resolution is determined by the nature of the ink and its interaction with the stamp and/or substrate, the resolution of the stamp, and the processing conditions that are used for printing or to convert the pattern of ink into a pattern of functional material. This chapter focuses on (1) printing techniques that are capable of micron and nanometer resolution, and (2) their use for fabricating key elements of active electronic or optical devices and subsystems. It begins with an overview of some methods for fabricating high-resolution stamps and then illustrates two different ways that these stamps can be used to print patterns of functional materials. Applications that highlight the capabilities of these techniques and the performances of systems that are constructed with them are also presented.

6.1 High-Resolution Stamps

The printing process can be separated into two parts: fabrication of the stamp and the use of this stamp to pattern features defined by the relief on its surface. These two processes are typically quite different, although it is possible in some cases to use patterns generated by a stamp to produce a replica of that stamp. The structure from which the stamp is derived, which is known as the *master*, can be fabricated with any technique that is capable of producing well-defined structures of relief on a surface. This master can then be used directly as the stamp, or to produce stamps via molding or printing procedures. It is important to note that the technique for producing the master does not need to be fast or low in cost. It also does not need to possess many other characteristics that might be desirable for a given patterning task: it is used just once to produce a master, which is directly or indirectly used to fabricate stamps. Each one of these stamps can then be used many times for printing.

In a common approach for the high-resolution techniques that are the focus of this chapter, an established lithographic technique, such as one of those developed for the microelectronics industry, defines the master. Figure 6.1 schematically illustrates typical processes. Here, photolithography patterns a thin layer of resist onto a silicon wafer. Stamps are generated from this structure in one of two ways: by casting against this master, or by etching the substrate with the patterned resist as a mask. In the first approach, the master itself can be used multiple times to produce many stamps, typically using a light or heat-curable prepolymer. In the second, the etched substrate serves as the stamp. Additional stamps can be generated either by repeating the lithography and etching, or by using the original stamp to print replica stamps. For minimum lateral feature sizes that are greater than ≈ 1–2 μm, contact- or proximity-mode photolithography with a mask produced by direct write photolithography represents a convenient method of fabricating the master. For features smaller than ≈ 2 μm, several different techniques can be used [6.2], including:

1. Projection mode photolithography [6.3]
2. Direct write electron beam (or focused ion beam) lithography [6.4, 5]

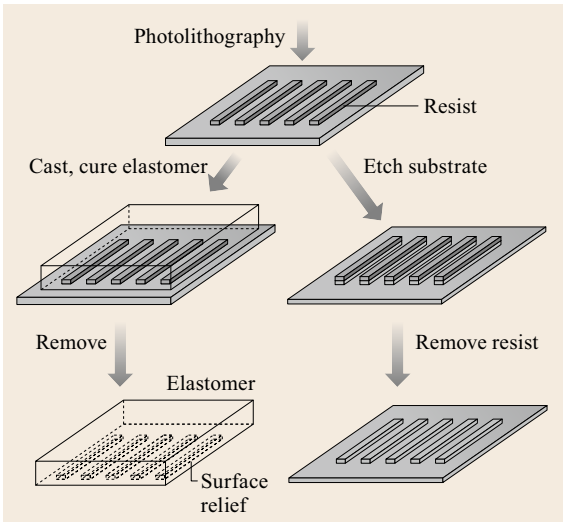


Fig. 6.1 Schematic illustration of two methods for producing high-resolution stamps. The first step involves patterning a thin layer of some radiation-sensitive material, known as the resist, on a flat substrate, such as a silicon wafer. It is convenient to use an established technique, such as photolithography or electron beam lithography, for this purpose. This structure, known as the *master*, is converted to a stamp either by etching or by molding. In the first case, the resist acts as a mask for etching the underlying substrate. Removing the resist yields a stamp. This structure can be used directly as a stamp to print patterns or to produce additional stamps. In the molding approach, a prepolymer is cast against the relief structure formed by the patterned resist on the substrate. Curing (thermally or optically) and then peeling the resulting polymer away from the substrate yields a stamp. In this approach, many stamps can be made with a single *master* and each stamp can be used many times

3. Scanning probe lithography [6.6–9]
4. Laser interference lithography [6.10].

The first approach requires a photomask generated by some other method, such as direct write photolithography or electron beam lithography. The reduction (typically 4×) provided by the projection optics relaxes the resolution requirements on the mask and enables features as small as ≈ 90 nm when deep ultraviolet radiation and phase shifting masks are used. The costs for these systems are, however, very high and their availability for general research purposes is limited. The second method is flexible in the geometry of patterns that can be produced, and the writing systems are highly developed: 30–50 nm features can be achieved with commercial systems [6.11], and < 10 nm features are possible with research tools, as first demonstrated more than 25 years ago by *Broers* [6.12]. The main drawbacks of this method are that it is relatively slow and it is difficult to pattern large areas. Like projection-mode photolithography, it can be expensive. The third

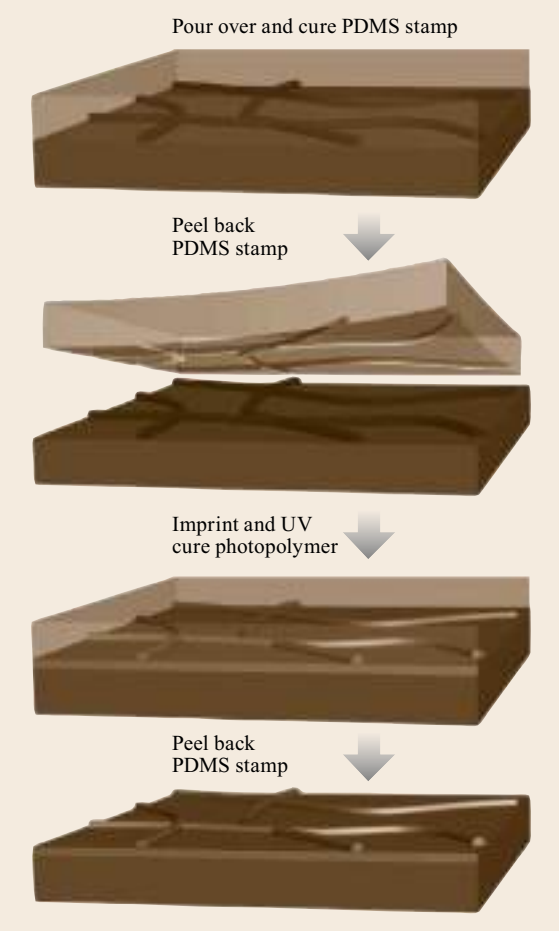


Fig. 6.2 Schematically illustrates a process for examining the ultimate limits in resolution of soft lithographic methods. The approach uses a SWNT master to create a PDMS mold with nanoscale relief features. Soft nanoimprint lithography transfers the relief on the PDMS to that on the surface of an ultraviolet curable photopolymer film

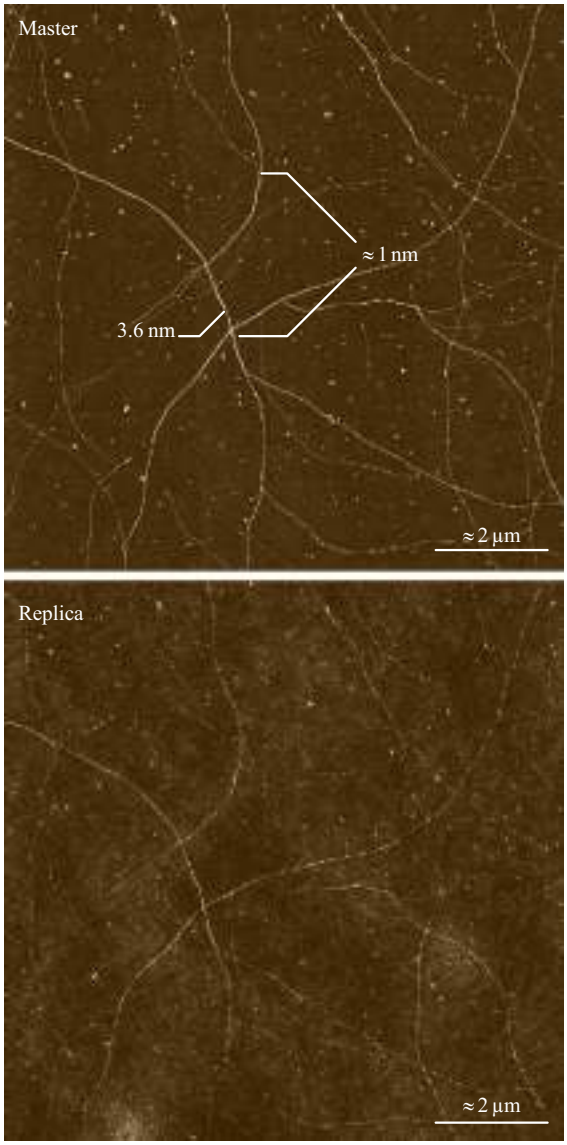


Fig. 6.3 Atomic force micrographs (*top* picture) of a *master* that consists of a submonolayer of single-walled carbon nanotubes (SWNTs; diameter between 0.5 and 5 nm) grown on a SiO_2/Si wafer. The *bottom* atomic force micrograph shows a replica of the relief structures in poly(urethane). These results indicate effective operation of a PDMS stamp for soft imprint lithography at the single nanometer scale

method, scanning probe lithography, is quite powerful in principle, but the tools are not as well established as those for other approaches. This technique has atomic resolution, but its writing speed can be lower and the areas that can be patterned are smaller than electron beam systems. Interference lithography provides a powerful, low-cost tool for generating periodic arrays of features with dimensions down to 100–200 nm; smaller sizes demand ultraviolet lasers, and patterns with aperiodic or nonregular features are difficult to produce.

In order to evaluate the ultimate resolution limit of the soft lithography methods, masters with relief structures in the single nanometer range must be fabricated. A simple method, presented in Fig. 6.2, uses submonolayer coverage of single-walled carbon nanotubes (SWNT) grown, by established chemical vapor deposition techniques, on an ultraflat silicon wafer. The SWNT, which have diameters (heights and widths) in the 0.5–5 nm range, are molded on the bottom surface of a poly(dimethylsiloxane) (PDMS) stamp generated by casting and curing against this master. Such a mold can be used to replicate the relief structure into a variety of photocurable polymers in a kind of soft nanoimprinting technique [6.13–15]. A single mold can give exceedingly high resolution, approaching the single nanometer scale range (comparable to a few bond lengths in the polymer backbone), as can be seen in Fig. 6.3 [6.16]. These results demonstrate the extreme efficiency of the basic soft lithographic procedure for generating and using elastomeric elements. The ultimate limits are difficult to predict, due to substantial uncertainties surrounding the polymer physics and chemistry that dominates in the nanometer regime.

6.2 Microcontact Printing

Microcontact printing (μCP) [6.17] is one of several soft lithographic techniques – replica molding, micromolding in capillaries, microtransfer molding, near-field conformal photolithography using an elastomeric phase-shifting mask, and so on – that have been developed as alternatives to established methods for micro- and nanofabrication [6.18–22]. μCP uses an elastomeric element (usually poly(dimethylsiloxane) – PDMS) with high-resolution features of relief as a stamp to print patterns of chemical inks. It was mainly developed for use with inks that form self-assembled monolayers (SAMs) of alkanethiolates on gold and silver. The procedure for carrying out μCP in these systems is remarkably simple: a stamp, inked with a solution of alkanethiol, is brought into contact with the surface of a substrate in order to transfer ink molecules to regions where the stamp and substrate contact. The resolution and effectiveness of μCP rely on conformal contact between the stamp and the surface of the substrate, rapid formation of highly ordered monolayers [6.23], and the autophobicity of the SAM, which effectively blocks the reactive spreading of the ink across the surface [6.24]. It can pattern SAMs over relatively large areas (≈ up to 0.25 ft^2 has been demonstrated in prototype electronic devices) in a single impression [6.25]. The edge resolution of SAMs printed onto thermally evaporated gold films is on the order of 50 nm, as determined by lateral force microscopy [6.26]. Microcontact printing has been used with a range of different SAMs on various substrates [6.18]. Of these, alkanethiolates on gold, silver, and palladium [6.27] presently give the highest resolution. In many cases, the mechanical properties of the stamp limit the sizes of the smallest features that can be achieved: the most commonly used elastomer (Sylgard 184, Dow Corning) has a low modulus, which can lead to mechanical collapse or sagging for features of relief with aspect ratios greater than ≈ 2 or less than ≈ 0.05. Stamps fabricated with high modulus elastomers avoid some of these problems [6.28, 29]. Conventional stamps are also susceptible to in-plane mechanical strains that can cause distortions in the printed patterns. Composite stamps that use thin elastomer layers on stiff supports are effective at minimizing this source of distortion [6.30]. Methods for printing that avoid direct mechanical manipulation of the stamp can reduce distortions with conventional and composite stamps [6.25]. This approach has proven effective in large-area flexible circuit applications that require accurate multilevel registration.

The patterned SAM can be used either as a resist in selective wet etching or as a template in selective deposition to form structures of a variety of materials: metals, silicon, liquids, organic polymers and even biological species. Figure 6.4 schematically illustrates the use of μCP and wet etching to pattern a thin film of Au. Figure 6.5 shows scanning electron micrograph (SEM) images of nanostructures of gold (20 nm thick, thermally evaporated with a 2.5 nm layer of Ti as an adhesion promoter) and silver (≈ 100 nm thick formed by electroless deposition using commercially available plating baths) [6.31] that were

fabricated using this approach. In the first and second examples, the masters for the stamps consisted of photoresist patterned on silicon wafers with projection and contact mode photolithography, respectively. Placing these masters in a desiccator for ≈ 1 h with a few drops of tridecafluoro-1,1,2,2-tetrahydrooctyl-1-trichlorosilane forms a silane monolayer on the exposed native oxide of the silicon. This monolayer prevents adhesion of the master to PDMS (Sylgard 184), which is cast and cured from a 10 : 1 mixture of prepolymer and curing agent. Placing a few drops of a ≈ 1 mM solution of hexadecanethiol (HDT) in ethanol on the surface of the stamps and then blowing them dry with a stream of nitrogen prepares them for printing. Contacting the metal film for a few seconds with the stamp produces a patterned self-assembled monolayer (SAM) of HDT. An aqueous etchant (1 mM $K_4Fe(CN)_6$, 10 mM $K_3Fe(CN)_6$, and 0.1 M $Na_2S_2O_3$) removes the unprinted regions of the silver [6.32]. A similar solution (1 mM $K_4Fe(CN)_6$, 10 mM $K_3Fe(CN)_6$, 1.0 M KOH, and 0.1 M $Na_2S_2O_3$) can be used to etch the bare gold [6.33]. The results in Fig. 6.5 show that the roughness on the edges of the patterns is ≈ 50–100 nm. The resolution is determined by the grain size of the metal films, the isotropic etching process, slight reactive spreading of the inks, and edge disorder in the patterned SAMs.

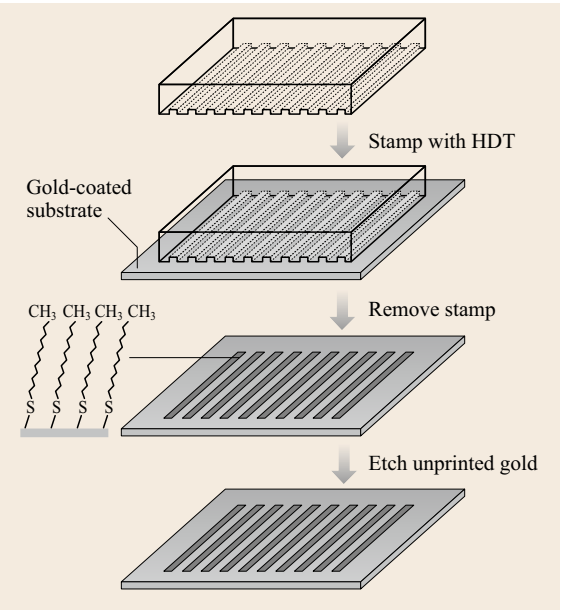


Fig. 6.4 Schematic illustration of microcontact printing. The first step involves *inking* a *stamp* with a solution of a material that is capable of forming a self-assembled monolayer (SAM) on a substrate that will be printed. In the case illustrated here, the ink is a millimolar concentration of hexadecanethiol (HDT) in ethanol. Directly applying the ink to the surface of the stamp with a pipette prepares the stamp for printing. Blowing the surface of the stamp dry and contacting it to a substrate delivers the ink to areas where the stamp contacts the substrate. The substrate consists of a thin layer of Au on a flat support. Removing the stamp after a few seconds of contact leaves a patterned SAM of HDT on the surface of the Au film. The printed SAM can act as a resist for the aqueous-based wet etching of the exposed regions of the Au. The resulting pattern of conducting gold can be used to build devices of various types

The structures of Fig. 6.5 were formed on the flat surfaces of silicon wafers (left image) and glass slides (right image). An attractive feature of μCP and certain other contact printing techniques is their ability to pattern features with high resolution on highly curved or rough surfaces [6.22, 34, 35]. This type of patterning task is difficult or impossible to accomplish with photolithography due to its limited depth of focus and the difficulty involved with casting uniform films of photoresist on nonflat surfaces. Figure 6.6 shows, as an example, a straightforward approach for high-resolution printing on the highly curved surfaces of optical fibers. Here, simply rolling the fiber over an inked stamp prints a pattern on the entire outer surface of the fiber. Simple staging systems allow alignment of features to the fiber axis; they also ensure registration of the pattern from one side of the fiber to the other [6.20]. Figure 6.7 shows 3 μm wide lines and spaces printed onto the surface of a single mode optical fiber (diameter 125 μm). The bottom frame shows a freestanding metallic struc-

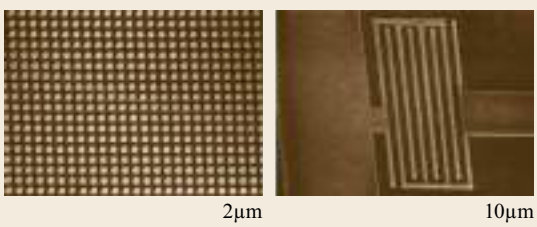


Fig. 6.5 Scanning electron micrographs of typical structures formed by microcontact printing a self-assembled monolayer ink of hexadecanethiol onto a thin metal film followed by etching of the unprinted areas of the film. The *left frame* shows an array of Au dots (20 nm thick) with ≈ 500 nm diameters. The *right frame* shows a printed structure of Ag (100 nm thick) in the geometry of interdigitated source/drain electrodes for a transistor in a simple inverter circuit. The edge resolution of patterns that can be easily achieved with microcontact printing is 50–100 nm

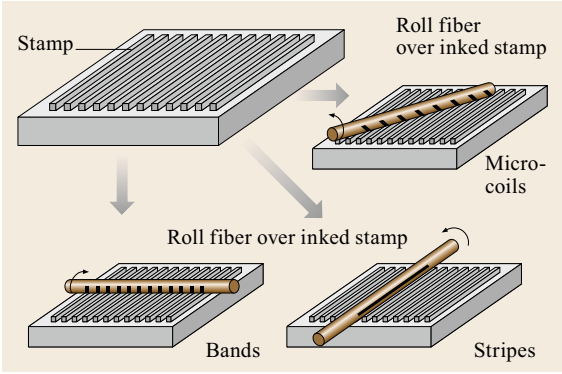


Fig. 6.6 Schematic illustration of a simple method to print lines on the surfaces of optical fibers. Rolling a fiber over the inked stamp prints a pattern onto the fiber surface. Depending on the orientation of the fiber axis with the line stamp illustrated here, it is possible, in a single rotation of the fiber, to produce a continuous microcoil, or arrays of bands or stripes

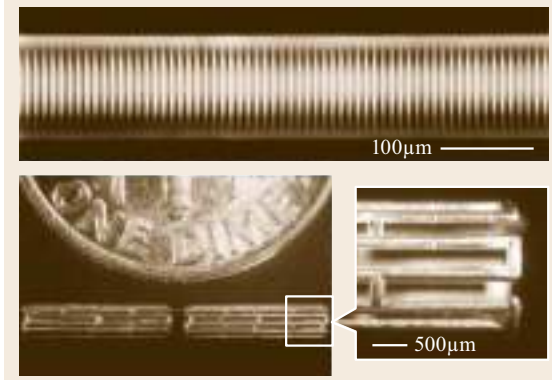


Fig. 6.7 Optical micrographs of some three-dimensional microstructures formed by microcontact printing on curved surfaces. The *top* frame shows an array of 3 μm lines of Au (20 nm)/Ti (1.5 nm) printed onto the surface of an optical fiber. This type of structure can be used as an integrated photomask for producing mode-coupling gratings in the core of the fiber. The *bottom* frames show a free-standing metallic microstructure formed by (a) microcontact printing and etching a thin film (100 nm thick) of Ag on the surface of a glass microcapillary tube, (b) electroplating the Ag to increase its thickness (to tens of micrometer) and (c) etching away the glass microcapillary with concentrated hydrofluoric acid. The structure shown here has the geometry and mechanical properties of an intravascular stent, which is a biomedical device commonly used in balloon angioplasty

ture with the geometry and mechanical properties of an intravascular stent, which is a biomedical device that is commonly used in balloon angioplasty procedures. In this latter case μCP followed by electroplating generated the Ag microstructure on a sacrificial glass cylinder that was subsequently etched away with concentrated hydrofluoric acid [6.36]. Other examples of microcontact printing on nonflat surfaces (low cost plastic sheets and optical ridge waveguides) appear in the Sect. 6.4.

6.3 Nanotransfer Printing

Nanotransfer printing (nTP) is a more recent high-resolution printing technique, which uses surface chemistries as interfacial *glues* and *release* layers (rather than *inks*, as in μCP) to control the transfer of solid material layers from relief features on a stamp to a substrate [6.37–39]. This approach is purely additive (material is only deposited in locations where it is needed), and it can generate complex patterns of single or multiple layers of materials with nanometer resolution over large areas in a single process step. It does not suffer from surface diffusion or edge disorder in the patterned inks of μCP, nor does it require post-printing etching or deposition steps to produce structures of functional materials. The method involves four components:

1. A stamp (rigid, flexible, or elastomeric) with relief features in the geometry of the desired pattern.
2. A method for depositing a thin layer of solid material onto the raised features of this stamp.
3. A means of bringing the stamp into intimate physical contact with a substrate.
4. Surface chemistries that prevent adhesion of the deposited material to the stamp and promote its strong adhesion to the substrate.

nTP has been demonstrated with SAMs and other surface chemistries for printing onto flexible and rigid substrates with hard inorganic and soft polymer stamps. Figure 6.8 presents a set of procedures for using nTP to pattern a thin metal bilayer of Au/Ti with a surface transfer chemistry that relies on a dehydration reaction [6.37]. The process begins with fabrication of a suitable stamp. Elastomeric stamps can be built using the same casting and curing procedures described for μCP. Rigid stamps can be fabricated by (1) patterning resist (such as electron beam resist or photoresist) on a substrate (such as Si or GaAs), (2) etching the exposed regions of the substrate with an anisotropic reactive ion etch, and (3) removing the resist, as illustrated in

Fig. 6.1. For both types of stamps, careful control of the lithography and the etching steps yields features of relief with nearly vertical or slightly reentrant sidewalls. The stamps typically have depths of relief $> 0.2\,\mu m$ for patterning metal films with thicknesses $< 50\,nm$.

Electron beam evaporation of Au (20 nm; 1 nm/s) and Ti (5 nm; 0.3 nm/s) generates uniform metal bilayers on the surfaces of the stamp. A vertical, collimated flux of metal from the source ensures uniform deposition only on the raised and recessed regions of relief. The gold adheres poorly to the surfaces of stamps made of GaAs, PDMS, glass, or Si. In the process of Fig. 6.8, a fluorinated silane monolayer acts to reduce the adhesion further when a Si stamp (with native oxide) is used. The Ti layer serves two purposes: (1) it promotes adhesion between the Au layer and the substrate after pattern transfer, and (2) it readily forms a ≈ 3 nm oxide layer at ambient conditions, which provides a surface where the dehydration reaction can take place. Exposing the titanium oxide (TiO_x) surface to an oxygen plasma breaks bridging oxygen bonds, thus creating defect sites where water molecules can adsorb. The result is a titanium oxide surface with some fractional coverage of hydroxyl (−OH) groups (titanol).

In the case of Fig. 6.8, the substrate is a thin film of PDMS (10–50 μm thick) cast onto a sheet of poly(ethylene terephthalate) (PET; 175 μm thick). Exposing the PDMS to an oxygen plasma produces surface (−OH) groups (silanol). Placing the plasma-oxidized, Au/Ti-coated stamp on top of these substrates leads to intimate, conformal contact between the raised regions of the stamp and the substrate, without the application of any external pressure. (The soft, conformable PDMS is important in this regard.) It is likely that a dehydration reaction takes place at the (−OH)-bearing interfaces during contact; this reaction results in permanent Ti−O−Si bonds that produce strong adhesion between the two surfaces. Peeling the substrate and stamp apart transfers the Au/Ti bilayer from the raised regions of the stamp (to which the metal has extremely poor adhesion) to the substrate. Complete pattern transfer from an elastomeric stamp to a thin elastomeric substrate occurs readily at room temperature in open air with contact times of less than 15 s. When a rigid stamp is employed, slight heating is needed to induce transfer. While the origin of this difference is unclear, it may reflect the comparatively poor contact when rigid stamps are used; similar differences are also observed in cold welding of gold films [6.40].

Figure 6.9 shows scanning electron micrographs of a pattern produced using a GaAs stamp generated by electron beam lithography and etching. The frames on the right show images of the metal-coated stamp before printing (top) and the transferred pattern (bottom).

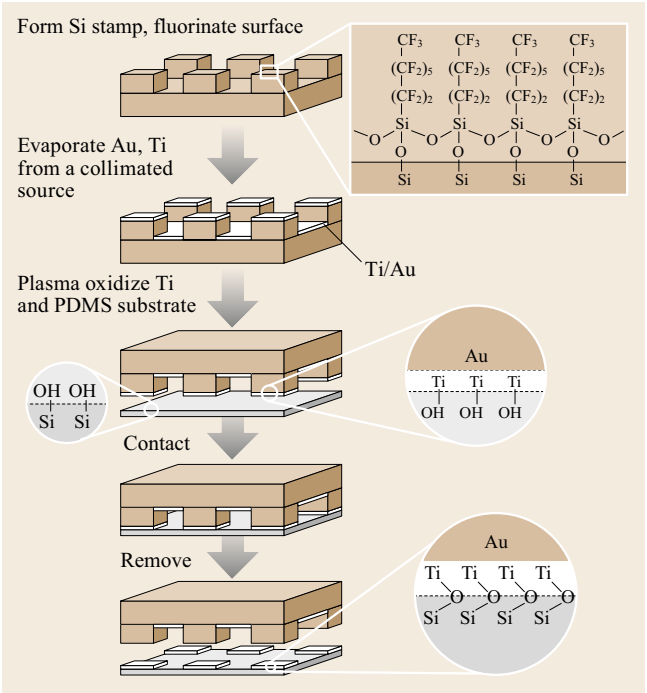


Fig. 6.8 Schematic illustration of nanotransfer printing procedure. Here, interfacial dehydration chemistries control the transfer of a thin metal film from a hard inorganic stamp to a conformable elastomeric substrate (thin film of poly(dimethylsiloxane) (PDMS) on a plastic sheet). The process begins with fabrication of a silicon stamp (by conventional lithography and etching) followed by surface functionalization of the native oxide with a fluorinated silane monolayer. This layer ensures poor adhesion between the stamp and a bilayer metal film (Au and Ti) deposited by electron beam evaporation. A collimated flux of metal oriented perpendicular to the surface of the stamp avoids deposition on the sidewalls of the relief. Exposing the surface Ti layer to an oxygen plasma produces titanol groups. A similar exposure for the PDMS produces silanol groups. Contacting the metal-coated stamp to the PDMS results in a dehydration reaction that links the metal to the PDMS. Removing the stamp leaves a pattern of metal in the geometry of the relief features

The resolution appears to be limited only by the resolution of the stamp itself, and perhaps by the grain size of the metal films. Although the accuracy in multilevel registration that is possible with nTP has not yet been quantified, its performance is likely similar to that of embossing techniques when rigid stamps are used [6.41].

A wide range of surface chemistries can be used for the transfer. SAMs are particularly attractive due to their chemical flexibility. Figure 6.10 illustrates the use of a thiol-terminated SAM and nTP for forming patterns of Au on a silicon wafer [6.38]. Here, the vapor phase

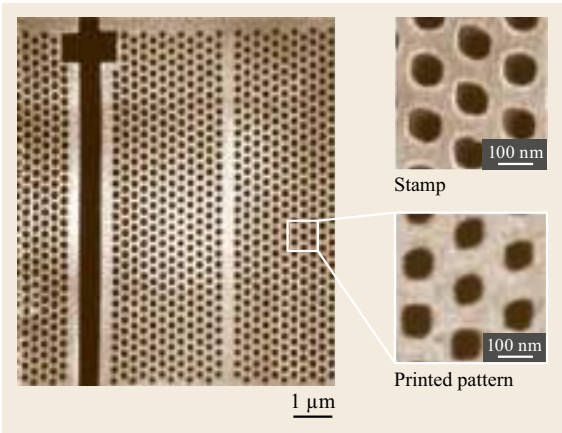


Fig. 6.9 Scanning electron micrograph (SEM) of a pattern produced by nanotransfer printing. The structure consists of a bilayer of Au (20 nm)/Ti (1 nm) (*white*) in the geometry of a photonic bandgap waveguide printed onto a thin layer of poly(dimethylsiloxane) on a sheet of plastic (*black*). Electron beam lithography and etching of a GaAs wafer produced the stamp that was used in this case. The transfer chemistry relied on condensation reactions between titanol groups on the surface of the Ti and silanol groups on the surface of the PDMS. The *frames* on the *right* show SEMs of the Au/Ti-coated stamp (*top*) before printing and on the substrate (*bottom*) after printing. The electron beam lithography and etching used to fabricate the stamp limit the minimum feature size (≈ 70 nm) and the edge resolution (≈ 5–10 nm) of this pattern

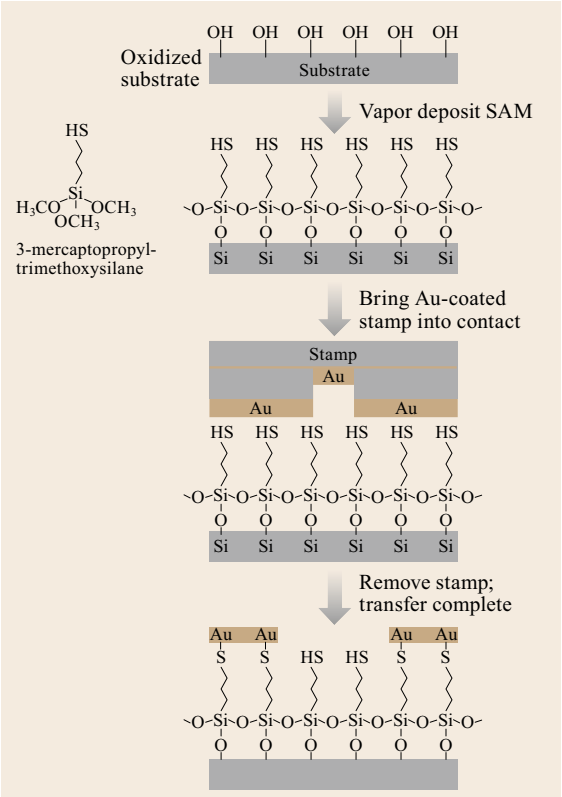


Fig. 6.10 Schematic illustration of steps involved in nanotransfer printing a pattern of a thin layer of Au onto a silicon wafer using a self-assembled monolayer (SAM) surface chemistry. Plasma oxidizing the surface of the wafer generates OH groups. Solution or vapor phase exposure of the wafer to 3-mercaptopropyltrimethoxysilane yields a SAM with exposed thiol groups. Contacting an Au-coated stamp to this surface produces thiol linkages that bond the gold to the substrate. Removing the stamp completes the transfer printing process

cocondensation of the methoxy groups of molecules of 3-mercaptopropyltrimethoxysilane (MPTMS) with the –OH-terminated surface of the wafer produces a SAM of MPTMS with exposed thiol (–SH) groups. PDMS stamps can be prepared for printing on this surface by coating them with a thin film (≈ 15 nm) of Au using conditions (thermal evaporation 1.0 nm/s; ≈ 10^{-7} Torr base pressure) that yield optically smooth, uniform films without the buckling that has been observed in the past with similar systems [6.42]. Nanocracking that sometimes occurs in the films deposited in this way can be reduced or eliminated by evaporating a small amount of Ti onto the PDMS before Au deposition and/or by exposing the PDMS surface briefly to an oxygen plasma. Bringing this coated stamp into contact with the MPTMS SAM leads to the formation of sulfur–gold bonds in the regions of contact. Removing the stamp after a few seconds efficiently transfers the gold from the raised regions of the stamp (Au does not adhere to the PDMS) to the substrate. Covalent bonding of the SAM glue to both the substrate and the gold leads to good adhesion of the printed patterns: They easily pass Scotch tape adhesion tests. Similar results can be obtained with other substrates containing surface –OH groups. For example, Au patterns can be printed onto ≈ 250 µm-thick sheets of poly(ethylene terephthalate) (PET) by first spin-casting and curing (130 °C for 24 h) a thin film of an organosilsesquioxane on the PET. Exposing the cured film (≈ 1 µm thick) to an oxygen plasma and then to air produces the necessary surface (–OH) groups. Figure 6.11 shows some optical micrographs of typical printed patterns in this case [6.38].

Similar surface chemistries can guide transfer to other substrates. Alkanedithiols, for example, are useful for printing Au onto GaAs wafers [6.39]. Immersing these substrates (freshly etched with 37% HCl for

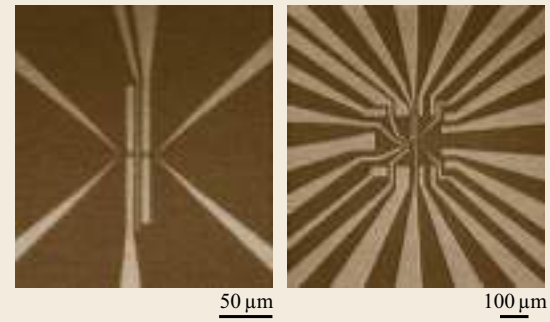


Fig. 6.11 Optical micrographs of patterns of Au (15 nm thick) formed on plastic (*left frame*) and silicon (*right frame*) substrates with nanotransfer printing. The transfer chemistries in both cases rely on self-assembled monolayers with exposed thiol groups. The minimum feature sizes and the edge resolution are both limited by the photolithography used to fabricate the stamps

≈ 2 min to remove the surface oxide) in a 0.05 M solution of 1,8-octanedithiol in ethanol for 3 h produces a monolayer of dithiol on the surface. Although the chemistry of this system is not completely clear, it is generally believed that the thiol end groups bond chemically to the surface. Surface spectroscopy suggests the formation of Ga−S and As−S bonds. Contacting an Au-coated PDMS stamp with the treated substrate causes the exposed thiol endgroups to react with Au in the regions of contact. This reaction produces permanent Au−S bonds at the stamp/substrate interface (see insets in Fig. 6.4 for idealized chemical reaction schemes). Figure 6.12 schematically illustrates how this procedure can generate continuous three-dimensional (3-D) metal patterns using stamps with nonvertical side walls. Several layers can be transferred on top of each other by successively cold-welding the different gold metal layers. Figure 6.13 shows high- and low-magnification scanning electron microscope images of nanotransfer printed single- and multilayer 3-D metal structures [6.43, 44]. The integrity of these free-standing 3-D metal structures is remarkable but depends critically on the careful optimization of the metal evaporation conditions and stamp & substrate surface chemistries.

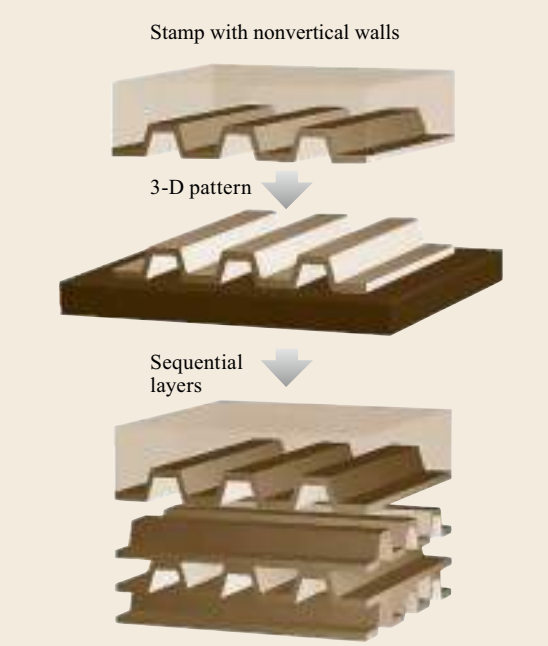


Fig. 6.12 Schematic illustration of the nanotransfer printing (nTP) process for generating continuous 3-D structures when the stamp relief side walls are not vertical. Successive transfer by cold welding the gold films on top of each other yields complex multilayer structures

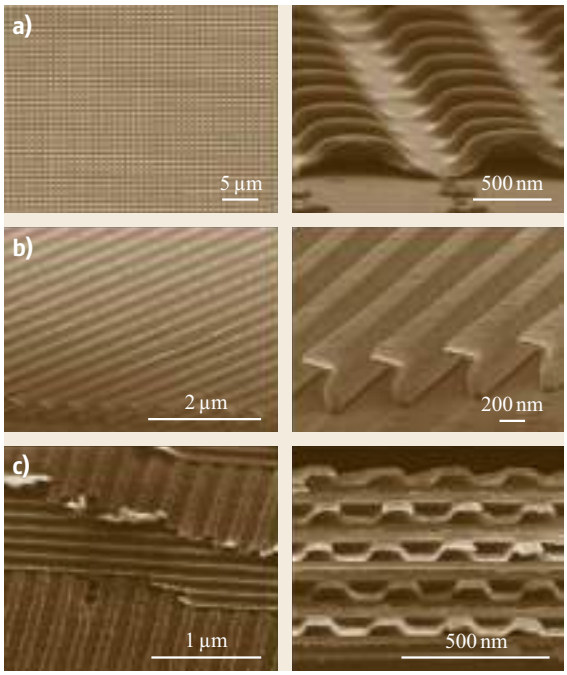


Fig. 6.13a–c Scanning electron micrographs of three-dimensional metal structures obtained by nanotransfer printing gold metal films. Part **(a)** shows closed gold nanocapsules. Part **(b)** shows free-standing L structures obtained using a stamp coated with a steeply angled flux of metal. Part **(c)** shows a multilayer 3-D structure obtained by the successive transfer and cold welding of continuous gold nanocorrugated films

6.4 Applications

Although conventional patterning techniques, such as photolithography or electron beam lithography, have the required resolution, they are not appropriate because they are expensive and generally require multiple processing steps with resists, solvents and developers that can be difficult to use with organic active materials and plastic substrates. Microcontact and nanotransfer printing are both particularly well suited for this application. They can be combined and matched with other techniques, such as ink-jet or screen printing, to form a complete system for patterning all layers in practical plastic electronic devices [6.45]. We have focused our efforts partly on unusual electronic systems such as flexible plastic circuits and devices that rely on electrodes patterned on curved objects such as microcapillaries and optical fibers. We have also explored photonic systems such as distributed feedback structures for lasers and other integrated optical elements that demand submicron features. The sections below highlight several examples in each of these areas.

6.4.1 Unconventional Electronic Systems

A relatively new direction in electronics research seeks to establish low-cost plastic materials, substrates and printing techniques for large-area flexible electronic devices, such as paperlike displays. These types of novel devices can complement those (including high-density memories and high-speed microprocessors) that are well suited to existing inorganic (such as silicon) electronics technologies. High-resolution patterning methods for defining the separation between the source and drain electrodes (the channel length) of transistors in these plastic circuits are particularly important because this dimension determines current output and other important characteristics [6.46].

Figure 6.14 illustrates schematically a cross-sectional view of a typical organic transistor. The frame on the right shows the electrical switching characteristics of a device that uses source/drain electrodes of Au patterned by μCP, a dielectric layer of an organosilsesquioxane, a gate of indium tin oxide (ITO), and a PET substrate. The effective semiconductor mobility extracted from these data is comparable to those measured in devices that use the same semiconductor (pentacene in this case) with inorganic substrates and dielectrics, and gold source/drain electrodes defined by photolithography. Our recent work [6.1, 31, 47] with μCP in the area of plastic electronics demonstrates:

1. Methods for using cylindrical *roller* stamps mounted on fixed axles for printing in a continuous reel-to-reel fashion, high-resolution source/drain electrodes in ultrathin gold and silver deposited from solution at room temperature using electroless deposition.
2. Techniques for performing registration and alignment of the printed features with other elements of a circuit over large areas.
3. Strategies for achieving densities of defects that are as good as those observed with photolithography when the patterning is performed outside of clean room facilities.
4. Methods for removing the printed SAMs to allow good electrical contact of the electrodes with organic semiconductors deposited on top of them.
5. Materials and fabrication sequences that can efficiently exploit these printed electrodes for working organic TFTs in large-scale circuits.

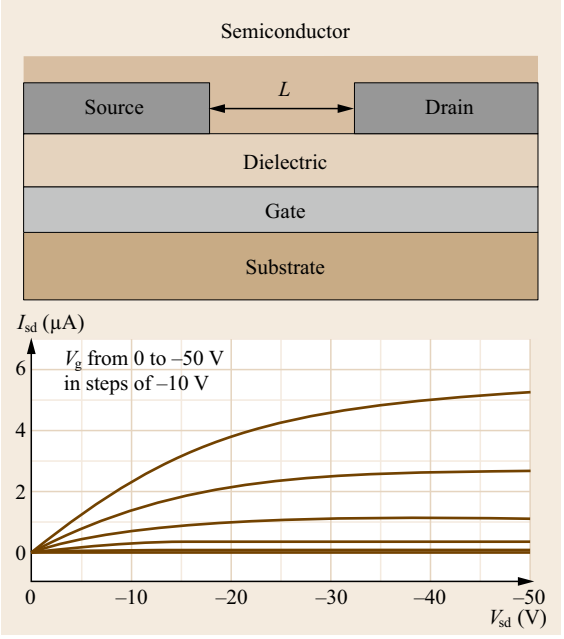


Fig. 6.14 Schematic cross-sectional view (*top*) and electrical performance (*bottom*) of an organic thin film transistor with microcontact printed source and drain electrodes. The structure consists of a substrate (PET), a gate electrode (indium tin oxide), a gate dielectric (spin-cast layer of organosilsesquioxane), source and drain electrodes (20 nm Au and 1.5 nm Ti), and a layer of the organic semiconductor pentacene. The electrical properties of this device are comparable to or better than those that use pentacene with photolithographically defined source/drain electrodes and inorganic dielectrics, gates and substrates

Figure 6.15 provides an image of a large-area plastic circuit with critical features defined by μCP. This circuit is a flexible active matrix backplane for a display. It consists of a square array of interconnected transistors, each of which serves as a switching element that controls the color of a display pixel [6.25, 48]. The transistors themselves have the layout illustrated in Fig. 6.11, and they use similar materials. The semiconductor in this image is blue (pentacene), the source/drain level is Au, the ITO appears green in the optical micrograph in the inset. Part of the circuit rests on the stamp that was used for μCP. The smallest features are the source and drain electrodes (≈ 15 μm lines), the interconnecting lines (≈ 15 μm lines), and the channel length of the transistor (≈ 15 μm). This circuit incorporates five layers of material patterned with good registration of the source/drain, gate, and semiconductor levels. The simple printing approach is illustrated in Fig. 6.16 [6.25]. Just before use, the surface of the stamp is cleaned using a conventional adhesive roller lint remover; this procedure removes dust from the stamp in such a way that does not contaminate or damage its surface. Inking the stamp and placing it face-up on a flat surface prepares it for printing. Matching the cross-hair alignment marks on the corners of one edge of the stamp with those patterned in the ITO brings the substrate into registration with the stamp. During this alignment, features on the stamp are viewed directly through the semitransparent substrate. By bending the PET sheet, contact with the stamp is initiated on the edge of the substrate that contains the cross-hair marks. Gradually unbending the sheet allows contact to progress across the rest of the surface. This printing procedure is attractive because it avoids distortions that can arise when directly manipulating the flexible rubber stamp. It also minimizes the number and size of trapped air pockets

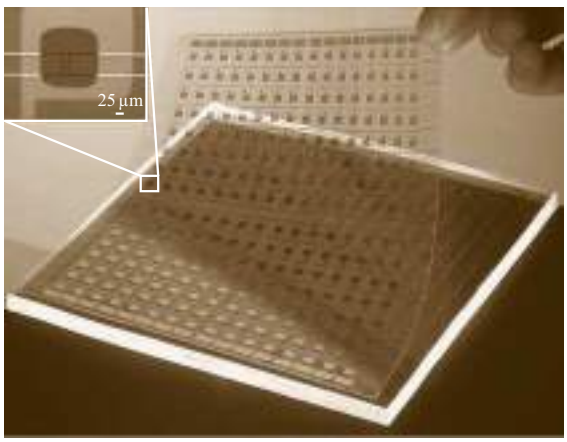


Fig. 6.15 Image of a flexible plastic active matrix backplane circuit whose finest features (transistor source/drain electrodes and related interconnects) are patterned by microcontact printing. The circuit rests partly on the elastomeric stamp that was used for printing. The circuit consists of a square array of interconnected organic transistors, each of which acts locally as a voltage-controlled switch to control the color of an element in the display. The *inset* shows an optical micrograph of one of the transistors

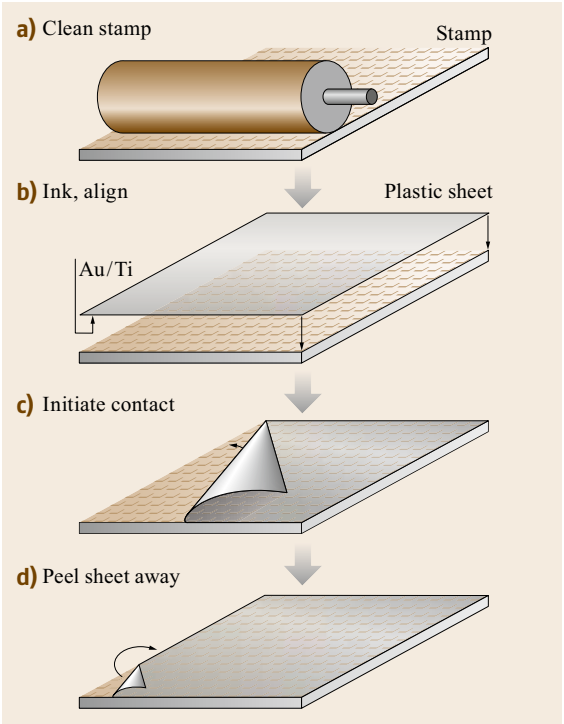


Fig. 6.16a–d Schematic illustration of fabrication steps for microcontact printing over large areas onto plastic sheets. The process begins **(a)** with cleaning the stamp using a conventional adhesive roller lint remover. This procedure effectively removes dust particles. To minimize distortions, the stamp rests face-up on a flat surface and it is not manipulated directly during the printing. Alignment and registration **(b)** are achieved with alignment marks on one side of the substrate and the stamp. By bending the plastic sheet, contact is initiated **(c)** on one side of the stamp; the contact line is then allowed to progress gradually across the stamp. This approach avoids formation of air bubbles that can frustrate good contact. After the substrate is in contact with the stamp for a few seconds, the plastic substrate is separated from the stamp **(d)** by peeling it away beginning in one corner. Good registration (maximum cumulative distortions of less than 50 μm over an area of 130 cm^2) and low defect density can be achieved with this simple approach. It is also well suited for use with rigid composite stamps designed to reduce the level of distortions even further

that can form between the stamp and substrate. Careful measurements performed after etching the unprinted areas of the gold show that over the entire $6'' \times 6''$ area of the circuit, (1) the overall alignment accuracy for positioning the stamp relative to the substrate (the offset of the center of the distribution of registration errors) is $\approx 50-100\,\mu m$, even with the simple approach used here, and (2) the distortion in the positions of features in the source/drain level, when referenced to the gate level, can be as small as $\approx 50\,\mu m$ (the full width at half maximum of the distribution of registration errors). These distortions represent the cumulative effects of deformations in the stamp and distortions in the gate and column electrodes that may arise during the patterning and processing of the flexible PET sheet. The density of defects in the printed patterns is comparable to (or smaller than) that in resist patterned by contact-mode photolithography when both procedures are performed outside of a clean-room facility (when dust is the dominant source of defects).

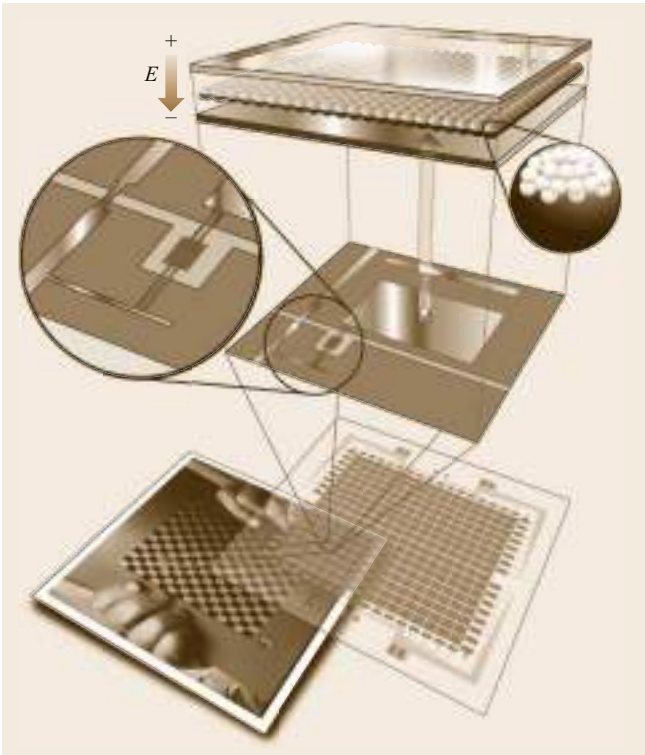


Fig. 6.17 Schematic exploded view of the components of a pixel in an electronic paperlike display (*bottom frame*) that uses a microcontact printed flexible active matrix backplane circuit (illustration near the *bottom frame*). The circuit is laminated against an unpatterned thin sheet of electronic ink (*top frame*) that consists of a monolayer of transparent polymer microcapsules (diameter $\approx 100\,\mu m$). These capsules contain a heavily dyed black fluid and a suspension of charged white pigment particles (see *right inset*). When one of the transistors turns on, electric fields develop between an unpatterned transparent frontplane electrode (indium tin oxide) and a backplane electrode that connects to the transistor. Electrophoretic flow drives the pigment particles to the front or the back of the display, depending on the polarity of the field. This flow changes the color of the pixel, as viewed from the front of the display, from black to white or vica versa

Figure 6.17 shows an *exploded* view of a paperlike display that consists of a printed flexible plastic backplane circuit, like the one illustrated in Fig. 6.16, laminated against a thin layer of *electronic ink* [6.25, 49]. The electronic ink is composed of a monolayer of transparent polymer microcapsules that contain a suspension of charged white pigment particles suspended in a black liquid. The printed transistors in the backplane circuit act as local switches, which control electric fields that drive the pigments to the front or back of the display. When the particles flow to the front of a microcapsule, it appears white; when they flow to the back, it appears black. Figure 6.18 shows a working sheet of active matrix electronic paper that uses this design. This prototype display has several hundred pixels and an optical contrast that is both independent of the viewing angle and significantly better than newsprint. The device is ≈ 1 mm thick, is mechanically flexible, and weighs $\approx 80\%$ less than a conventional liquid crystal display of similar size. Although these displays have only a relatively coarse resolution, all of the processing techniques, the μCP method, the materials, and the electronic inks, are suitable for the large numbers of pixels required for high-information content electronic newspapers and other systems.

Like μCP, nTP is well suited to forming high-resolution source/drain electrodes for plastic electronics. nTP of Au/Ti features in the geometry of the drain and source level of organic transistors, and with appropriate interconnects on a thin layer of PDMS on PET it yields a substrate that can be used in an unusual but powerful way for building circuits: soft, room temperature lamination of such a structure against a plastic substrate that supports the semiconductor, gate dielectric, and gate levels yields a high performance circuit embedded between two plastic sheets [6.37, 50]. (Details of this lamination procedure are presented elsewhere.) The left frame of Fig. 6.19 shows the current–voltage characteristics of a laminated n-channel transistor that uses the organic semiconductor copper hexadecafluorophthalocyanine (n-type) and source/drain electrodes patterned with nTP. The inset shows an optical micrograph of the printed interdigitated source/drain electrodes of this device. The bottom frame of Fig. 6.16 shows the transfer characteristics of a laminated complementary organic inverter circuit whose electrodes



Fig. 6.18 Electronic paperlike display showing two different images. The device consists of several hundred pixels controlled by a flexible active matrix backplane circuit formed by microcontact printing. The relatively coarse resolution of the display is not limited by material properties or by the printing techniques. Instead, it is set by practical considerations for achieving high pixel yields in the relatively uncontrolled environment of the chemistry laboratory in which the circuits were fabricated

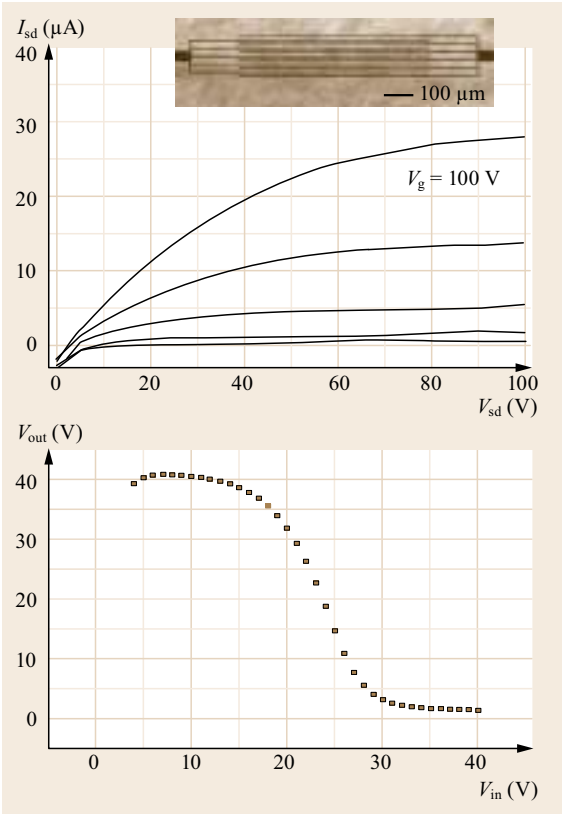


Fig. 6.19 The *upper* frame shows current–voltage characteristics of an n-channel transistor formed with electrodes patterned by nanotransfer printing that are laminated against a substrate that supports an organic semiconductor, a gate dielectric and a gate. The *inset* shows an optical micrograph of the interdigitated electrodes. The *lower* frame shows the transfer characteristics of a simple CMOS inverter circuit that uses this device and a similar one for the p-channel transistor

and connecting lines are defined by nTP. The p-channel transistor in this circuit used pentacene for the semiconductor [6.37].

In addition to high-resolution source/drain electrodes, it is possible to use nTP to form complex multilayer devices with electrical functionality on plastic substrates [6.38]. Figure 6.20 shows a metal/insulator/metal (MIM) structure of Au (50 nm), SiN_x (100 nm; by plasma enhanced vapor deposition, PECVD), Ti (5 nm) and Au (50 nm) formed by transfer printing with a silicon stamp that is coated sequentially with these layers. In this case, a short reactive ion etch (with CF_4) after the second Au deposition removes the SiN_x from the sidewalls of the stamp. nTP transfers these layers in a patterned geometry to a substrate of Au (15 nm)/Ti (1 nm)-coated PDMS (50 μm)/PET (250 μm). Interfacial cold-welding between the Au on the surfaces of the stamp and substrate bonds the multilayers to the substrate. Figure 6.8 illustrates the procedures, the structures (lateral dimensions of $250 \times 250\,\mu m^2$, for ease of electrical probing), and their electrical characteristics. These MIM capacitors have performances similar to devices fabricated on silicon wafers by photolithography and lift-off. This example illustrates the ability of nTP to print patterns of materials whose growth conditions (high-temperature SiN_x by PECVD, in this case) prevent their direct deposition or processing on the substrate of interest (PET, in this case). The cold-welding transfer approach has also been exploited in other ways for patterning components for plastic electronics [6.51, 52].

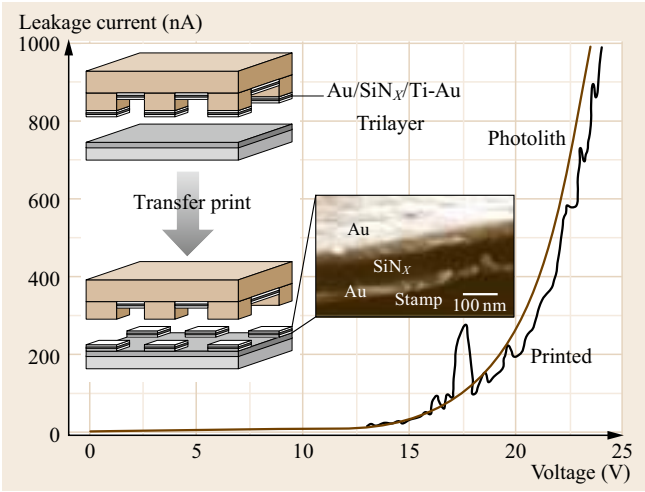


Fig. 6.20 Multilayer thin film capacitor structure printed in a single step onto a plastic substrate using the nanotransfer printing technique. A multilayer of Au/SiN_x/Ti/Au was first deposited onto a silicon stamp formed by photolithography and etching. Contacting this stamp to a substrate of Au/PDMS/PET forms a cold weld that bonds the exposed Au on the stamp to the Au-coating on the substrate. Removing the stamp produces arrays of square ($250 \times 250\,mm^2$) metal/insulator/metal capacitors on the plastic support. The *black line* shows the measured current–voltage characteristics of one of these printed capacitors. The *brown line* corresponds to a similar structure formed on a rigid glass substrate using conventional photolithographic procedures. The characteristics are the same for these two cases. The slightly higher level of noise in the printed devices results, at least partly, from the difficulties involved with making good electrical contacts to structures on the flexible plastic substrate

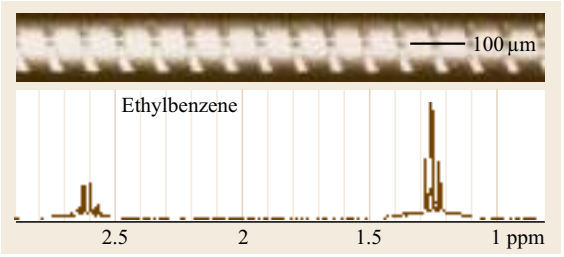


Fig. 6.21 The *top frame* shows an optical micrograph of a continuous conducting microcoil formed by microcontact printing onto a microcapillary tube. This type of printed microcoil is well suited for excitation and detection of nuclear magnetic resonance spectra from nanoliter volumes of fluid housed in the bore of the microcapillary. The *bottom frame* shows a spectra trace collected from an $\approx 8\,nL$ volume of ethyl benzene using a structure similar to the one shown in the *top frame*

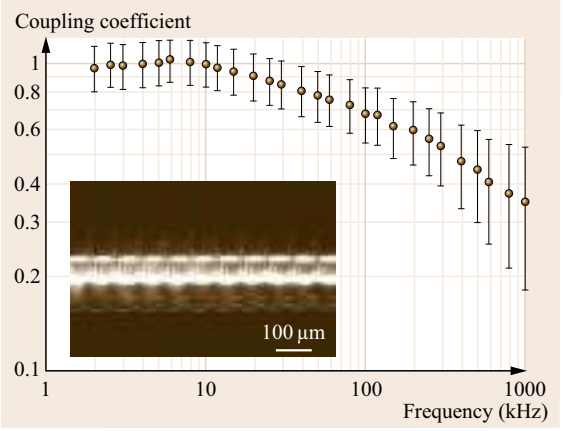


Fig. 6.22 The *inset* shows a concentric microtransformer formed using microcoils printed onto two different microcapillary tubes. The smaller of the tubes (outer diameter 135 µm) has a ferromagnetic wire threaded through its core. The larger one (outer diameter 350 µm) has the smaller tube threaded through its core. The resulting structure is a microtransformer that shows good coupling coefficients at frequencies up to $\approx 1\,MHz$. The *graph* shows its performance

Another class of unusual electronic/optoelectronic devices relies on circuits or circuit elements on curved surfaces. This emerging area of research was stimulated primarily by the ability of µCP to print high-resolution features on fibers and cylinders. Figure 6.21 shows a conducting microcoil printed with µCP on a microcapillary tube using the approach illustrated in Fig. 6.4. The coil serves as the excitation and detection element for high-resolution proton nuclear magnetic resonance of nanoliter volumes of fluid that are housed in the bore of the microcapillary [6.53]. The high fill factor and other considerations lead to extremely high sensitivity with such printed coils. The bottom frame of Fig. 6.21 shows the spectrum of an $\approx 8\,nL$ volume of ethylbenzene. The narrow lines demonstrate the high resolution that is possible with this approach. Similar coils can be used as magnets [6.54], springs [6.36], and electrical transformers [6.55]. Figure 6.22 shows an optical micrograph and the electrical measurements from a concentric cylindrical microtransformer that uses a microcoil printed on a microcapillary tube with a ferromagnetic wire threaded through its core. Inserting this structure into the core of a larger microcapillary that also supports a printed microcoil completes the transformer [6.55]. This type of device shows good coupling coefficients up to relatively high frequencies. Examples of other optoelectronic components appear in fiber optics where microfabricated on-fiber structures serve as integrated photomasks [6.20] and distributed thermal actuators [6.22].

6.4.2 Lasers and Waveguide Structures

In addition to integral components of unconventional electronic systems, useful structures for integrated optics can be built by using μCP and nTP to print sacrificial resist layers for etching glass waveguides. These printing techniques offer the most significant potential value for this area when they are used to pattern features that are smaller than those that can be achieved with contact mode photolithography. Mode coupling gratings and distributed laser resonators are two such classes of structures. We have demonstrated μCP for forming distributed feedback (DFB) and distributed Bragg reflector (DBR) lasers that have narrow emission line widths [6.56]. This challenging fabrication demonstrates the suitability of μCP for building structures that have (1) feature sizes of significantly less than 1 μm (≈ 300 nm), *and* (2) long-range spatial coherence (≈ 1 mm). The lasers employ optically pumped gain material deposited onto DFB or DBR resonators formed from periodic relief on a transparent substrate. The gain media confines light to the surface of the structure; its thickness is chosen to support a single transverse mode. To generate the required relief, lines of gold formed by μCP on a glass slide act as resists for reactive ion etching of the glass. Removing the gold leaves a periodic pattern of relief (600 nm period, 50 nm depth) on the surface of the glass (Fig. 6.23). Figure 6.24 shows the performance of plastic lasers that use printed DFB and DBR resonators with gain media consisting of thin films of PBD doped with 1 wt% of coumarin 490 and DCMII, photopumped with 2 ns pulses from a nitrogen laser with intensities $> 5\,\mathrm{kW/cm^2}$ [6.56]. Multimode lasing at resolution-limited line widths was observed at wavelengths corresponding to the third harmonic of the gratings. These characteristics are similar to those observed in lasers that use resonators generated with photolithography and are better than those that use imprinted polymers [6.57].

Contact printing not only provides a route to low-cost equivalents of gratings fabricated with other approaches, but also allows the fabrication of structures that would be difficult or impossible to generate with photolithography. For example, μCP can be used to form DFB resonators directly on the top surfaces of ridge waveguides [6.58]. Figure 6.25 illustrates the procedures. The bottom left frame shows an optical micrograph of the printed gold lines. Sublimation of a ≈ 200 nm film of tris(8-hydroxyquinoline) aluminum (Al) doped with 0.5–5.0 wt% of the laser dye DCMII onto the resonators produces waveguide DFB lasers. The layer of gain material itself provides a planar waveguide with air and polymer as the cladding layers. The relief waveguide provides lateral confinement of the light. Photopumping these devices with the output of a pulsed nitrogen laser (≈ 2 ns, 337 nm) causes lasing due to Bragg reflections induced by the DFB structures on the top surfaces of the ridge waveguides. Some of the laser emission scatters out of the plane of the waveguide at an angle allowed by phase matching conditions. In this way, the grating also functions as an output coupler and offers a convenient way to characterize the laser emission. The bottom right frame of Fig. 6.22 shows the emission profile.

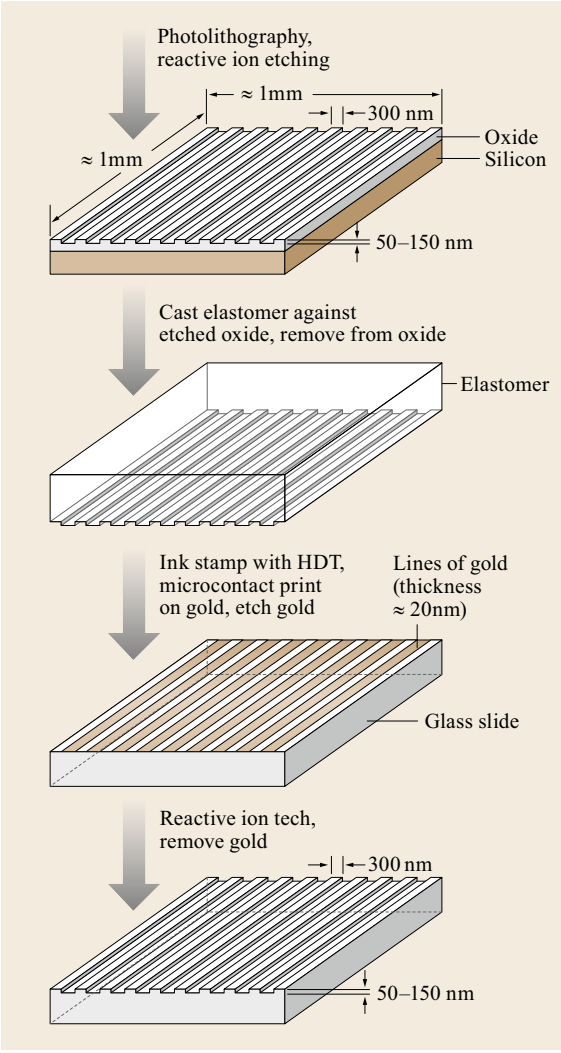
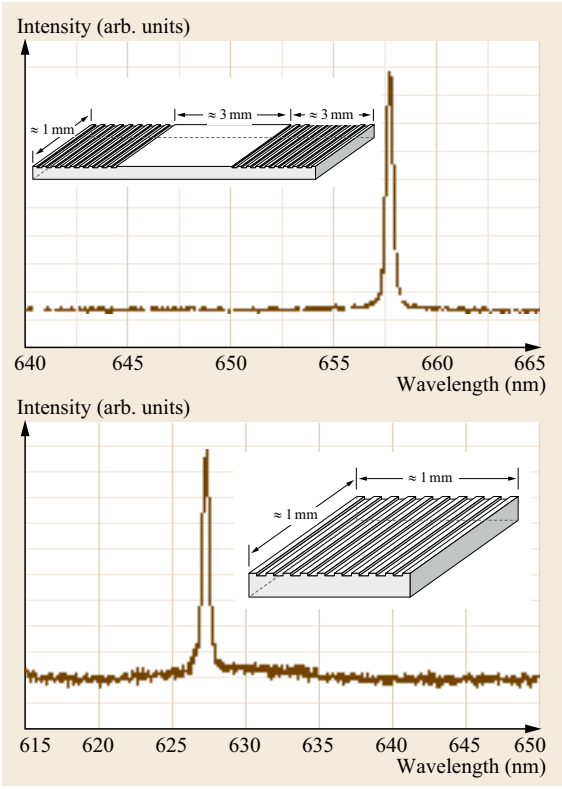


Fig. 6.23 Schematic illustration of the use of microcontact printing (μCP) for fabricating high-resolution gratings that can be incorporated into distributed planar laser structures or other components for integrated optics. The geometries illustrated here are suitable for third-order distributed feedback (DFB) lasers that operate in the red

Fig. 6.24 The *top frames* gives a schematic illustration of steps for microcontact printing high-resolution gratings directly onto the top surfaces of ridge waveguides. The printing defines a sacrificial etch mask of gold which is subsequently removed. Producing this type of structure with photolithography is difficult because of severe thickness nonuniformities that appear in photoresist spin-cast on this type of nonplanar substrate. The *upper bottom* frame shows a top view optical micrograph of printed gold lines on the ridge waveguides. The *lower bottom* frame shows the emission output of a plastic photopumped laser that uses the printed structure and a thin evaporated layer of gain media ►



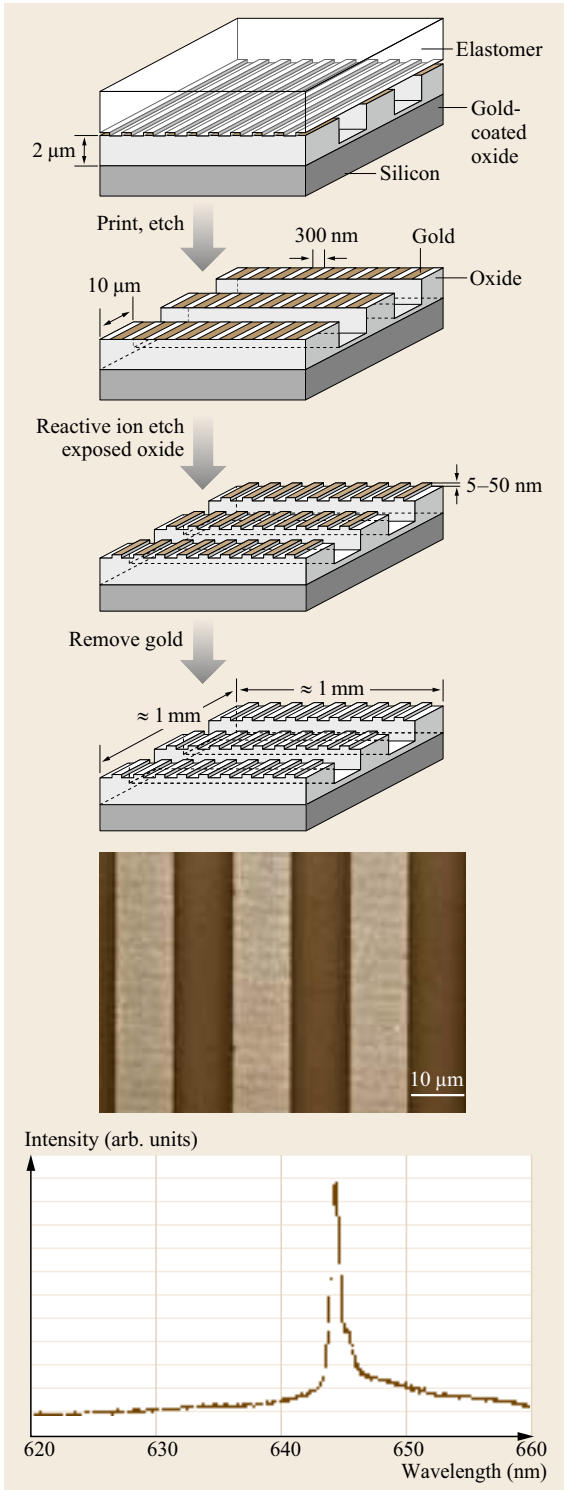


Fig. 6.25 Schematic illustrations and lasing spectra of plastic lasers that use microcontact printed resonators based on surface relief distributed Bragg reflectors (DBRs) and distributed feedback gratings (DFBs) on glass substrates. The grating periods are ≈ 600 nm in both cases. The lasers use thin film plastic gain media deposited onto the printed gratings. This layer forms a planar waveguide that confines the light to the surface of the substrate. The laser shows emission over a narrow wavelength range, with a width that is limited by the resolution of the spectrometer used to characterize the output. In both cases, the emission profiles, the lasing thresholds and other characteristics of the devices are comparable to similar lasers that use resonators formed by high-resolution projection-mode photolithography

6.5 Conclusions

This chapter provides an overview of two contact printing techniques that are capable of micrometer and submicrometer resolution. It also illustrates some applications of these methods that may provide attractive alternatives to more established lithographic methods. The growing interest in nanoscience and technology makes it crucial to develop new methods for fabricating the relevant test structures and devices. The simplicity of these techniques together with the interesting and subtle materials science, chemistry, and physics associated with them make this a promising area for basic and applied study.

Acknowledgments. The authors extend their deepest thanks to all of the collaborators who contributed the work described here.

References

6.1 C.A. Mirkin, J.A. Rogers: Emerging methods for micro- and nanofabrication, MRS Bulletin **26**, 506–507 (2001)

6.2 H.I. Smith, H.G. Craighead: Nanofabrication, Phys. Today **43**, 24–43 (1990)

6.3 W.M. Moreau (Ed.): *Semiconductor Lithography: Principles and Materials* (Plenum, New York 1988)

6.4 S. Matsui, Y. Ochiai: Focused ion beam applications to solid state devices, Nanotechnology **7**, 247–258 (1996)

6.5 J.M. Gibson: Reading and writing with electron beams, Phys. Today **50**, 56–61 (1997)

6.6 L.L. Sohn, R.L. Willett: Fabrication of nanostructures using atomic-force microscope-based lithography, Appl. Phys. Lett. **67**, 1552–1554 (1995)

6.7 E. Betzig, K. Trautman: Near-field optics – Microscopy, spectroscopy, and surface modification beyond the diffraction limit, Science **257**, 189–195 (1992)

6.8 A.J. Bard, G. Denault, C. Lee, D. Mandler, D.O. Wipf: Scanning electrochemical microscopy: A new technique for the characterization and modification of surfaces, Acc. Chem. Res. **23**, 357 (1990)

6.9 J.A. Stroscio, D.M. Eigler: Atomic and molecular manipulation with the scanning tunneling microscope, Science **254**, 1319–1326 (1991)

6.10 J. Nole: Holographic lithography needs no mask, Laser Focus World **33**, 209–212 (1997)

6.11 A.N. Broers, A.C.F. Hoole, J.M. Ryan: Electron beam lithography – Resolution limits, Microelectron. Eng. **32**, 131–142 (1996)

6.12 A.N. Broers, W. Molzen, J. Cuomo, N. Wittels: Electron-beam fabrication of 80 Å metal structures, Appl. Phys. Lett. **29**, 596 (1976)

6.13 G.D. Aumiller, E.A. Chandross, W.J. Tomlinson, H.P. Weber: Submicrometer resolution replication of relief patterns for integrated optics, J. Appl. Phys. **45**, 4557–4562 (1974)

6.14 Y. Xia, J.J. McClelland, R. Gupta, D. Qin, X.-M. Zhao, L.L. Sohn, R.J. Celotta, G.M. Whiteside: Replica molding using polymeric materials: A practical step toward nanomanufacturing, Adv. Mater. **9**, 147–149 (1997)

6.15 T. Borzenko, M. Tormen, G. Schmidt, L.W. Molenkamp, H. Janssen: Polymer bonding process for nanolithography, Appl. Phys. Lett. **79**, 2246–2248 (2001)

6.16 H. Hua, Y. Sun, A. Gaur, M.A. Meitl, L. Bilhaut, L. Rotinka, J. Wang, P. Geil, M. Shim, J.A. Rogers: Polymer imprint lithography with molecular-scale resolution, Nano Lett. **4**(12), 2467–2471 (2004)

6.17 A. Kumar, G.M. Whitesides: Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol ink followed by chemical etching, Appl. Phys. Lett. **63**, 2002–2004 (1993)

6.18 Y. Xia, G.M. Whitesides: Soft lithography, Angew. Chem. Int. Ed. **37**, 550–575 (1998)

6.19 Y. Xia, J.A. Rogers, K.E. Paul, G.M. Whitesides: Unconventional methods for fabricating and patterning nanostructures, Chem. Rev. **99**, 1823–1848 (1999)

6.20 J.A. Rogers, R.J. Jackman, J.L. Wagener, A.M. Vengsarkar, G.M. Whitesides: Using microcontact printing to generate photomasks on the surface of optical fibers: A new method for producing in-fiber gratings, Appl. Phys. Lett. **70**, 7–9 (1997)

6.21 B. Michel, A. Bernard, A. Bietsch, E. Delamarche, M. Geissler, D. Juncker, H. Kind, J.P. Renault, H. Rothuizen, H. Schmid, P. Schmidt-Winkel, R. Stutz, H. Wolf: Printing meets lithography: soft approaches to high-resolution printing, IBM J. Res. Dev. **45**, 697–719 (2001)

6.22 J.A. Rogers: Rubber stamping for plastic electronics and fiber optics, MRS Bulletin **26**, 530–534 (2001)

6.23 N.B. Larsen, H. Biebuyck, E. Delamarche, B. Michel: Order in microcontact printed self-assembled monolayers, J. Am. Chem. Soc. **119**, 3017–3026 (1997)

6.24 H.A. Biebuyck, G.M. Whitesides: Self-organization of organic liquids on patterned self-assembled monolayers of alkanethiolates on gold, Langmuir **10**, 2790–2793 (1994)

6.25 J.A. Rogers, Z. Bao, K. Baldwin, A. Dodabalapur, B. Crone, V.R. Raju, V. Kuck, H. Katz, K. Amundson, J. Ewing, P. Drzaic: Paper-like electronic displays: Large area, rubber stamped plastic sheets of electronics and electrophoretic inks, Proc. Natl. Acad. Sci. USA **98**, 4835–4840 (2001)

6.26 J.L. Wilbur, H.A. Biebuyck, J.C. MacDonald, G.M. Whitesides: Scanning force microscopies can image patterned self-assembled monolayers, Langmuir **11**, 825–831 (1995)

6.27 J.C. Love, D.B. Wolfe, M.L. Chabinyc, K.E. Paul, G.M. Whitesides: Self-assembled monolayers of alkanethiolates on palladium are good etch resists, J. Am. Chem. Soc. **124**, 1576–1577 (2002)

6.28 H. Schmid, B. Michel: Siloxane polymers for high-resolution, high-accuracy soft lithography, Macromolecules **33**, 3042–3049 (2000)

6.29 K. Choi, J.A. Rogers: A photocurable poly(dimethylsiloxane) chemistry for soft lithography in the nanometer regime, J. Am. Chem. Soc. **125**, 4060–4061 (2003)

6.30 J.A. Rogers, K.E. Paul, G.M. Whitesides: Quantifying distortions in soft lithography, J. Vac. Sci. Technol. B **16**, 88–97 (1998)

6.31 J. Tate, J.A. Rogers, C.D.W. Jones, W. Li, Z. Bao, D.W. Murphy, R.E. Slusher, A. Dodabalapur, H.E. Katz, A.J. Lovinger: Anodization and microcontact printing on electroless silver: Solution-based fabrication procedures for low voltage organic electronic systems, Langmuir **16**, 6054–6060 (2000)

6.32 Y. Xia, E. Kim, G.M. Whitesides: Microcontact printing of alkanethiols on silver and its application to microfabrication, J. Electrochem. Soc. **143**, 1070–1079 (1996)

6.33 Y.N. Xia, X.M. Zhao, E. Kim, G.M. Whitesides: A selective etching solution for use with patterned self-assembled monolayers of alkanethiolates on gold, Chem. Mater. **7**, 2332–2337 (1995)

6.34 R.J. Jackman, J. Wilbur, G.M. Whitesides: Fabrication of submicrometer features on curved substrates by microcontact printing, Science **269**, 664–666 (1995)

6.35 R.J. Jackman, S.T. Brittain, A. Adams, M.G. Prentiss, G.M. Whitesides: Design and fabrication of topologically complex, three-dimensional microstructures, Science **280**, 2089–2091 (1998)

6.36 J.A. Rogers, R.J. Jackman, G.M. Whitesides: Microcontact printing and electroplating on curved substrates: A new means for producing free-standing three-dimensional microstructures with possible applications ranging from micro-coil springs to coronary stents, Adv. Mater. **9**, 475–477 (1997)

6.37 Y.-L. Loo, R.W. Willett, K. Baldwin, J.A. Rogers: Additive, nanoscale patterning of metal films with a stamp and a surface chemistry mediated transfer process: applications in plastic electronics, Appl. Phys. Lett. **81**, 562–564 (2002)

6.38 Y.-L. Loo, R.W. Willett, K. Baldwin, J.A. Rogers: Interfacial chemistries for nanoscale transfer printing, J. Am. Chem. Soc. **124**, 7654–7655 (2002)

6.39 Y.-L. Loo, J.W.P. Hsu, R.L. Willett, K.W. Baldwin, K.W. West, J.A. Rogers: High-resolution transfer printing on GaAs surfaces using alkane dithiol self-assembled monolayers, J. Vac. Sci. Technol. B **20**, 2853–2856 (2002)

6.40 G.S. Ferguson, M.K. Chaudhury, G.B. Sigal, G.M. Whitesides: Contact adhesion of thin gold-films on elastomeric supports – cold welding under ambient conditions, Science **253**, 776–778 (1991)

6.41 W. Zhang, S.Y. Chou: Multilevel nanoimprint lithography with submicron alignment over 4 in Si wafers, Appl. Phys. Lett. **79**, 845–847 (2001)

6.42 N. Bowden, S. Brittain, A.G. Evans, J.W. Hutchinson, G.M. Whitesides: Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer, Nature **393**, 146–149 (1998)

6.43 E. Menard, L. Bilhaut, J. Zaumseil, J.A. Rogers: Improved surface chemistries, thin film deposition techniques, and stamp designs for nanotransfer printing, Langmuir **20**, 6871–6878 (2004)

6.44 J. Zaumseil, M.A. Meitl, J.W.P. Hsu, B. Acharya, K.W. Baldwin, Y.-L. Loo, J.A. Rogers: Three-dimensional and multilayer nanostructures formed by nanotransfer printing, Nano Lett. **3**, 1223–1227 (2003)

6.45 Z. Bao, J.A. Rogers, H.E. Katz: Printable organic and polymeric semiconducting materials and devices, J. Mater. Chem. **9**, 1895–1904 (1999)

6.46 J.A. Rogers, Z. Bao, A. Dodabalapur, A. Makhija: Organic smart pixels and complementary inverter circuits formed on plastic substrates by casting, printing and molding, IEEE Electron Dev. Lett. **21**, 100–103 (2000)

6.47 J.A. Rogers, Z. Bao, A. Makhija: Non-photolithographic fabrication sequence suitable for reel-to-reel production of high performance organic transistors and circuits that incorporate them, Adv. Mater. **11**, 741–745 (1999)

6.48 P. Mach, S. Rodriguez, R. Nortrup, P. Wiltzius, J.A. Rogers: Active matrix displays that use printed organic transistors and polymer dispersed liquid crystals on flexible substrates, Appl. Phys. Lett. **78**, 3592–3594 (2001)

6.49 J.A. Rogers: Toward paperlike displays, Science **291**, 1502–1503 (2001)

6.50 Y.-L. Loo, T. Someya, K.W. Baldwin, P. Ho, Z. Bao, A. Dodabalapur, H.E. Katz, J.A. Rogers: Soft, conformable electrical contacts for organic transistors: High resolution circuits by lamination, Proc. Natl. Acad. Sci. USA **99**, 10252–10256 (2002)

6.51 C. Kim, P.E. Burrows, S.R. Forrest: Micropatterning of organic electronic devices by cold-welding, Science **288**, 831–833 (2000)

6.52 C. Kim, M. Shtein, S.R. Forrest: Nanolithography based on patterned metal transfer and its application to organic electronic devices, Appl. Phys. Lett. **80**, 4051–4053 (2002)

6.53 J.A. Rogers, R.J. Jackman, G.M. Whitesides, D.L. Olson, J.V. Sweedler: Using microcontact printing to fabricate microcoils on capillaries for high resolution ^{1}H-NMR on nanoliter volumes, Appl. Phys. Lett. **70**, 2464–2466 (1997)

6.54 J.A. Rogers, R.J. Jackman, G.M. Whitesides: Constructing single and multiple helical microcoils and characterizing their performance as components of microinductors and microelectromagnets, J. Microelectromech. Syst. **6**, 184–192 (1997)

6.55 R.J. Jackman, J.A. Rogers, G.M. Whitesides: Fabrication and characterization of a concentric, cylindrical microtransformer, IEEE Trans. Magn. **33**, 2501–2503 (1997)

6.56 J.A. Rogers, M. Meier, A. Dodabalapur: Using stamping and molding techniques to produce distributed feedback and Bragg reflector resonators for plastic lasers, Appl. Phys. Lett. **73**, 1766–1768 (1998)

6.57 M. Berggren, A. Dodabalapur, R.E. Slusher, A. Timko, O. Nalamasu: Organic solid-state lasers with imprinted gratings on plastic substrates, Appl. Phys. Lett. **72**, 410–411 (1998)

6.58 J.A. Rogers, M. Meier, A. Dodabalapur: Distributed feedback ridge waveguide lasers fabricated by nanoscale printing and molding on non-planar substrates, Appl. Phys. Lett. **74**, 3257–3259 (1999)

7. Materials Aspects of Micro- and Nanoelectromechanical Systems

Christian A. Zorman

One of the more significant technological achievements during the last twenty years has been the development of the field of microelectromechanical systems (MEMS) and its offshoot, nanoelectromechanical systems (NEMS). These developments were made possible by significant advancements in the materials and processing technologies used in the fabrication of MEMS and NEMS devices. While initial developments capitalized on a mature Si infrastructure built for the integrated circuit (IC) industry, recent advances have come about using materials and processes not typically associated with IC fabrication, a trend that is likely to continue as new application areas emerge.

A well-rounded understanding of MEMS and NEMS technology requires a basic knowledge of the materials used to construct the devices, since material properties often govern device performance and dictate fabrication approaches. An understanding of the materials used in MEMS and NEMS involves an understanding of material systems, since such devices are rarely constructed of a single material, but rather a collection of materials working in conjunction with each other to provide critical functions. It is from this perspective that the following chapter is constructed. This chapter is not a summary of all materials used in MEMS and NEMS, as such a work would itself constitute a single text of significant size. It does, however, present a selection of some of the more popular materials, as well as those that illustrate the importance of viewing MEMS and NEMS in terms of material systems.

7.1 **Silicon and its Commonly-Used Derivatives** 163
7.1.1 Single-Crystal Silicon 164
7.1.2 Polycrystalline and Amorphous Silicon 165
7.1.3 Porous Silicon 168
7.1.4 Silicon Dioxide 168
7.1.5 Silicon Nitride 169
7.2 **Germanium-Based Materials** 170
7.2.1 Polycrystalline Germanium 170
7.2.2 Polycrystalline Silicon Germanium 171
7.3 **Metals** 171
7.4 **Semiconductors for Harsh Environment Applications** 173
7.4.1 Silicon Carbide 173
7.4.2 Diamond 176
7.5 **GaAs, InP and Related III-V Materials** 178
7.6 **Piezoelectric Materials** 180
7.6.1 Lead Zirconate Titanate 180
7.6.2 Aluminum Nitride 181
7.6.3 Gallium Nitride 182
7.7 **Polymer Materials** 183
7.7.1 Polyimide 183
7.7.2 SU-8 183
7.7.3 Parylene 184
7.7.4 Liquid Crystal Polymer 184
7.8 **Future Trends** 185
References 185

7.1 Silicon and its Commonly-Used Derivatives

The use of silicon (Si) as a material for microfabricated sensors dates back to the middle of the last century when the piezoresistive effect in germanium (Ge) and Si was first identified [7.1]. It was discovered that the piezoresistive coefficients of Si were significantly higher than those associated with metals used in conventional strain gages; this finding initiated the development of Si-based strain gage devices, and, along

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_7

with Si bulk micromachining techniques, piezoresistive Si pressure sensors during the 1960s and 1970s. The subsequent development of Si surface micromachining techniques along with the recognition that micromachined Si structures could be integrated with Si IC devices marked the advent of MEMS with Si firmly positioned as the primary MEMS material.

7.1.1 Single-Crystal Silicon

For MEMS applications, single-crystal Si serves several key functions. Single-crystal Si is one of the most versatile materials for bulk micromachining, due to the availability of anisotropic etching processes in conjunction with good mechanical properties. Single-crystal Si has favorable mechanical properties (i. e., a Young's modulus of about 190 GPa), enabling its use as a material for membranes, bridges, cantilevers and other popular micromachined structures. For surface micromachining applications, single-crystal Si substrates are used primarily as mechanical platforms on which device structures are fabricated, although the advent of silicon-on-insulator (SOI) substrates enables the fabrication of single-crystal Si surface-micromachined structures by using the buried oxide as a sacrificial layer. The use of high-quality single-crystal wafers enables the fabrication of integrated MEMS devices, at least for materials and processes that are compatible with Si ICs.

From the materials perspective, single-crystal Si is a relatively easy material to bulk micromachine, due to the availability of anisotropic etchants such as potassium hydroxide (KOH) and tetramethyl-aluminum hydroxide (TMAH) that attack the (100) and (110) Si crystal planes significantly faster than the (111) crystal planes. For example, the etching rate ratio of (100) to (111) planes in Si is about 400:1 for a typical KOH/water solution. Silicon dioxide (SiO_2), silicon nitride (Si_3N_4), and some metallic thin films (e.g., Cr, Au, etc.) provide good etch masks for most Si anisotropic etchants. Heavily boron-doped Si is an effective etch stop for some liquid reagents. Boron-doped etch stops are often less than 10 μm thick, since the boron concentration in Si must exceed 7×10^{19} cm^3 for the etch stop to be effective and the doping is done by thermal

Table 7.1 Distinguishing characteristics and application examples of selected materials for MEMS and NEMS

Material	Distinguishing Characteristics	Application Examples
Single-crystal silicon (Si)	High quality electronic material, selective anisotropic etching	Bulk micromachining, piezoresistive sensing
Polycrystalline Si (polysilicon)	Doped Si films on sacrificial layers	Surface micromachining, electrostatic actuation
Silicon dioxide (SiO_2)	Insulating, etched by HF, compatible with polysilicon	Sacrificial layer in polysilicon surface micromachining, passivation layer for devices
Silicon nitride (Si_3N_4, Si_xN_y)	Insulating, chemically resistant, mechanically durable	Isolation layer for electrostatic devices, membrane and bridge material
Polycrystalline germanium (polyGe), Polycrystalline silicon-germanium (poly-Si-Ge)	Deposited at low temperatures	Integrated surface micromachined MEMS
Gold (Au), aluminum (Al)	Conductive thin films, flexible deposition techniques	Interconnect layers, masking layers, electromechanical switches
Nickel-iron (NiFe)	Magnetic alloy	Magnetic actuation
Titanium-nickel (TiNi)	Shape-memory alloy	Thermal actuation
Silicon carbide (SiC) Diamond	Electrically and mechanically stable at high temperatures, chemically inert, high Young's modulus to density ratio	Harsh-environment MEMS, high-frequency MEMS/NEMS
Gallium arsenide (GaAs), indium phosphide (InP), indium arsenide (InAs) and related materials	Wide bandgap, epitaxial growth on related ternary compounds	Radio-frequency MEMS, optoelectronic devices, single-crystal bulk and surface micromachining
Lead zirconate titanate (PZT), aluminum nitride, gallium nitride	Piezoelectric materials	Mechanical sensors and actuators
Polyimide	Chemically resistant, high-temperature polymer	Mechanically flexible MEMS, bioMEMS
SU-8	Thick, photodefinable resist	Micromolding, High-aspect-ratio structures.
Parylene	Biocompatible polymer, deposited at room temperature by CVD	Protective coatings, molded polymer structures

diffusion. Ion implantation can be used to create a subsurface etch stop layer; however, the practical limit is a few microns.

In contrast to anisotropic etching, isotropic etching exhibits no selectivity to the various crystal planes. Commonly used isotropic Si etchants consist of hydrofluoric (HF) and nitric (HNO_3) acid mixtures in water or acetic acid (CH_3COOH), with the etch rate dependent on the ratio of HF to HNO_3. From a processing perspective, isotropic etching of Si is commonly used for removal of work-damaged surfaces, creation of structures in single-crystal slices, and patterning of single-crystal or polycrystalline films.

Well-established dry etching processes are routinely used to pattern single-crystal Si. The process spectrum ranges from physical techniques such as sputtering and ion milling to chemical techniques such as plasma etching. Reactive ion etching (RIE) is the most commonly used dry etching technique for Si patterning. By combining both physical and chemical processes, RIE is a highly effective anisotropic Si etching technique that can be used to generate patterns that are independent of crystalline orientation. Fluorinated compounds such as CF_4, SF_6, and NF_3, or chlorinated compounds such as CCl_4 or Cl_2, sometimes mixed with He, O_2 or H_2, are commonly used in Si RIE. The RIE process is highly directional, which enables direct lateral pattern transfer from an overlying masking material to the etched Si surface. SiO_2 thin films are often used as masking and sacrificial layers, owing to the material's chemical durability under these plasma conditions. Process limitations (i. e., etch rates) restrict the etch depths of conventional Si RIE to less than 10 μm, however, a process called deep reactive ion etching (DRIE) has extended the use of anisotropic dry etching to depths well beyond several hundred microns.

Using the aforementioned processes and techniques, a wide variety of microfabricated devices have been made from single-crystal Si, such as piezoresistive pressure sensors, accelerometers, and mechanical resonators to name a few. Using nearly the same approaches but on a smaller scale, *top-down* nanomachining techniques have been used to fabricate nanoelectromechanical devices from single-crystal Si. Single-crystal Si is particularly well suited for nanofabrication because high crystal quality substrates with very smooth surfaces are readily available. By coupling electron-beam (e-beam) lithographic techniques with conventional Si etching, device structures with submicron dimensions have been fabricated. Submicron, single-crystal Si nanomechanical structures have been successfully micromachined from bulk Si wafers [7.2] and silicon-on-insulator (SOI) wafers [7.3]. In the former case, an isotropic Si etch was performed to release the device structures, whereas in the latter case, the 50–200 nm structures were released by dissolving the underlying oxide layer in HF. An example of nanoelectromechanical beam structures fabricated from a single-crystal Si substrate is shown in Fig. 7.1.

7.1.2 Polycrystalline and Amorphous Silicon

Surface micromachining is a process where a sequence of thin films, often of different materials, is deposited and selectively etched to form the desired micromechanical (or microelectromechanical) structure. In contrast to bulk micromachining, the substrate serves primarily as a device-supporting platform. For Si-based surface micromachined MEMS, polycrystalline Si (polysilicon) is most often used as the structural material, SiO_2 as the sacrificial material, silicon nitride (Si_3N_4) for electrical isolation of device structures, and single-crystal Si as the substrate. Like single-crystal Si, polysilicon can be doped during or after film deposition. SiO_2 can be thermally grown or deposited on polysilicon over a broad temperature range (e.g., 200–1150 °C) to meet various process and material requirements. SiO_2 is readily dissolvable in hydrofluoric acid (HF), which does not etch polysilicon and thus can be used to dissolve SiO_2 sacrificial layers. Si_3N_4 is an insulating film that is highly resistant to oxide etchants. The polysilicon micromotor shown in Fig. 7.2 was surface micromachined using a process that included these materials.

For MEMS and IC applications, polysilicon films are commonly deposited using a process known as low-pressure chemical vapor deposition (LPCVD). The



Fig. 7.1 A collection of Si nanoelectromechanical beam resonators fabricated from a single-crystal Si substrate (courtesy M. Roukes, Caltech)



Fig. 7.2 Scanning electron microscope (SEM) micrograph of a surface micromachined polysilicon micromotor fabricated using a SiO_2 sacrificial layer

typical polysilicon LPCVD reactor is based on a hot-wall, resistance-heated furnace. Typical processes are performed at temperatures ranging from 580 to 650 °C and pressures from 100 to 400 mtorr. The most commonly used source gas is silane (SiH_4). The microstructure of polysilicon thin films consists of a collection of small grains whose microstructure and orientation is a function of the deposition conditions [7.4]. For typical LPCVD processes (e.g., 200 mtorr), the amorphous-to-polycrystalline transition temperature is about 570 °C, with polycrystalline films deposited above the transition temperature. At 600 °C, the grains are small and equiaxed, while at 625 °C, the grains are large and columnar [7.4]. The crystal orientation is predominantly (110) Si for temperatures between 600–650 °C, while the (100) orientation is dominant for temperatures between 650–700 °C.

The resistivity of polysilicon can be modified using the doping methods developed for single-crystal Si. Diffusion is an effective method for doping polysilicon films, especially for heavy doping of thick films. Phosphorous, which is the most commonly used dopant in polysilicon MEMS, diffuses significantly faster in polysilicon than in single-crystal Si due primarily to enhanced diffusion rates along grain boundaries. The diffusivity of phosphorus in polysilicon thin films with small equiaxed grains is about $1 \times 10^{12}\,cm^2/s$. Ion implantation is also used to dope polysilicon films. A high-temperature annealing step is usually required to electrically activate the implanted dopants as well as to repair implant-related damage in the polysilicon films. In general, the conductivity of implanted polysilicon films is not as high as films doped by diffusion.

In-situ doping of polysilicon is performed by simply including a dopant gas, usually diborane (B_2H_6) or phosphine (PH_3), in the CVD process. The addition of dopants during the deposition process not only modifies the conductivity, but also affects the deposition rate of the polysilicon films. As shown in Fig. 7.3, the inclusion of boron generally increases the deposition rate of polysilicon relative to undoped films [7.5], while phosphorus (not shown) reduces the rate. In situ doping can be used to produce conductive films with uniform doping profiles without requiring the high-temperature steps commonly associated with diffusion or ion implantation. Although commonly used to produce doped polysilicon for electrostatic devices, *Cao* et al. [7.6] have used in situ phosphorus doped polysilicon films in piezoresistive strain gages, achieving gage factors as high as 15 for a single strip sensor.

The thermal conductivity of polysilicon is a strong function of its microstructure, and therefore the conditions used during deposition [7.4]. For fine-grain films, the thermal conductivity is about 25% of the value of single-crystal Si. For thick films with large grains, the thermal conductivity ranges from 50–85% of the single-crystal value.

Like the electrical and thermal properties of polysilicon, the as-deposited residual stress in polysilicon films depends on microstructure. For films deposited under typical conditions (200 mtorr, 625 °C), the as-deposited polysilicon films have compressive residual stresses. The highest compressive stresses are found in amorphous Si films and polysilicon films with a strong, columnar (110) texture. For films with fine-grained microstructures, the stress tends to be tensile. Annealing can be used to reduce the compressive stress in as-deposited polysilicon films. For instance, compressive residual stresses on the order of 500 MPa can be reduced to less than 10 MPa by annealing the as-deposited films at 1000 °C in ambient N_2 [7.7, 8]. Rapid thermal annealing (RTA) provides an effective method of stress reduction in polysilicon films on temperature-sensitive substrates. *Zhang* et al. [7.9] reported that a 10 s anneal at 1100 °C was sufficient to completely relieve the stress in films that originally had a compressive stress of about 340 MPa. RTA is particularly attractive in situations where the process parameters require a low thermal budget.

As an alternative to high temperature annealing, *Yang* et al. [7.10] have developed an approach that actually utilizes the residual stress characteristics of polysilicon deposited under various conditions to construct polysilicon multilayers that have the desired thickness and stress values. The multilayers are composed of alternating tensile and compressive polysilicon layers that are deposited in a sequential manner. The tensile layers consist of fine-grained polysilicon grown at a temperature of 570 °C, while the compressive layers are made

up of columnar polysilicon deposited at 615 °C. The overall stress in the composite film depends on the number of alternating layers and the thickness of each layer. With the proper set of parameters, a composite polysilicon multilayer can be deposited with near zero residual stress and no stress gradient. The process achieves stress reduction without high-temperature annealing; a considerable advantage for integrated MEMS processes.

Along similar lines, *Kang* et al. [7.11] reported a method that uses nickel silicides (Ni_xSi_y) to compensate for residual stress gradients in polysilicon thin films. Silicides are used in complementary metal-oxide-semiconductor (CMOS) technology to reduce sheet and contact resistances in polysilicon gates and thus would be available in an integrated polysilicon MEMS process. Ni_xSi_y films were formed on polysilicon surfaces by first depositing a Ni film by thermal evaporation, followed by a low-temperature (< 400 °C) thermal anneal. The researchers found that the sheet resistance of the resulting film can be reduced from 20 000 to 10 Ω/□ and the stress gradient, as determined by vertical deflection of surface micromachined, 2 μm-thick micromachined cantilevers, can be completely relaxed at an annealing temperature of 290 °C. The cantilevers were fabricated using a tetraethoxysilane-based silicon dioxide sacrificial layer. Release of the beams required that a direct current (DC) bias be placed on the Ni_xSi_y layer to suppress electrochemical corrosion of the layer in the HF release solution.

Many device designs require polysilicon thicknesses that are not readily achievable using conventional LPCVD polysilicon due to the low deposition rates associated with such systems. For these applications, epitaxial Si reactors can be used to grow polysilicon films. Unlike conventional LPCVD processes, which typically have deposition rates of less than 10 nm/min, epitaxial processes have deposition rates on the order of 1 μm/min [7.12]. The high deposition rates result from the much higher substrate temperatures (> 1000 °C) and deposition pressures (> 50 torr) used in these processes. The polysilicon films are usually deposited on SiO_2 sacrificial layers to enable surface micromachining. An LPCVD polysilicon seed layer is sometimes used in order to control nucleation, grain size, and surface roughness. As with conventional polysilicon, the microstructure and residual stress of the epipoly films, as they are known, is related to deposition conditions. Compressive films generally have a mixture of [110] and [311] grains [7.13, 14] while tensile films have a random mix of [110], [100], [111], and [311] grains [7.13]. The Young's modulus of epipoly measured from micromachined test structures is comparable with LPCVD polysilicon [7.14]. Mechanical

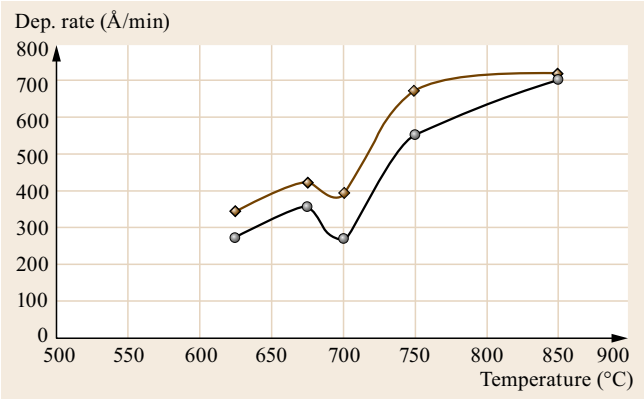


Fig. 7.3 Deposition rate versus substrate temperature for in situ boron doped (◇) and undoped (●) polysilicon films grown by atmospheric pressure chemical vapor deposition (after [7.5])

test structures [7.12–14], thermal actuators [7.12], electrostatically actuated accelerometers [7.12] and gyroscopes [7.15] have been fabricated from these films.

As a low-temperature alternative to LPCVD polysilicon, physical vapor deposition (PVD) techniques have been developed to produce Si thin films on temperature-sensitive substrates. *Abe* et al. [7.16] and *Honer* et al. [7.17] have developed sputtering processes for polysilicon. Early work [7.16] emphasized the ability to deposit very smooth (2.5 nm) polysilicon films on thermally oxidized wafers at reasonable deposition rates (19 nm/min) and with low residual compressive stresses. The process involved DC magnetron sputtering from a Si target using an Ar sputtering gas, a chamber pressure of 5 mtorr, and a power of 100 W. The authors reported that a postdeposition anneal at 700 °C in N_2 for 2 h was needed to crystallize the deposited film and perhaps lower the stress. *Honer* et al. [7.17] sought to develop a polymer-friendly, Si-based surface micromachining process based on polysilicon sputtered onto polyimide and phosphosilicate glass (PSG) sacrificial layers. To improve the conductivity of the micromachined Si structures, the sputtered Si films were sandwiched between two TiW cladding layers. The device structures on polyimide were released using oxygen plasma etching. The processing step with the highest temperature was in fact the polyimide cure at 350 °C. To test the robustness of the process, sputter-deposited Si microstructures were fabricated on substrates containing CMOS devices. As expected from thermal budget considerations, the authors reported no measurable degradation of device performance.

Plasma-enhanced chemical vapor deposition (PECVD) has emerged as an alternative to LPCVD for the production of Si-based surface micromachined structures on temperature-sensitive substrates. *Gaspar*

et al. [7.18], reported on the development of surface micromachined microresonators fabricated from hydrogenated amorphous Si (a-Si:H) thin films deposited by PECVD. The vertically actuated resonators consisted of doubly clamped microbridges suspended over fixed Al electrodes. The a-Si:H films were deposited using SiH_4 and H_2 precursors and PH_3 as a doping gas. The substrate temperature was held to around 100 °C, which enabled the use of photoresist as a sacrificial layer. The microbridges consisted of a large paddle suspended by two thin paddle supports, with the paddle providing a large reflective surface for optical detection of resonant frequency. The megahertz-frequency resonators exhibited quality factors in the 1×10^5 range when tested in vacuum. This work has since been extended to using polynorbornene as a sacrificial layer material [7.19]. A polynorbornene known as Unity 4011 thermally decomposes into vapor at 425 °C, enabling a dry, nonplasma-based release process. Surface micromachined resonators were successfully fabricated from a-Si:H films deposited by PECVD using this method. Cracking was observed near the anchor points of some structures and was attributed to differences in thermal expansion coefficients, an issue that is likely resolvable by modifying the design of the anchor regions.

7.1.3 Porous Silicon

Porous Si is produced by room-temperature electrochemical etching of Si in HF. If configured as an electrode in an HF-based electrochemical circuit, positive charge carriers (holes) at the Si surface facilitate the exchange of F atoms with H atoms that terminate the Si surface. The exchange continues in the subsurface region, leading to the eventual removal of the fluorinated Si. The quality of the etched surface is related to the density of holes at the surface, which is controlled by the applied current density. For high current densities, the density of holes is high and the etched surface is smooth. For low current densities, the hole density is low and clustered in highly localized regions associated with surface defects. Surface defects become enlarged by etching, which leads to the formation of pores. Pore size and density are related to the type of Si used and the conditions of the electrochemical cell. Both single-crystal and polycrystalline Si can be converted to porous Si.

The large surface-to-volume ratios make porous Si attractive for gaseous and liquid applications, including filter membranes and absorbing layers for chemical and mass sensing [7.20]. When single-crystal substrates are used, the unetched porous layer remains single crystalline and is suitable for epitaxial Si growth. It has been shown that CVD coatings do not generally penetrate the porous regions, but rather overcoat the pores at the surface of the substrate [7.21]. The formation of localized Si-on-insulator structures is therefore possible by simply combining pore formation with epitaxial growth, followed by dry etching to create access holes to the porous region, and thermal oxidation of the underlying porous region. A third application uses porous Si as a sacrificial layer for polysilicon and single-crystalline Si surface micromachining. As shown by *Lang* et al. [7.21], the process involves the electrical isolation of the solid structural Si layer by either pn-junction formation through selective doping, or use of electrically insulating thin films since the formation of pores only occurs on electrically charged surfaces. A weak Si etchant will aggressively attack the porous regions with little damage to the structural Si layers, and can be used to release the devices.

Porous polysilicon is also being developed as a structural material for chip-level vacuum packaging [7.22]. In this example, a 1.5 μm-thick polysilicon is deposited deposited onto a supporting PSG sacrificial layer, electrochemically etched in an HF solution to render it porous and then annealed by RTA to reduce stress in the porous layer. When fabricated locally over a prefabricated device structure (prior to release), the porous Si forms a localized shell, which will serve as a mechanical support for the main packaging structure. The porous structure enables an HF etch to remove the supporting PSG layer as well as any sacrificial oxide layers associated with the prefabricated MEMS device. After the sacrificial etch, the packaging sequence is completed by depositing a polysilicon film by LPCVD at 179 mtorr on the porous shell, thus fully encapsulating the device under vacuum conditions. This technique was used to package a microfabricated Pirani vacuum gage, which enabled an in situ measurement of pressure versus time. The authors found no detectable change in pressure over a three-month period. This process has matured to the point where it has successfully been incorporated into the well-known multiuser MEMS process (MUMPS) foundry service for polysilicon surface micromachining [7.23].

7.1.4 Silicon Dioxide

Silicon dioxide (SiO_2) is one of the most widely used materials in the fabrication of MEMS. In polysilicon surface micromachining, SiO_2 is used as a sacrificial material since it can be easily dissolved using etchants that do not attack polysilicon. SiO_2 is widely used as an etch mask for dry etching of thick polysilicon films, since it is chemically resistant to dry etching processes for polysilicon. SiO_2 films are also used as passivation

layers on the surfaces of environmentally sensitive devices.

The most common processes used to produce SiO_2 films for polysilicon surface micromachining are thermal oxidation and LPCVD. Thermal oxidation of Si is performed at temperatures of 900 to 1200 °C in the presence of oxygen or steam. Since thermal oxidation is a self-limiting process, the maximum practical film thickness that can be obtained is about 2 μm, which is sufficient for many sacrificial applications. As noted by its name, thermal oxidation of Si can only be performed on Si surfaces.

SiO_2 films can be deposited on a wide variety of substrate materials by LPCVD. In general, LPCVD provides a means for depositing thick (> 2 μm) SiO_2 films at temperatures much lower than thermal oxidation. Known as low-temperature oxides or LTOs for short, these films have a higher etch rate in HF than thermal oxides, which translates to significantly faster release times when LTO films are used as sacrificial layers. Phosphosilicate glass (PSG) can be formed using nearly the same deposition process as LTO by adding a phosphorus-containing gas to the precursor flows. PSG films are useful as sacrificial layers since they generally have higher etching rates in HF than LTO films.

PSG and LTO films are deposited in hot-wall, low-pressure, fused silica furnaces in systems similar to those described previously for polysilicon. Precursor gases include SiH_4 as a Si source, O_2 as an oxygen source, and, in the case of PSG, PH_3 as a source of phosphorus. LTO and PSG films are typically deposited at temperatures of 425–450 °C and pressures ranging from 200 to 400 mtorr. The low deposition temperatures result in LTO and PSG films that are slightly less dense than thermal oxides due to the incorporation of hydrogen in the films. LTO films can, however, be densified by an annealing step at high temperature (1000 °C). The low density of LTO and PSG films is partially responsible for the increased etch rate in HF.

Thermal SiO_2 and LTO are electrical insulators used in numerous MEMS applications. The dielectric constants of thermal oxide and LTO are 3.9 and 4.3 respectively. The dielectric strength of thermal SiO_2 is 1.1×10^6 V/cm, and for LTO it is about 80% of that value [7.24]. The stress in thermal SiO_2 is compressive with magnitude of about 300 MPa [7.24]. For LTO, however, the typical as-deposited residual stress is tensile, with a magnitude of about 100–400 MPa [7.24]. The addition of phosphorous to LTO decreases the tensile residual stress to about 10 MPa for phosphorus concentrations of 8% [7.25]. As with polysilicon, the properties of LTO and PSG are dependent on processing conditions.

Plasma-enhanced chemical vapor deposition (PECVD) is another common method to produce oxides of silicon. Using a plasma to dissociate the gaseous precursors, the deposition temperatures needed to deposit PECVD oxide films is lower than for LPCVD films. For this reason, PECVD oxides are quite commonly used as masking, passivation, and protective layers especially on devices that have been coated with metals.

Quartz is the crystalline form of SiO_2 and has interesting properties for MEMS. Quartz is optically transparent, piezoelectric, and electrically insulating. Like single-crystal Si, quartz substrates are available as high-quality, large-area wafers that can be bulk micromachined using anisotropic etchants. A short review of the basics of quartz etching was written by *Danel* et al. [7.26] and is recommended for those interested in the subject. Quartz is a popular substrate material for microfluidic devices due to its optical, electronic and chemical properties.

Another SiO_2-related material that has found uses in MEMS is spin-on-glass (SOG). SOG is a polymeric material with a viscosity suitable for spin coating. Two recent publications illustrate the potential for SOG in MEMS fabrication. In the first example, *Yasseen* et al. [7.27] detailed the development of SOG as a thick-film sacrificial molding material for thick polysilicon films. The authors reported a process to deposit, polish and etch SOG films that were 20 μm thick. The thick SOG films were patterned into molds and filled with 10 μm-thick LPCVD polysilicon films, planarized by selective chemical mechanical polishing (CMP), and subsequently dissolved in a wet etchant containing HCl, HF, and H_2O to reveal the patterned polysilicon structures. The cured SOG films were completely compatible with the polysilicon deposition process. In the second example, *Liu* et al. [7.28] fabricated high aspect ratio channel plate microstructures from SOG. Electroplated nickel (Ni) was used as a molding material, with Ni channel plate molds fabricated using a conventional Lithography, Galvanoforming, and Abformung (LIGA) process. The Ni molds were then filled with SOG, and the sacrificial Ni molds were removed in a reverse electroplating process. In this case, the fabricated SOG structures (over 100 μm tall) were micromachined glass structures fabricated using a molding material more commonly used for structural components.

7.1.5 Silicon Nitride

Silicon nitride (Si_3N_4) is widely used in MEMS for electrical isolation, surface passivation, etch masking, and as a mechanical material typically for membranes and other suspended structures. Two deposition meth-

ods are commonly used to deposit Si_3N_4 thin films, LPCVD and PECVD. PECVD silicon nitride is generally nonstoichiometric (sometimes denoted as Si_xN_y:H) and may contain significant concentrations of hydrogen. The use of PECVD silicon nitride in micromachining applications is somewhat limited because it has a high etch rate in HF (e.g., often higher than that of thermally grown SiO_2). However, PECVD offers the ability to deposit nearly stress-free silicon nitride films, an attractive property for encapsulation and packaging. Recent experiments using bulk micromachined square and rectangular membranes have shown that Si_xN_y:H thin films ($\approx$ 300–700 nm) deposited by PECVD have a Young's modulus that increases with increasing deposition temperature, from 79 GPa at 125 °C to 151 at 205 °C [7.29]. The residual stresses in these films are below 30 MPa for each deposition temperature.

Unlike its PECVD counterpart, LPCVD Si_3N_4 is extremely resistant to chemical attack, thereby making it the material of choice for many Si bulk and surface micromachining applications. LPCVD Si_3N_4 is commonly used as an insulating layer because it has a resistivity of 10^{16} Ω cm and field breakdown limit of 10^7 V/cm. LPCVD Si_3N_4 films are deposited in horizontal furnaces similar to those used for polysilicon deposition. Typical deposition temperatures and pressures range between 700–900 °C and 200–500 mtorr respectively. The standard source gases are dichlorosilane (SiH_2Cl_2) and ammonia (NH_3). To produce stoichiometric Si_3N_4 a NH_3 to SiH_2Cl_2 ratio 10:1 is commonly used. The microstructure of films deposited under these conditions is amorphous.

The residual stress in stoichiometric Si_3N_4 is large and tensile, with a magnitude of about 1 GPa. Such a large residual stress causes films thicker than a few hundred nanometers to crack. Nonetheless thin stoichiometric Si_3N_4 films have been used as mechanical support structures and electrical insulating layers in piezoresistive pressure sensors [7.30]. To enable the use of Si_3N_4 films for applications that require micron-thick, durable and chemically resistant membranes, Si_xN_y films can be deposited by LPCVD. These films, often referred to as Si-rich or low-stress nitride, are intentionally deposited with an excess of Si by simply decreasing the ratio of NH_3 to SiH_2Cl_2 during deposition. Nearly stress-free films can be deposited using a NH_3 to SiH_2Cl_2 ratio of 1:6, a deposition temperature of 850 °C and a pressure of 500 mtorr [7.31]. The increase in Si content not only leads to a reduction in tensile stress, but also a decrease in the etch rate in HF. Such properties have enabled the development of fabrication techniques that would otherwise not be feasible with stoichiometric Si_3N_4. For example, low-stress silicon nitride has been bulk micromachined using silicon as the sacrificial material [7.32]. In this case, Si anisotropic etchants (TMAH) were used for dissolving the sacrificial silicon. *French* et al. [7.33] used PSG as a sacrificial layer to surface micromachine low-stress nitride, capitalizing on the HF resistance of the nitride films.

7.2 Germanium-Based Materials

7.2.1 Polycrystalline Germanium

Like Si, Ge has a long history as a semiconductor device material, dating back to the development of the earliest transistors and semiconductor strain gages. Issues related to germanium oxide, however, stymied the development of Ge for microelectronic devices. Nonetheless, there is a renewed interest in using Ge in surface micromachined devices due to the relatively low processing temperatures required to deposit the material and its compatibility with Si.

Thin polycrystalline Ge (poly-Ge) films can be deposited by LPCVD at temperatures as low as 325 °C on Si, Ge and SiGe substrates [7.34]. Ge does not nucleate on SiO_2 surfaces, which prohibits the use of thermal oxides and LTO films as sacrificial layers, but enables the use of these films as sacrificial molds. Residual stress in poly-Ge films deposited on Si substrates can be reduced to nearly zero after short anneals at modest temperatures (30 s at 600 °C). Poly-Ge is essentially impervious to KOH, TMAH, and BOE, enabling the fabrication of Ge membranes on Si substrates [7.34]. The mechanical properties of poly-Ge are comparable to polysilicon, having a Young's modulus of 132 GPa and a fracture stress ranging between 1.5 GPa and 3.0 GPa [7.35]. Mixtures of HNO_3, H_2O, and HCl and H_2O, H_2O_2, and HCl, as well as the RCA SC-1 cleaning solution isotropically etch Ge. Since these mixtures do not etch Si, SiO_2, Si_3N_4 and SiN, poly-Ge can be used as a sacrificial substrate layer in polysilicon surface micromachining. Using these techniques, poly-Ge-based thermistors and Si_3N_4 membrane-based pressure sensors made using poly-Ge sacrificial layers have been fabricated [7.34]. *Franke* et al. [7.35] found no performance degradation in Si CMOS devices following the fabrication of surface micromachined poly-Ge structures, thus demonstrating the potential for on-chip integration of Ge electromechanical devices with Si circuitry.

7.2.2 Polycrystalline Silicon Germanium

Like poly-Ge, polycrystalline SiGe (poly-SiGe) is a material that can be deposited at temperatures lower than polysilicon. Deposition processes include LPCVD, APCVD, and rapid thermal CVD (RTCVD) using SiH_4 and GeH_4 as precursor gases. Deposition temperatures range between 450 °C for LPCVD [7.36] to 625 °C for RTCVD [7.37]. In general, the deposition temperature is related to the concentration of Ge in the films, with higher Ge concentrations resulting in lower deposition temperatures. Like polysilicon, poly-SiGe can be doped with boron and phosphorus to modify its conductivity. In situ boron doping can be performed at temperatures as low as 450 °C [7.36]. *Sedky* et al. [7.38] showed that the deposition temperature of conductive films doped with boron could be further reduced to 400 °C if the Ge content was kept at or above 70%. Recent work [7.39] has shown that films grown at temperatures around 400 °C exhibit a strain gradient in the range of $\approx 1\times10^{-5}/\mu m$. This strain gradient is attributed to the columnar microstructure of the films combined with a high compressive stress at the film–substrate interface. Boron doping enhances uniformity in the columnar microstructure throughout the thickness of the film and thus reduces the strain gradient. It has been reported that deposition temperatures for poly-SiGe can be reduced to 210 °C, however to achieve strain gradients on the order of $1\times10^{-6}/\mu m$, the films must be annealed following deposition [7.40]. Since the objective of this effort is to minimize thermal budget, localized annealing by way of excimer laser illumination is required.

Unlike poly-Ge, poly-SiGe can be deposited on a number of sacrificial substrates, including SiO_2 [7.37], PSG [7.35] and poly-Ge [7.35]. For Ge-rich films, a thin polysilicon seed layer is sometimes used on SiO_2 surfaces since Ge does not readily nucleate on oxide surfaces. Like many compound materials, variations in film composition can change the physical properties of the material. For instance, etching of poly-SiGe by H_2O_2 becomes significant for Ge concentrations over 70%. *Sedky* et al. [7.38] have shown that the microstructure, film conductivity, residual stress, and residual stress gradient are related to the concentration of Ge in the material. With respect to residual stress, *Franke* et al. [7.36] produced in situ boron-doped films with residual compressive stresses as low as 10 MPa.

The poly-SiGe, poly-Ge material system is particularly attractive for surface micromachining since H_2O_2 can be used as a release agent. It has been reported that poly-Ge etches at a rate of 0.4 μm/min in H_2O_2, while poly-SiGe with Ge concentrations below 80% has no observable etch rate after 40 h [7.41]. The ability to use H_2O_2 as a sacrificial etchant makes the combination of poly-SiGe and poly-Ge extremely attractive for surface micromachining from processing, safety and materials compatibility points of view. Due to the conformal nature of LPCVD processing, poly-SiGe structural elements, such as gimbal-based microactuator structures, have been made by high-aspect-ratio micromolding [7.41]. Capitalizing on the low deposition temperatures, poly-SiGe MEMS structures integrated with Si ICs have been demonstrated [7.36]. In this process, CMOS structures are first fabricated on Si wafers. Poly-SiGe mechanical structures are then surface micromachined using a poly-Ge sacrificial layer. A significant advantage of this design lies in the fact that the MEMS structure is positioned directly above the CMOS structure, thus reducing the parasitic capacitance and contact resistance characteristic of interconnects associated with side-by-side integration schemes. The use of H_2O_2 as the sacrificial etchant eliminates the need for layers to protect the underlying CMOS structure during release. In addition to its utility as a material for integrated MEMS devices, poly-SiGe has been identified as a material well suited for micromachined thermopiles [7.42] due to its lower thermal conductivity relative to Si.

7.3 Metals

It can be argued that of all the material categories associated with MEMS, metals may be among the most enabling, since metallic thin films are used in many different capacities, ranging from etch masks used in device fabrication to interconnects and structural elements in microsensors and microactuators. Metallic thin films can be deposited using a wide range of techniques, including evaporation, sputtering, CVD, and electroplating. Since a complete review of the metals used in MEMS is far beyond the scope of this chapter, the examples presented in this section were selected to represent a broad cross-section where metals have found uses in MEMS.

Aluminum (Al) and gold (Au) are among the most widely employed metals in microfabricated electronic and electromechanical devices as a result of their use as interconnect and packaging materials. In addition to these critical electrical functions, Al and Au are also de-

sirable as electromechanical materials. One such example is the use of Au micromechanical switches for radio frequency MEMS (RF MEMS. For conventional RF applications, chip-level switching is currently performed using FET- and PIN diode-based solid-state devices fabricated from gallium arsenide (GaAs) substrates (FET: field effect transistor; PIN: p-type/intrinsic/n-type). Unfortunately, these devices suffer from insertion losses and poor electrical isolation. In an effort to develop replacements for GaAs-based solid-state switches, *Hyman* et al. [7.43] reported the development of an electrostatically actuated, cantilever-based micromechanical switch fabricated on GaAs substrates. The device consisted of a silicon nitride-encased Au cantilever constructed on a sacrificial silicon dioxide layer. The silicon nitride and silicon dioxide layers were deposited by PECVD, and the Au beam was electroplated from a sodium sulfite solution inside a photoresist mold. A thin multilayer of Ti and Au was sputter deposited in the mold prior to electroplating. The trilayered cantilever structure was chosen to minimize the deleterious effects of thermal and process-related stress gradients in order to produce unbent and thermally stable beams. After deposition and pattering, the cantilevers were released in HF. The processing steps proved to be completely compatible with GaAs substrates. The released cantilevers demonstrated switching speeds of better than 50 μs at 25 V with contact lifetimes exceeding 10^9 cycles.

In a second example from RF MEMS, *Chang* et al. [7.44] reported the fabrication of an Al-based micromachined switch as an alternative to GaAs FETs and PIN diodes. In contrast to the work by *Hyman* et al. [7.43], this switch utilizes the differences in the residual stresses in Al and Cr thin films to create bent cantilever switches that capitalize on the stress differences in the materials. Each switch is comprised of a series of linked bimorph cantilevers designed in such a way that the resulting structure bends significantly out of the plane of the wafer due to the stress differences in the bimorph. The switch is drawn closed by electrostatic attraction. The bimorph consists of metals that can easily be processed with GaAs wafers, thus making integration with GaAs devices possible. The released switches were relatively slow, at 10 ms, but an actuation voltage of only 26 V was needed to close the switch.

A recent study published by *Reddy* et al. documents the use of micromachined structures to characterize the mechanical properties of thin Al and Ni films [7.45]. The structures, fabricated from sputter-deposited films, consisted of bulk micromachined suspended metal membranes and surface micromachined metal-coated lateral resonators. The researchers found that the Young's modulus for the Al films was 70 GPa and that for the Ni films was 200 GPa, values that are in agreement with bulk samples.

Direct bulk micromachining of metal substrates is being developed for MEMS applications requiring structures with the dimensional complexity associated with Si DRIE and the physical properties of metals. One such example is Ti, which has a higher fracture toughness, a greater biocompatibility, and a more stable passivating oxide than Si. A process to fabricate high-aspect-ratio, three-dimensional structures from bulk Ti substrates has been developed [7.46]. This process involves inductively coupled plasma (ICP) etching of a TiO_2-capped Ti substrate. The TiO_2 capping layer is deposited by DC reactive sputtering and photolithographically patterned using a CHF_3-based dry etch. The deep Ti etch is then performed using a Cl/Ar-based plasma that exhibits a selectivity of 40:1 with the masking TiO_2 layer. The etch process consists of a series of two-step sequences, where the first step involves Ti removal by the Cl/Ar plasma while the second step involves sidewall passivation using an oxygen plasma. After the prescribed etch period, the masking thin film can be removed by HF etching. High-aspect-ratio comb-drive actuators and other beam-based structures have been fabricated directly from bulk Ti using this method. An effort to develop Ti microneedles that incorporate microfluidic channels for drug delivery applications fabricated using this technique has been published [7.47].

Thin-film metallic alloys that exhibit the shape-memory effect are of particular interest to the MEMS community for their potential in microactuators. The shape-memory effect relies on the reversible transformation from a ductile martensite phase to a stiff austenite phase in the material with the application of heat. The reversible phase change allows the shape-memory effect to be used as an actuation mechanism since the material changes shape during the transition. It has been found that high forces and strains can be generated from shape-memory thin films at reasonable power inputs, thus enabling shape-memory actuation to be used in MEMS-based microfluidic devices such as microvalves and micropumps. Titanium-nickel (TiNi) is among the most popular of the shape-memory alloys owing to its high actuation work density, ($50\,MJ/m^3$), and large bandwidth (up to 0.1 kHz) [7.48]. TiNi is also attractive because conventional sputtering techniques can be employed to deposit thin films, as detailed in a recent report by *Shih* et al. [7.48]. In this study, TiNi films were deposited by cosputtering elemental Ti and Ni targets, and cosputtering TiNi alloy and elemental Ti targets. It was reported that cosputtering from TiNi and Ti targets produced better films due to process variations related to roughening of the

Ni target in the case of Ti and Ni cosputtering. The TiNi/Ti co-puttering process has been used to produce a shape-memory material for a silicon spring-based microvalve [7.49].

The use of thin-film metal alloys in magnetic actuator systems is another example of the versatility of metallic materials in MEMS. Magnetic actuation in microdevices generally requires the magnetic layers to be relatively thick (tens to hundreds of microns) to generate magnetic fields of sufficient strength to generate the desired actuation. To this end, magnetic materials are often deposited by thick-film methods such as electroplating. The thicknesses of these layers exceeds what can feasibly be patterned by etching, so plating is often performed in microfabricated molds made from materials such as polymethylmethacrylate (PMMA). The PMMA mold thickness can exceed several hundred microns, so x-rays are used as the exposure source during the patterning steps. When necessary, a metallic thin-film seed layer is deposited prior to plating. After plating, the mold is dissolved, which frees the metallic component. Known as LIGA (short for Lithography, Galvanoforming, and Abformung), this process has been used to produce a wide variety of high-aspect ratio structures from plateable materials, such as nickel-iron (NiFe) magnetic alloys [7.50] and Ni [7.51].

In addition to elemental metals and simple compound alloys, more complex metallic alloys commonly used in commercial macroscopic applications are finding their way into MEMS applications. One such example is an alloy of titanium known as Ti-6Al-4V. Composed of 88% titanium, 6% aluminum and 4% vanadium, this alloy is widely used in commercial aviation due to its weight, strength and temperature tolerance. *Pornsin-Sirirak* et al. [7.52] have explored the use of this alloy in the manufacture of MEMS-based winged structures for micro-aerial vehicles. The authors considered this alloy not only because of its weight and strength, but also because of its ductility and its etching rate at room temperature. The designs for the wing prototype were modeled after the wings of bats and various flying insects. For this application, Ti-alloy structures patterned from bulk (250 μm-thick) material by an $HF/HO_3/H_2O$ etching solution were used rather than thin films. Parylene-C (detailed in a later section) was deposited on the patterned alloy to serve as the wing membrane. The miniature micromachined wings were integrated into a test setup and several prototypes actually demonstrated short-duration flight.

7.4 Semiconductors for Harsh Environment Applications

7.4.1 Silicon Carbide

Silicon Carbide (SiC) has long been recognized as the leading semiconductor for use in high-temperature and high-power electronics, and is currently being developed as a material for harsh environment MEMS. SiC is a polymorphic material that exists in cubic, hexagonal, and rhombehedral polytypes. The cubic polytype, called 3C-SiC, has an electronic bandgap of 2.3 eV, which is over twice that of Si. Numerous hexagonal and rhombehedral polytypes have been identified, with the two most common being 4H-SiC and 6H-SiC. The electronic bandgaps of 4H- and 6H-SiC are even higher than 3C-SiC, being 2.9 and 3.2 eV respectively. SiC films can be doped to create n-type and p-type materials. The Young's modulus of SiC is still the subject of research, but most reported values range from 300 GPa to 450 GPa, depending on the microstructure and measurement technique. SiC is not etched in any wet Si etchants and is not attacked by XeF_2, a popular dry Si etchant used for releasing device structures [7.53]. SiC is a material that does not melt, but rather sublimes at temperatures in excess of 1800 °C. Single-crystal 4H- and 6H-SiC wafers are commercially available, but are generally smaller in diameter (3 in) and much more expensive than Si wafers.

SiC thin films can be grown or deposited using a number of different techniques. For high-quality single-crystal films, APCVD and LPCVD processes are most commonly employed. Homoepitaxial growth of 4H- and 6H-SiC yields high-quality films suitable for microelectronic applications but typically only on substrates of the same polytype. These processes usually employ dual precursors, such as SiH_4 and C_3H_8, and are performed at temperatures ranging from 1500 to 1700 °C. Epitaxial films with p-type or n-type conductivity can be grown using Al and B for p-type films and N and P for n-type films. Nitrogen is so effective at modifying the conductivity of SiC that growth of undoped SiC films is extremely challenging because the concentrations of residual nitrogen in typical deposition systems are sufficient for n-type doping.

APCVD and LPCVD can also be used to deposit 3C-SiC on Si substrates. Heteroepitaxy is possible despite a 20% lattice mismatch because 3C-SiC and Si have the same lattice structure. The growth process involves two key steps. The first step, called carbonization, converts the near-surface region of the Si

substrate to 3C-SiC by simply exposing it to a hydrocarbon/hydrogen mixture at high substrate temperatures (> 1200 °C). The carbonized layer forms a crystalline template on which a 3C-SiC film can be grown by adding a silicon-containing gas to the hydrogen/hydrocarbon mix. The lattice mismatch between Si and 3C-SiC results in the formation of crystalline defects in the 3C-SiC film, with the density being highest in the carbonization layer and decreasing with increasing thickness. The crystal quality of 3C-SiC films is nowhere near that of epitaxially grown 4H- and 6H-SiC films; however, the fact that 3C-SiC can be grown on Si substrates enables the use of Si bulk micromachining techniques for fabrication of a host of 3C-SiC-based mechanical devices. These include microfabricated pressure sensors [7.54] and nanoelectromechanical resonant structures [7.55]. For designs that require electrical isolation from the substrate, 3C-SiC devices can be made directly on SOI substrates [7.54] or by wafer bonding and etchback, such as the capacitive pressure sensor developed by *Young* et al. [7.56].

Polycrystalline SiC (poly-SiC) is a more versatile material for SiC MEMS than its single-crystal counterparts. Unlike single-crystal versions of SiC, poly-SiC can be deposited on a variety of substrate types, including common surface micromachining materials such as polysilicon, SiO_2 and Si_3N_4. Commonly used deposition techniques include LPCVD [7.53, 57, 58] and APCVD [7.59, 60]. The deposition of poly-SiC requires much lower substrate temperatures than epitaxial films, ranging from roughly 700 to 1200 °C. Amorphous SiC can be deposited at even lower temperatures (25–400 °C) by PECVD [7.61] and sputtering [7.62]. The microstructure of poly-SiC films is temperature-, substrate-, and process-dependent. For amorphous substrates such as SiO_2 and Si_3N_4, APCVD poly-SiC films deposited from SiH_4 and C_3H_8 are randomly oriented with equiaxed grains [7.60], whereas for oriented substrates such as polysilicon, the texture of the poly-SiC film matches that of the substrate itself [7.59]. By comparison, poly-SiC films deposited by LPCVD from SiH_2Cl_2 and C_2H_2 are highly textured (111) films with a columnar microstructure [7.57], while films deposited from disilabutane have a distribution of orientations [7.53]. This variation suggests that device performance can be tailored by selecting the proper substrate and deposition conditions.

SiC films deposited by AP- and LPCVD generally suffer from large residual stresses that are tensile and on the order of several 100 MPa. Moreover, the residual stress gradients in these films tend to be large, leading to significant out-of-plane bending of structures anchored at a single location. The thermal stability of SiC makes a postdeposition annealing step impractical for films deposited on Si substrates, since the temperatures needed to significantly modify the film are likely to exceed the melting temperature of the wafer. For LPCVD processes using SiH_2Cl_2 and C_2H_2 precursors, *Fu* et al. [7.63] have described a relationship between deposition pressure and residual stress that enables the deposition of undoped poly-SiC films with nearly zero residual stresses and negligible stress gradients. This work has been extended to include films doped with nitrogen [7.64].

Direct bulk micromachining of SiC is very difficult, due to its chemical inertness. Although conventional wet chemical techniques are not effective, several electrochemical etch processes have been demonstrated and used in the fabrication of 6H-SiC pressure sensors [7.65]. The etching processes are selective to the conductivity of the material, so dimensional control of the etched structures depends on the ability to form doped layers, which can only be formed by in situ, or ion implantation processes since solid source diffusion is not possible at reasonable processing temperatures. This constraint somewhat limits the geometrical complexity of the patterned structures as compared to conventional plasma-based etching. To fabricate thick (100s of microns), three-dimensional (3-D), high-aspect-ratio SiC structures, a molding technique has been developed [7.51]. The molds are fabricated from Si substrates using deep reactive ion etching and then filed with SiC using a combination of thin epitaxial and thick polycrystalline film CVD processes. The thin-film process is used to protect the mold from pitting during the more aggressive mold-filling SiC growth step. The mold-filling process coats all surfaces of the mold with a SiC film as thick as the mold is deep. To release the SiC structure, the substrate is first mechanically polished to expose sections of the Si mold, then the substrate is immersed in a Si etchant to completely dissolve the mold. This process has been used to fabricate solid SiC fuel atomizers [7.51], and a variant has been used to fabricate SiC structures for micropower systems [7.66]. *Min* et al. [7.67] reported a process to fabricate reusable glass press molds made from SiC structures that were patterned using Si molding masters. SiC was selected as the material for the glass press mold, because the application requires a hard, mechanically strong, and chemically stable material that can withstand and maintain its properties at temperatures between 600–1400 °C.

In addition to CVD processes, bulk micromachined SiC structures can be fabricated using sintered SiC powders. *Tanaka* et al. [7.68] describe a process where SiC components, such as microscale gas turbine engine rotors, can be fabricated from SiC powders using a microreaction-sintering process. The molds are mi-

crofabricated from Si using DRIE, and filled with SiC and graphite powders mixed with a phenol resin. The molds are then reaction-sintered using a hot isostatic pressing technique. The SiC components are then released from the Si mold by wet chemical etching. The authors reported that the component shrinkage was less than 3%. The bending strength and Vickers hardness of the microreaction-sintered material was roughly 70–80% of commercially available reaction-sintered SiC, the difference being attributed to the presence of unreacted Si in the microscale components.

In a related process, *Liew* et al. [7.69] detail a technique to create silicon carbon nitride (SiCN) MEMS structures by molding injectable polymer precursors. Unlike the aforementioned processes, this technique uses SU-8 photoresists for the molds. To be detailed later in this chapter, SU-8 is a versatile photodefinable polymer in which thick films (100s of microns) can be patterned using conventional ultraviolet (UV) photolithographic techniques. After patterning, the molds are filled with the SiCN-containing polymer precursor, lightly polished, and then subjected to a multistep heat-treating process. During the thermal processing steps, the SU-8 mold decomposes and the SiCN structure is released. The resulting SiCN structures retain many of the same properties of stoichiometric SiC.

Although SiC cannot be etched using conventional wet etch techniques, SiC can be patterned using conventional dry etching techniques. RIE processes using fluorinated compounds such as CHF_3 and SF_6 combined with O_2 and sometimes with an inert gas or H_2 are used to pattern thin films. The high oxygen content in these plasmas generally prohibits the use of photoresist as a masking material; therefore hard masks made of Al, Ni, and ITO are often used. RIE-based SiC surface micromachining processes with polysilicon and SiO_2 sacrificial layers have been developed for single-layer devices [7.70, 71]. ICP RIE of SiC using SF_6 plasmas and Ni or ITO etch masks has been extended to bulk micromachine SiC substrates by DRIE, with structural depths in excess of 100 µm reported [7.72]. *Sharpe* et al. reported on the use of bulk micromachined tensile specimens to characterize the tensile strength of single-crystalline SiC structures at temperatures up to 1000 °C [7.73]. The structures were made from commercially available 6H-SiC wafers and micromachined by DRIE. They found that the the testing specimens exhibited a fracture strength between 0.45 and 1.2 GPa at room temperature depending on specimen geometry, which varied only slightly with increasing temperature.

Initially, multilayer thin-film structures were very difficult to fabricate by direct RIE because the etch rates of the sacrificial layers were much higher than the SiC structural layers, making dimensional control very difficult. To address this issue, a micromolding process for patterning SiC films on sacrificial layer substrates was developed [7.74]. In essence, the micromolding technique is the thin-film analog to the molding-based, bulk micromachining technique presented earlier. The micromolding process utilizes polysilicon and SiO_2 films as both molds and sacrificial substrate layers, with SiO_2 molds used with polysilicon sacrificial layers and vice versa. These films are deposited and patterned using conventional methods, thus leveraging the well-characterized and highly selective processes developed for polysilicon MEMS. Poly-SiC films are simply deposited into the micromolds and mechanical polishing is used to remove poly-SiC from atop the molds. Appropriate etchants are then used to dissolve the molds and sacrificial layers. The micromolding method utilizes the differences in chemical properties of the three materials in this system in a way that bypasses the difficulties associated with chemical etching of SiC. This technique has been developed specifically for multilayer processing, and has been used successfully to fabricate SiC micromotors [7.74] and the lateral resonant structure shown in Fig. 7.4 [7.75].

Recent advancements in the area of SiC RIE show that significant progress has been made in developing etch recipes with selectivities to nonmetal mask and sacrificial layers that are suitable for multilayer SiC surface micromachining. For instance, *Gao* et al. [7.76] have developed a transformer coupled RIE process using a HBr-based chemistry for thin-film poly-SiC etching. The recipe exhibits a SiC-to-SiO_2 selectivity of 20:1 and a SiC-to-Si_3N_4 selectivity of 22:1, which are the highest reported thus far. In addition, the anisotropy of the etch was quite high, and micromasking, a common problem when metal masks are used, was

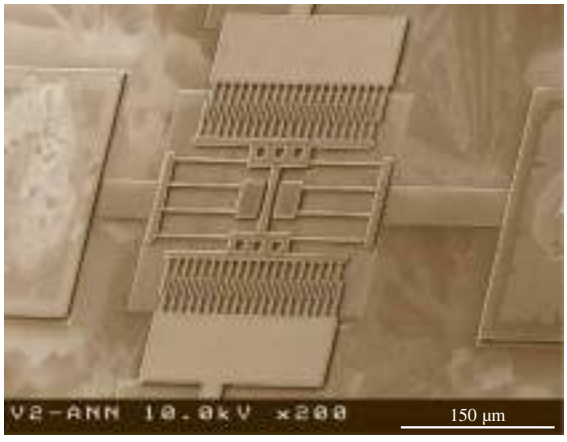


Fig. 7.4 SEM micrograph of a poly-SiC lateral resonant structure fabricated using a multilayer, micromolding-based micromachining process (after [7.75])

not an issue. This process has since been used to fabricate multilayered lateral resonant structures that utilize poly-SiC as the main structural material and polysilicon as a conducting plane that lies underneath the resonating shuttle [7.77].

Yang et al. [7.55] have shown that the chemical inertness of SiC facilitates the fabrication of NEMS devices. In this work, the authors present a fabrication method to realize SiC mechanical resonators with submicron thickness and width dimensions. The resonators were fabricated from ≈ 260 nm-thick 3C-SiC films epitaxially grown on (100) Si wafers. The films were patterned into 150 nm-wide beams ranging in length from 2 to 8 μm. The beams were etched in a $NF_3/O_2/Ar$ plasma using an evaporated Cr etch mask. After patterning, the beams were released by etching the underlying Si isotropically using a NF_3/Ar plasma. The inertness of the SiC film to the Si etchant enables the dry release of the nanomechanical beams. An example of a 3C-SiC nanomechanical beam is shown in Fig. 7.5. The first nanomechanical resonator to exhibit a fundamental resonant frequency in excess of 1 GHz was fabricated from a 3C-SiC film using the aforementioned process [7.78].

7.4.2 Diamond

Diamond is commonly known as nature's hardest material, making it ideal for high-wear environments. Diamond has a very large electronic bandgap (5.5 eV), which makes it attractive for high-temperature electronics. Undoped diamond is a high-quality insulator with a dielectric constant of 5.5; however it can be relatively easily doped with boron to create p-type conductivity. Diamond has a very high Young's modulus (1035 GPa) making it suitable for high-frequency micromachined resonators, and it is among nature's most chemically inert materials, making it well suited for harsh chemical environments.

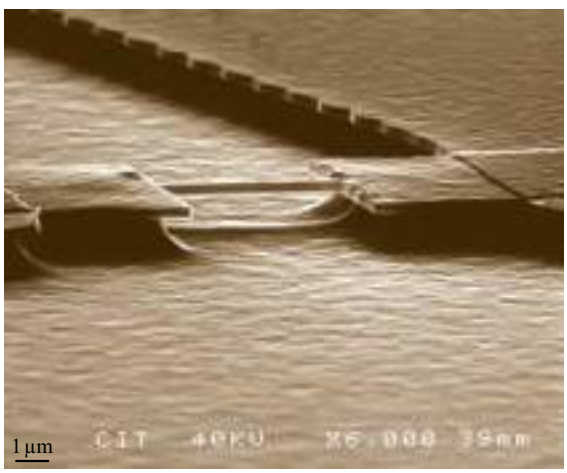


Fig. 7.5 SEM micrograph of a 3C-SiC nanomechanical beam resonator fabricated by electron-beam lithography and dry etching processes (courtesy M. Roukes, Caltech)

Unlike SiC, fabrication of diamond MEMS is currently restricted to polycrystalline and amorphous material, since single-crystal diamond wafers are not yet commercially available. Polycrystalline diamond films can be deposited on Si and SiO_2 substrates by CVD methods, but the surfaces must often be seeded by diamond powders or biased with a negative charge to initiate growth. In general, diamond nucleates much more readily on Si surfaces than on SiO_2 surfaces, an effect that has been used to selectively pattern diamond films into micromachined atomic force microscope (AFM) cantilever probes using SiO_2 molding masks [7.79].

Bulk micromachining of diamond using wet and dry etching is extremely difficult given its extreme chemical inertness. Diamond structures have nevertheless been fabricated using bulk micromachined Si molds to pattern the structures [7.80]. The Si molds were fabricated using conventional micromachining techniques and filed with polycrystalline diamond deposited by hot filament chemical vapor deposition (HFCVD). The HFCVD process uses H_2 and CH_4 precursors. The process was performed at a substrate temperature of 850–900 °C and a pressure of 50 mtorr. The Si substrate was seeded prior to deposition using a diamond particle/ethanol solution. After deposition, the top surface of the structure was polished using a hot iron plate. After polishing, the Si mold was removed in a Si etchant, leaving behind the micromachined diamond structure. This process was used to produce high-aspect-ratio capillary channels for microfluidic applications [7.81], and components for diffractive optics, laser-to-fiber alignment, and power device cooling structures [7.82].

Due to the nucleation processes associated with diamond film growth, surface micromachining of polycrystalline diamond thin films requires modifications to conventional micromachining to facilitate film growth on sacrificial substrates. Initially, conventional RIE methods were generally ineffective, so work was focused on developing selective deposition techniques. One early method used selective seeding to form patterned templates for diamond nucleation. The selective seeding process employed the lithographic patterning of photoresist that contained diamond powders [7.83]. The diamond-loaded photoresist was deposited and patterned onto a Cr-coated Si wafer. During the onset of diamond growth, the patterned photoresist rapidly evaporates, leaving behind the diamond seed particles in the desired locations. A patterned diamond film is then selectively grown on these locations.

A second process utilized selective deposition directly on sacrificial substrate layers. This processes combined conventional diamond seeding with photolithographic patterning and etching to fabricate micromachined diamond structures on SiO_2 sacrificial layers [7.84]. The process was performed in one of two ways. The first approach begins with the seeding of an oxidized Si wafer. The wafer is coated with photoresist, and photolithographically patterned. Unmasked regions of the seeded SiO_2 film are then partially etched, forming a surface unfavorable for diamond growth. The photoresist is then removed and a diamond film is deposited on the seeded regions. The second approach also begins with an oxidized Si wafer. The wafer is coated with photoresist, photolithographically patterned and then seeded with diamond particles. The photoresist is removed, leaving behind a patterned seed layer suitable for selective growth. These techniques have been successfully used to fabricate cantilever beams and bridge structures.

A third method to surface micromachine polycrystalline diamond films follows the conventional approach of film deposition, dry etching, and release. The chemical inertness of diamond renders most conventional plasma chemistries useless; however, oxygen-based ion beam plasmas can be used to etch diamond thin films [7.85]. A simple surface micromachining process begins with the deposition of a polysilicon sacrificial layer on a Si_3N_4-coated Si wafer. The polysilicon layer is seeded using diamond slurry and a diamond film is deposited by HFCVD. Since photoresist is not resistant to O_2 plasmas, an Al masking film is deposited and patterned. The diamond films are then etched in the O_2 ion beam plasma, and the structures are released by etching the polysilicon with KOH. This process has been used to create lateral resonant structures, but a significant stress gradient in the films rendered the devices inoperable.

In general, conventional HFCVD requires that the substrate be pretreated with a seeding layer prior to diamond film growth. However, a method called biased enhanced nucleation (BEN) has been developed that enables the growth of diamond on unseeded Si surfaces. *Wang* et al. [7.86] have shown that if Si substrates are masked with patterned SiO_2 films, selective diamond growth will occur primarily on the exposed Si surfaces, and a slight HF etch is sufficient to remove the adventitious diamond from the SiO_2 mask. This group was able to use this method to fabricate diamond micromotor rotors and stators on Si surfaces.

Diamond is a difficult, but not impossible material to etch using conventional RIE techniques. It is well known that diamond will be etched in oxygen plasmas, but these plasmas can be problematic for device fabrication because the etching tends to be isotropic. Fabrication of MEMS structures has been aided significantly by the development of RIE processes for diamond thin films. *Wang* et al. [7.87] describe a process to fabricate a vertically actuated, doubly clamped micromechanical diamond beam resonator using RIE. The process outlined in this paper addresses two key issues related to diamond surface micromachining, namely residual stress gradients in the diamond films and diamond patterning techniques. A microwave plasma CVD (MPCVD) reactor was used to grow the diamond films on sacrificial SiO_2 layers that were pretreated with a nanocrystalline diamond powder, resulting in a uniform nucleation density at the diamond/SiO_2 interface. The diamond films were etched in a CF_4/O_2 plasma using Al as a hard mask. Reasonably straight sidewalls were created, with roughness attributable to the surface roughness of the faceted diamond film. An Au/Cr drive electrode beneath the sacrificial oxide remained covered throughout the diamond patterning steps and thus was undamaged during the diamond etching process. This work has since been extended to develop a 1.51 GHz diamond micromechanical disk resonator [7.88]. In this instance, the nanocrystalline diamond film was deposited by MPCVD, coated with an oxide film that was patterned into an etch mask, and then etched in a O_2/CF_4 RIE plasma under conditions that yielded a fairly anisotropic etch with a diamond-to-oxide selectivity of 15:1. The disk was suspended over the substrate on a polysilicon stem using an oxide sacrificial layer. Polysilicon was also used as the drive and sense electrodes. The material mismatch between the step and the resonating disk substantially reduced anchor losses, thus allowing for very high quality factors (11 500) for 1.5 GHz resonators testing in vacuum. *Sepulveda* et al. have successfully fabricated cantilever-beam resonators by surface micromachining using silicon dioxide as a sacrificial layer material [7.89]. The diamond films were grown by MPCVD using 25 nm diamond powders as seeds for nucleation. Seeding was not done preferentially, so the diamond film was patterned by dry etching using a 300 nm-thick Ti film as an etch mask. Along similar lines, *Sekaric* et al. fabricated NEMS structures such as paddle oscillators, doubly clamped beams and mesh membranes from nanocrystalline diamond films grown on oxide-coated Si wafers by MPCVD using nanoparticle seeds [7.90].

In conjunction with recent advances in RIE and micromachining techniques, work is being performed to develop diamond deposition processes specifically for MEMS applications. Diamond films grown using conventional techniques, especially processes that require pregrowth seeding, tend to have high residual

stress gradients and roughened surface morphologies as a result of the highly faceted, large-grain polycrystalline films that are produced by these methods (Fig. 7.6). The rough surface morphology degrades the patterning process, resulting in roughened sidewalls in etched structures and roughened surfaces of films deposited over these layers. Unlike polysilicon and SiC, a postdeposition polishing process is not technically feasible for diamond due to its extreme hardness. For the fabrication of multilayer diamond devices, methods to reduce the surface roughness of the as-deposited films are highly desirable. Along these lines, *Krauss* et al. [7.91] have reported on the development of an ultrananocrystalline diamond (UNCD) film that exhibits a much smoother surface morphology than comparable diamond films grown using conventional methods. Unlike conventional CVD diamond films that are grown using a mixture of H_2 and CH_4, the ultrananocrystalline diamond films are grown from mixtures of Ar, H_2 and C_{60} or Ar, H_2 and CH_4. Films produced by this method have proven to be effective as conformal coatings on Si surfaces, and have been used successfully in several surface micromachining processes. This group has extended the UNCD deposition technology to low deposition temperatures, with high-quality nanocrystalline diamond films being deposited at rates of 0.2 μm/h at substrate temperatures of 400 °C, making these films compatible from a thermal budget perspective with Si IC technology [7.92]. A wide range of microdevice structures have been successfully fabricated using UNCD diamond films, including UNCD cantilevers that incorporate PZT piezoelectric actuators [7.93], suspended membranes [7.94], and microfabricated retinal implants [7.95]. In addition to achieving a smooth surface morphology, it has been shown using micromachined cantilevers that nanocrystalline and ultrananocrystalline diamond films to not suffer from the high stress gradients associated with their microcrystalline counterparts [7.96].

Another alternative deposition method that is proving to be well suited for diamond MEMS is based on pulsed laser deposition [7.97]. The process is performed in a high-vacuum chamber and uses a pulsed excimer laser to ablate a pyrolytic graphite target. Material from the ejection plume deposits on a substrate, which is kept at room temperature. Background gases composed of N_2, H_2 and Ar can be introduced to adjust the deposition pressure and film properties. The as-deposited films consist of tetrahedrally bonded carbon that is amorphous in microstructure, hence the name amorphous diamond. Nominally stress-free films can be deposited by proper selection of deposition parameters [7.98] or by a short postdeposition annealing step [7.97]. The amorphous diamond films exhibit many of the properties of single-crystal diamond, such as a high hardness (88 GPa) a high Young's modulus (1100 GPa) and chemical inertness. Many single-layer surface micromachined structures have been fabricated using these films; in part because the films can readily be deposited in oxide sacrificial layers and they can be etched in an oxygen plasma. Amorphous diamond films have been used as a dielectric isolation layer in vertically actuated microbridges in micromachined RF capacitive switches [7.99]. The diamond films sit atop W fixed electrodes and provide dielectric isolation from an Au microbridge that spans the fixed electrode structure. The diamond films are particularly attractive for such applications since the surfaces are hydrophobic and thus do not suffer from stiction and are highly resistant to wear over repeated use.

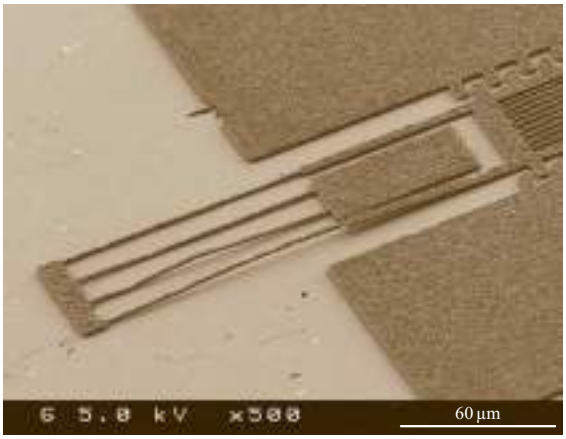


Fig. 7.6 SEM micrograph of the folded beam truss of a diamond lateral resonator. The diamond film was deposited using a seeding-based hot filament CVD process. The micrograph illustrates the challenges facing MEMS structures made from polycrystalline material, namely roughened surfaces and residual stress gradients

7.5 GaAs, InP and Related III–V Materials

Gallium arsenide (GaAs), indium phosphide (InP) and related III-V compounds have favorable piezoelectric and optoelectric properties, high piezoresistive constants and wide electronic bandgaps relative to Si, making them attractive for various sensor and optoelectronic applications. Like Si, significant research in bulk crystal growth has led to the development of GaAs and InP substrates that are commercially available as

high-quality, single-crystal wafers. Unlike compound semiconductors such as SiC, III-V materials can be deposited as ternary and quaternary alloys with lattice constants that closely match the binary compounds from which they are derived (i. e., $Al_xGa_{1-x}As$ and GaAs), thus permitting the fabrication of a wide variety of heterostructures that facilitate device performance.

Crystalline GaAs has a zinc blend crystal structure with an electronic bandgap of 1.4 eV, enabling GaAs electronic devices to function at temperatures as high as 350 °C [7.100]. High-quality, single-crystal wafers are commercially available, as are well-developed metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE) growth processes for epitaxial layers of GaAs and its alloys. GaAs does not outperform Si in terms of mechanical properties, however its stiffness and fracture toughness are still suitable for micromechanical devices.

Micromachining of GaAs is relatively straightforward, since many of its lattice-matched ternary and quaternary alloys have sufficiently different chemical properties to allow their use as sacrificial layers [7.101]. For example, the most common ternary alloy for GaAs is $Al_xGa_{1-x}As$. For values of x less than or equal to 0.5, etchants containing mixtures of HF and H_2O will etch $Al_xGa_{1-x}As$ without attacking GaAs, while etchants containing NH_4OH and H_2O_2 attack GaAs isotropically but do not etch $Al_xGa_{1-x}As$. Such selectivity enables the micromachining of GaAs wafers using lattice-matched etch stops and sacrificial layers. Devices fabricated using these methods include comb-drive lateral resonant structures [7.101], pressure sensors [7.102, 103], thermopile sensors [7.104], Fabry–Perot detectors [7.105], and cantilever-based sensors and actuators [7.106, 107]. In addition, nanoelectromechanical devices, such as suspended micromechanical resonators [7.108] and tethered membranes [7.109] have been fabricated using these techniques. An example of a nanoelectromechanical beam structure fabricated from GaAs is shown in Fig. 7.7.

In addition to using epitaxial layers as etch stops, ion implantation methods can also be used to produce etch stops in GaAs layers. *Miao* et al. [7.110] describe a process that uses electrochemical etching to selectively remove n-type GaAs layers. The process relies on the creation of a highly resistive near-surface GaAs layer on an n-type GaAs substrate by low-dose nitrogen implantation in the MeV energy range. A pulsed electrochemical etch method using a H_2PtCl_6, H_3PO_4, H_2SO_4 platinum electrolytic solution at 40 °C with 17 V, 100 ms pulses is sufficient to selectively remove n-type GaAs at about 3 µm/min. Using this method, stress-free, tethered membranes could readily be fabricated from the highly resistive GaAs layer. The high



Fig. 7.7 SEM micrograph of a GaAs nanomechanical beam resonator fabricated by epitaxial growth, electron-beam lithography and selective etching (courtesy of M. Roukes, Caltech)

implant energies enable the fabrication of membranes several microns in thickness. Moreover, the authors demonstrated that if the GaAs wafer was etched in such a way as to create an undulating surface prior to ion implantation, corrugated membranes could be fabricated. These structures can sustain much higher deflection amplitudes than flat structures.

Micromachining of InP closely resembles the techniques used for GaAs. Many of the properties of InP are similar to GaAs in terms of crystal structure, mechanical stiffness and hardness; however, the optical properties of InP make it particularly attractive for micro-optomechanical devices to be used in the 1.3−1.55 µm wavelength range [7.111]. Like GaAs, single-crystal wafers of InP are readily available and ternary and quaternary lattice-matched alloys, such as InGaAs, InAlAs, InGaAsP and InGaAlAs, can be used as either etch stop and/or sacrificial layers depending on the etch chemistry [7.101]. For instance, InP structural layers deposited on $In_{0.53}Al_{0.47}As$ sacrificial layers can be released using etchants containing $C_6H_8O_7$, H_2O_2, and H_2O. In addition, InP films and substrates can be etched in solutions containing HCl and H_2O using $In_{0.53}Ga_{0.47}As$ films as etch stops. Using InP-based micromachining techniques, multi-air gap filters [7.112], bridge structures [7.111], and torsional membranes [7.105] have been fabricated from InP and its related alloys.

In addition to GaAs and InP, materials such as indium arsenide (InAs) can be micromachined into device structures. Despite a 7% lattice mismatch between InAs and (111) GaAs, high-quality epitaxial layers can be grown on GaAs substrates. As described by *Yamaguchi*

et al. [7.113], the surface Fermi level of InAs/GaAs structures is pinned in the conduction band, enabling the fabrication of very thin conductive membranes. In fact the authors have successfully fabricated free-standing InAs structures that range in thickness from 30 to 300 nm. The thin InAs films were grown directly on GaAs substrates by MBE and etched using a solution containing H_2O, H_2O_2, and H_2SO_4. The structures, mainly doubly clamped cantilevers, were released by etching the GaAs substrate using a $H_2O/H_2O_2/NH_4OH$ solution.

Ternary III-V compounds have also been utilized in MEMS structures. *Deng* et al. report on the development of disk-type microresonators based on epitaxial AlGaAs films [7.114]. The resonator consists of a surface micromachined disk that is tethered to the substrate by four supporting beams. The disk is made from a Si-doped $Al_{0.3}Ga_{0.7}As$ layer that clads an undoped $Al_{0.3}Ga_{0.7}As$ layer. This three-layer stack is grown on a sacrificial $Al_{0.7}Ga_{0.3}As$ layer that is epitaxially grown on a commercially available GaAs wafer. Reactive ion etching is used to pattern the disk resonator and the structure is released by dissolving the sacrificial layer in HF, which is highly selective to the Al-rich AlGaAs layer. The device capitalizes on the piezoelectric properties of AlGaAs for actuation, which is realized in the suspended disk due to the excellent lattice matching between each epitaxial layer.

7.6 Piezoelectric Materials

7.6.1 Lead Zirconate Titanate

Piezoelectric materials play an important role in MEMS technology for sensing and mechanical actuation applications. In a piezoelectric material, mechanical stress produces a polarization, and conversely a voltage-induced polarization produces a mechanical stress. Many asymmetric materials, such as quartz, GaAs, and zinc oxide (ZnO), exhibit varying levels of piezoelectric behavior. Recent work in MEMS has focused on the development of ferroelectric compounds such as lead zirconate titanate, $Pb(Zr_xTi_{1-x})O_3$, or PZT for short, because such compounds have high piezoelectric constants that result in high mechanical transduction. It is relatively straightforward to fabricate a PZT structure on top of a thin free-standing structural layer (i. e., cantilever, diaphragm). Such a capability enables the piezoelectric material to be used in sensor applications or actuator applications where piezoelectric materials are particularly well suited. Like Si, PZT films can be patterned using dry etch techniques based on chlorine chemistries, such as Cl_2/CCl_4, as well as ion beam milling using inert gases like Ar.

PZT has been successfully deposited in thin-film form using cosputtering, CVD, and sol-gel processing. Sol-gel processing is particularly attractive because the composition and homogeneity of the deposited material over large surface areas can be readily controlled. The sol-gel process outlined by *Lee* et al. [7.115] uses PZT solutions made from liquid precursors containing Pb, Ti, Zr, and O. The solution is deposited by spin coating on a Si wafer that has been coated with a $Pt/Ti/SiO_2$ thin-film multilayer. The process is executed to produce a PZT film in layers, with each layer consisting of a spin-coated layer that is dried at 110 °C for 5 min and then heat treated at 600 °C for 20 min. After building up the PZT layer to the desired thickness, the multilayer was heated at 600 °C for up to 6 h. Prior to this anneal, a PbO top layer was deposited on the PZT surface. An Au/Cr electrode was then sputter deposited on the surface of the piezoelectric stack. This process was used to fabricate a PZT-based force sensor. *Xu* et al. [7.116] describe a similar sol-gel process to produce 12 μm-thick, crack-free PZT films on Pt-coated Si wafers and 5 μm-thick films on insulating ZrO_2 layers to produce micromachined MHz-range two-dimensional transducer arrays for acoustic imaging. A publication by *Hong* et al. describes the development of large-deflection circular diaphragm actuators made using a solution-cast PZT film [7.117]. The PZT film was deposited by spin casting onto a sol-gel ZrO_2 film deposited on a SiO_2-coated Si wafer. A sputter-deposited Pt film was patterned into interdigitated electrodes on the PZT film and the diaphragm was created by bulk micromachining the silicon dioxide-coated substrate by DRIE. The PZT film had a tensile residual stress of 50 to 100 MPa. For a ≈ 3 μm-thick, 800 μm-diameter membrane, a center deflection of ≈ 7 μm could be generated with an applied voltage of 100 V.

Thick-film printing techniques for PZT have been developed to produce thick films in excess of 100 μm. Such thicknesses are desired for applications that require actuation forces that cannot be achieved with the much thinner sol-gel films. *Beeby* et al. [7.118] describe a thick-film printing process whereby a PZT paste is made from a mixture of 95% PZT powder, 5% lead borosilicate powder and an organic carrier. The paste was then printed through a stainless steel screen using a thick-film printer. Printing was performed on an oxidized Si substrate that is capped with a Pt elec-

trode. After printing, the paste was dried and then fired at 850–950 °C. Printing could be repeated to achieve the desired thickness. The top electrode consisted of an evaporated Al film. The authors found that it was possible to perform plasma-based processing on the printed substrates, but that the porous nature of the printed PZT films made them unsuitable for wet chemical processing.

7.6.2 Aluminum Nitride

Aluminum nitride (AlN) has emerged as an alternative to PZT in MEMS and NEMS applications owing to its favorable piezoelectric, mechanical and thermal properties combined with its process compatibility with other MEMS materials, in particular silicon, but also other MEMS materials like SiC and GaN. To produce thin films of AlN suitable for MEMS and NEMS, both chemical and physical vapor deposition techniques can be used, with the two most common deposition techniques being metalorganic chemical vapor deposition (MOCVD) and magnetron sputtering. MOCVD is the preferred method to produce single-crystalline and high-quality polycrystalline films. The most commonly used precursor gases are trimethylaluminum (TMA) and ammonium (NH_3). Substrate temperatures are typically in excess of 1000 °C. Under these conditions, high-quality crystalline films can be produced on (111) Si substrates. In fact, one of the most common uses of AlN thin films is as a transition layer for epitaxial growth of GaN on Si, however the excessively high growth temperatures associated with AlN growth by MOCVD inhibit integration with prefabricated Si CMOS devices.

Fortunately for MEMS and NEMS applications, the piezoelectric properties of micro- and nanocrystalline AlN films produced by magnetron sputtering are sufficient for many sensor/actuator applications. *Tonish* et al. compared the d_{33} piezoelectric coefficient of AlN films deposited on (111) Si by both magnetron sputtering and MOCVD [7.119]. The sputtered films were produced by reactive sputtering of an Al target in a nitrogen environment with the substrate temperature ranging from 350 to 500 °C. The MOCVD films were grown using TMA and NH_3 at substrate temperatures of 1050 to 1190 °C. Films ranged in thickness between 130 and 250 nm. The sputtered films exhibited a nanocrystalline microstructure with a grain size of $\approx$ 20 nm while the MOCVD films were polycrystalline with a grain size of 100 nm. Of the two deposition methods, the sputtered films exhibited the highest d_{33} coefficient ($\approx$ 5 pm/V), which was attributed to the favorable microstructure of these films in supporting the piezoelectric effect in spite of the small grain size. The low substrate temperatures associated with the sputter deposition of AlN make it compatible for use as a piezoelectric actuator on structural components made from polysilicon [7.120]. Bulge testing of bulk micromachined diaphragms made from 600 nm-thick AlN films deposited by magnetron sputtering on a silicon substrate yielded a Young's modulus of $\approx$ 250 GPa and a residual stress of 350 MPa; values that are similar to other MEMS materials [7.121]. These diaphragms were fabricated by bulk micromachining the silicon substrate by conventional deep reactive ion etching.

Piazza et al. have exploited the favorable piezoelectric properties of sputter-deposited AlN films to develop a new class of vibrating contour-mode MEMS resonators [7.122]. These MHz-frequency rectangular and ring-shaped resonators consist of structural 2 µm-thick AlN films that are sandwiched between a bottom Pt electrode and a top Al electrode. These devices are fabricated on Si_3N_4-coated Si wafers by conventional sputter deposition, photolithographic patterning and chemical etching. Large area contacts for the bottom Pt electrode are etched through the AlN layer using a phosphoric acid etch at 160 °C, whereas the AlN resonators are fabricated simultaneous to the pattering of the top Al electrode in a self-aligned process using a Cl_2-based RIE etch. The device is released by an isotropic dry etch of the underlying Si in XeF_2. Testing in air revealed quality factors of 4300 and exceptionally low motional resistances ($< 700\,\Omega$), at the time the lowest ever reported for contour-mode MEMS devices.

Sputter-deposited AlN retains its favorable piezoelectric properties when scaled down to the NEMS regime. *Sinha* et al. demonstrated NEMS-based bimorph actuators based on nanometer-scale AlN films [7.123]. These NEMS-based cantilevers were comprised of a multimaterial stack consisting of two 100 nm-thick AlN layers sandwiched between three 50 nm Pt electrode layers. All the layers were deposited by physical vapor deposition, with the AlN layers deposited by alternating current (AC) reactive sputtering. To control the residual stresses in the film in situ, substrate heating during a portion of each AlN deposition was used. Cantilevers ranging in length from 5 to 18 µm were fabricated using an isotropic silicon etchant for release of doubly clamped bridges followed by focused ion beam etching to cut away one of the bridge anchors. The three-electrode configuration allowed for actuation voltages to be applied independently to each AlN layer, thus enabling bimorph-type actuation due to differing voltage-related displacements in the two piezoelectric AlN films. Vertical displacements as large as 40 nm at 2 V were realized from these devices.

Karabalin et al. also developed NEMS-based devices from sputter-deposited AlN films [7.124]. Their

devices were NEMS-based cantilevers and doubly clamped bridges consisting of a single 100 nm AlN film sandwiched between two 100 nm Mo thin-film electrodes. A thin (20 nm) AlN seed layer was sputter deposited onto the (100) Si wafer prior to the deposition of the bottom Mo electrode to ensure that the Mo film was of proper microstructure to produce highly *c*-axis AlN films. A multistep electron beam-based photolithography and etch process was used to pattern the thin-film stack into cantilevers that were nominally 6 μm in length, 900 nm in width and 320 nm in thickness. An anisotropic Ar/Cl_2 inductively couple plasma etch was used to pattern the AlN and Mo films, while an isotropic Ar/NF_3 reactive ion etch was used to selectively remove the underlying silicon and release the structures. The resonators demonstrated fundamental flexural modes up to 80 MHz, which were enabled by the AlN layers retaining a d_{31} coefficient of 2.4 pm/V.

7.6.3 Gallium Nitride

Gallium nitride is an attractive material for MEMS applications because it is a wide bandgap semiconductor ($E_g = 3.4$ eV) that also possesses piezoelectric properties that while not as strong as PZT and AlN are useful in MEMS. GaN is a binary semiconductor that belongs to a family of III-N ternary compound semiconductors of similar lattice constant but differing bandgaps, thus enabling the fabrication of single-crystal heterostructures that cannot be realized in SiC and diamond. These heterostructures enable the direct integration of GaN MEMS with AlGaN/GaN high electron mobility transistors (HEMTs) that can be used for sensing the mechanical motion of the MEMS structure in a manner that may be more straightforward than similar structures in Si. Although wafer-scale bulk GaN is not currently available, high-quality epitaxial films can readily be grown on sapphire, 4H-SiC, and Si wafers. As such, GaN has become an extremely popular semiconductor for high-frequency and optoelectronic devices.

Like AlN, the most popular method to produce device-quality GaN is by MOCVD with the most common precursors being trimethylgalium and ammonia and substrate temperatures in the neighborhood of 1000 °C. For MEMS applications, Si is the most convenient substrate owing to the compatibility of GaN to Si micromachining. Like SiC, GaN is chemically stable, which translates to a lack of viable wet chemical etchants for surface and bulk micromachining. *Ramesh* et al. describe the use of electrochemical etching techniques for the release of GaN devices [7.125]. In this case, a p-type GaN layer was grown on a sacrificial InGaN layer and photoelectrochemical etching of the sacrificial layer was performed in a KOH solution while illuminating the structure with UV radiation at a wavelength below the bandgap of GaN but above that of InGaN. Similarly, a bias-assisted dark electrochemical etch was developed to selectively remove p-type GaN. This etch requires a relatively large bias voltage (5 V) and is highly isotropic, therefore requiring proper design of anchor regions.

Like nearly all MEMS structural materials, GaN can be patterned by plasma etching. *Lv* et al. describe a surface micromachining process to fabricate GaN-based accelerometers [7.126]. The device actually consists of a 20 nm-thick AlGaN, 1 μm-thick GaN heterostructure that serves as both the structural element and sensing component, with the interface of the epitaxial layers serving as a two-dimensional (2-D) electron gas that is sensitive to the displacement of the structure; essentially a piezoresistive effect based on the piezoelectric polarization at this interface. The AlGaN/GaN layers are first grown onto a (111) Si substrate by MOCVD. The AlGaN and GaN layers are etched using inductively coupled plasma etching using Cl_2 gas, with the thinner AlGaN layer etched at a lower power than the thicker GaN film. A Ti/Al/Ni/Au stack was used for ohmic contacts. The device was released by first etching the underlying silicon using an anisotropic $SF_6/O_2/C_4F_8$ deep reactive ion etching process followed by an isotropic SF_6/O_2 etch. Assessment of residual stress by Raman spectroscopy in released and unreleased sections of the device showed a reduction in stress by 87% after release. Other plasma-based etches that are effective in patterning GaN include mixes of Cl_2 and BCl_3 mixed with Ar [7.127].

Like almost all of the other semiconductor materials used in MEMS, the mechanical properties of GaN thin films has been characterized using micromachined structures. *Amar* et al. used vibrating micromachined cantilevers comprised of AlN/GaN/AlGaN multilayers to experimentally determine the Young's modulus of GaN and AlN from resonant frequency data [7.128]. They found that the Young's modulus of their GaN to be ≈ 260 GPa while that for AlN was ≈ 340 GPa; values that compare favorably with measurements made by nanoindentation. *Zimmerman* et al. also characterized AlGaN and GaN cantilevers by static load-displacement and found the Young's modulus of GaN to be ≈ 250 GPa [7.129]. Testing to fracture yielded a fracture strength of 350 MPa.

7.7 Polymer Materials

7.7.1 Polyimide

Polyimides comprise an important class of durable polymers that are well suited for many of the techniques used in conventional MEMS processing. In general, polyimides can be acquired in bulk or deposited as thin films by spin coating. Polyimides can be patterned using conventional dry etching techniques and processed at relatively high temperatures. These attributes make polyimides an attractive group of polymers for MEMS that require polymer structural and/or substrate layers, such as microfabricated biomedical devices where inertness and flexibility are important parameters.

Shearwood et al. [7.130] explored the use of polyimide as a robust mechanical material for microfabricated audio membranes. The authors fabricated 7 μm-thick, 8 mm-diameter membranes on GaAs substrates by bulk micromachining the GaAs substrate using a NH_3/H_2O_2 solution. They realized 100% yield and despite a low Young's modulus (≈ 3 GPa), observed flat membranes to within 1 nm after fabrication.

Jiang et al. [7.131] capitalized on the strength and flexibility of polyimide to fabricate a flexible sheer-stress sensor array based on Si sensors. The sensor array consisted of a collection of Si islands linked by two polyimide layers. Each Si sensor island was 250 × 250 μm^2 in area and 80 μm in thickness. Al was used as an electrical interconnect layer. The two polyimide layers served as highly flexible hinges giving the sensor array the ability to be mounted on curved surfaces. The sensor array was successful in profiling the shear-stress distribution along the leading edge of a rounded delta wing.

Yousef et al. report on the development of MEMS-based thermopiles made from micromachined polyimide substrates [7.132]. The thermopiles consisted of vertically oriented Sb and Ni pillars traversing the thickness of the polyimide substrate. The pillars, known as thermolegs, are formed by electroplating in vias micromachined into the substrate. The vias were micromachined by selectively irradiating the polyimide substrate with 8.3 MeV Xe ions and etching the ion-damaged regions in a sodium hypochlorite solution.

The chemical and temperature durability of polyimides enables their use as a sacrificial layer for a number of commonly used materials, such as evaporated or sputter-deposited metals. *Memmi* et al. [7.133] developed a fabrication process for capacitive micromechanical ultrasonic transducers using polyimide as a sacrificial layer. The authors showed that the polyimide could withstand the conditions used to deposit silicon monoxide by evaporation and silicon nitride by PECVD at 400 °C. Recent work by *Bagolini* et al. [7.134] has shown that polyimides can even be used as sacrificial layers for micromachined devices made from PECVD SiC thin films.

In the area of microfabricated biomedical devices, polyimide is receiving attention as a substrate material for implantable devices, owing to its potential biocompatiblity and mechanical flexibility. *Stieglitz* [7.135] reported on the fabrication of multichannel microelectrodes on polyimide substrates. Instead of using polyimide sheets as starting substrates, Si carrier wafers coated with a 5 μm-thick polyimide film were used. Pt microelectrodes were then fabricated on these substrates using conventional techniques. Thin polyimide layers were deposited between various metal layers to serve as insulating layers. A capping polyimide layer was then deposited on the top of the substrates, and then the entire polyimide/metal structure was peeled off the Si carrier wafers. Backside processing was then performed on the free-standing polyimide structures to create devices that have exposed electrodes on both surfaces. In later papers, *Stieglitz* et al. [7.136, 137] describe variations of this process for advanced polyimide-based neural prostheses.

7.7.2 SU-8

SU-8 is a negative-tone epoxy-like photoresist that is receiving much attention for its versatility in MEMS processing. It is a high-aspect-ratio, UV-sensitive resist designed for applications requiring single-coat resists with thicknesses on the order of 500 μm [7.138]. SU-8 has favorable chemical properties, which enable it to be used as a molding material for high-aspect-ratio electroplated structures (as an alternative to LIGA), and as a structural material for microfluidics [7.138]. In terms of mechanical properties, *Lorenz* et al. [7.139] reported that SU-8 has a modulus of elasticity of 4.02 GPa, which compares favorably with a commonly used polyamide (3.4 GPa).

In addition to the above-mentioned conventional uses for SU-8, several interesting alternative uses are beginning to appear in the literature. *Conradie* et al. [7.140] have used SU-8 to trim the mass of silicon paddle oscillators as a means to adjust the resonant frequency of the beams. The trimming process involves the patterning of SU-8 posts on Si paddles. The process capitalizes on the relative chemical stability of the SU-8 resin in conjunction with the relatively large masses that can be patterned using standard UV exposure processes.

SU-8 is also of interest as a bonding layer material for wafer bonding processes using patterned bonding

layers. *Pan* et al. [7.141] compared several UV photodefinable polymeric materials and found that SU-8 exhibited the highest bonding strength (20.6 MPa) for layer thicknesses up to 100 μm.

7.7.3 Parylene

Parylene (poly-paraxylylene) is another emerging polymeric MEMS material due in large part to its biocompatibility. It is particularly attractive from the fabrication perspective because it can be deposited by CVD at room temperature. Moreover the deposition process is conformal, which enables parylene coatings to be applied to prefabricated structures, such as Si microneedles [7.142], low-stress silicon nitride membrane particle filters [7.143], and micromachined polyimide/Au optical scanners [7.144]. In the former cases, the parylene coating served to strengthen the microfabricated structures, while in the latter case it served to protect the structure from condensing water vapor.

In addition to its function as a protective coating, parylene can actually be bulk micromachined into free-standing components. *Noh* et al. [7.145] demonstrated a method to create bulk micromachined parylene microcolumns for miniature gas chromatographs. The structure is fabricated using a micromolding technique where Si molds are fabricated by DRIE and coated with parylene to form three sides of the microcolumn. A second wafer is coated with parylene and the two are bonded together via a fusion bonding process. After bonding, the structure is released from the Si mold by KOH etching. In a second example, *Suzuki* et al. describe a process to fabricate high-aspect-ratio parylene springs for use in low-frequency accelerometers [7.146]. The process involves the fabrication of deep trench etching into a Si substrate in locations where the parylene springs are desired. The trench walls are coated with SiO_2 by thermal oxidation. Parylene is then deposited onto the wafer, filling the trenches and coating the field areas. Unwanted parylene in the field is removed by O_2 plasma etching, leaving behind isolated parylene structures embedded into the Si wafer. Si DRIE is then used from the backside to free up the parylene springs and release the Si proof mass. The paper reports the fabrication and successful testing of a prototype device with a resonant frequency of only 37 Hz.

Like many vapor deposited materials, parylene structures can be fabricated by surface micromachining. *Yao* et al. [7.147] describe a dry release process for parylene surface micromachining. In this process, sputtered Si is used as a sacrificial layer onto which a thick sacrificial photoresist is deposited. Parylene is then deposited on the photoresist and patterned into the desired structural shape. The release procedure is a two-step process. First the photoresist is dissolved in acetone. This results in the parylene structure sticking to the sputtered Si. Next, a dry BrF_3 etch is performed, which dissolves the Si and releases the parylene structures. Parylene beams that were 1 mm long and 4.5 μm thick were successfully fabricated using this technique. The use of photoresist as a sacrificial layer for parylene surface micromachining has since been extended to fabricate parylene-based valves for microfluidic applications [7.148].

Parylene is a United States Pharmacopoeia Class IV material that is commonly used for biomedical device packaging. This attribute, combined with the fact that parylene can be deposited by highly conformal CVD, makes parylene an ideal candidate for micromachined neural prostheses. *Rodger* et al. have developed retinal prostheses based on parylene [7.149]. Their basic structure consists of parylene base and capping layers that sandwich a Pt multi-electrode (i. e., 256) array. The total thickness of the parylene layers is ≈ 13 μm, giving the structure a very high degree of mechanical flexibility. Access to the Pt electrodes is easily made by RIE. Due in part to the highly conformal nature of the coating process combined with the low substrate temperatures required for film growth, prepackaged application specific integrated circuits (ASICS) can be incorporated into the electrode structure. Devices are fabricated on Si handle wafers using photoresist as a sacrificial release layer.

Ziegler et al. describe a method to fabricate neural probes with built-in microfluidic channels for drug delivery [7.150]. The device is fabricated from two parylene layers, a base layer that supports the metal electrodes and a capping layer that forms the top of the fluidic channel and covers the electrode traces. The base layer is fabricated on a Cr/Al-coated Si wafer while the capping layer is formed on a detergent-coated glass substrate. The key processing step is the bonding step, which joins the base and capping layers forming a leak-free seal. This processing step is performed in a standard convection oven at 160 °C over 30 min with an applied pressure of only 4.9 MPa. After bonding the glass wafer is simply removed due to the weak glass-to-parylene bond as a result of the detergent coating. The Si wafer is then removed by anodic aluminum etching in a NaCl solution. Undoubtedly, the bonding process is greatly enhanced by the fact that parylene can be vapor deposited in extremely smooth Si wafers.

7.7.4 Liquid Crystal Polymer

Liquid crystal polymer (LCP) is a high-performance thermoplastic currently being used in printed circuit board and electronics packaging applications, and has

been investigated for use in MEMS applications requiring a material that is mechanically flexible, electrically insulating, chemically durable and impermeable to moisture. LCP can be bonded to itself and other substrate materials such as glass and Si by thermal lamination. It can be micromachined using an oxygen plasma and yet it is highly resistant to HF and many metal etchants [7.151]. The moisture absorption is less than 0.02% as compared with about 1% for polyimide [7.152], making it well suited as a packaging material.

Applications where LCPs have been used as a key component in a MEMS device are beginning to emerge. *Faheem* et al. [7.153] reported on the use of LCP for encapsulation of variable RF MEMS capacitors. In this example, LCP, dispensed in liquid form, was used to join and seal a glass microcap to a prefabricated, microbridge capacitor. LCP was chosen in part because in addition to the aforementioned properties, it has very low RF loss characteristics, making it very well suited as an RF MEMS packaging material. *Wang* et al. [7.151] showed that LCP is a very versatile material that is highly compatible with many standard Si-based processing techniques, and that these techniques can be used to make LCP cantilever flow sensors that incorporate metal strain gages and LCP membrane tactile sensors using NiCr strain gages. *Lee* et al. [7.152] have developed a mechanically flexible, multichannel microelectrode array structure for neural stimulation and recording using LCP as a base substrate and *Palasagaram* et al. have developed a capacitive pressure sensor from LCP [7.154]. The capacitive pressure sensor was fabricated by thermally laminating three layers of LCP. The bottom layer consisted of 30 mil-thick LCP coated 0.5 μm-thick Cu that was patterned into an electrode. The middle layer consisted of an unmetalized, 2 mil-thick LCP layer that was patterned with a through hole that forms the pressure sensor cavity. The pressure sensor membrane was formed from a 2 mil-thick LCP layer that supports a 0.5 μm-thick Cu electrode. The sensor membrane was roughly 1.6 mm in diameter. This sensor exhibited a relative capacitance change of 0.277 pF over an applied pressure range of 0–100 kPa.

7.8 Future Trends

The rapid expansion of MEMS in recent years is due in large part to the inclusion of new materials that have expanded the functionality of microfabricated devices beyond what is achievable in silicon. This trend will certainly continue as new application areas for micro- and nanofabricated devices are identified. Many of these applications will likely require both new materials and new processes to fabricate the micro- and nanomachined devices for these yet-to-be-identified applications. Currently, conventional micromachining techniques employ a *top-down* approach that begins with either bulk substrates or thin films. Future MEMS and NEMS will likely incorporate materials that are created using a *bottom-up* approach. A significant challenge facing device design and fabrication engineers alike will be how to marry top-down and bottom-up approaches to create devices and systems that cannot be made using either process alone.

Acknowledgments. Prof. Zorman thanks Prof. Mehran Mehregany of Case Western Reserve University for his numerous contributions to this manuscript.

References

7.1 C.S. Smith: Piezoresistive effect in germanium and silicon, Phys. Rev. **94**, 1–10 (1954)

7.2 A.N. Cleland, M.L. Roukes: Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals, Appl. Phys. Lett. **69**, 2653–2655 (1996)

7.3 D.W. Carr, H.G. Craighead: Fabrication of nanoelectromechanical systems in single-crystal silicon using silicon on insulator substrates and electron beam lithography, J. Vacuum Sci. Technol. B **15**, 2760–2763 (1997)

7.4 T. Kamins: *Polycrystalline Silicon for Integrated Circuits and Displays*, 2nd edn. (Kluwer Academic, Boston 1988)

7.5 J.J. McMahon, J.M. Melzak, C.A. Zorman, J. Chung, M. Mehregany: Deposition and characterization of in-situ boron doped polycrystalline silicon films for microelectromechanical systems applications, Mater. Res. Symp. Proc. **605**, 31–36 (2000)

7.6 L. Cao, T.S. Kin, S.C. Mantell, D. Polla: Simulation and fabrication of piezoresistive membrane type MEMS strain sensors, Sens. Actuat. **80**, 273–279 (2000)

7.7 H. Guckel, T. Randazzo, D.W. Burns: A simple technique for the determination of mechanical strain in thin films with application to polysilicon, J. Appl.

Phys. **57**, 1671–1675 (1983)
7.8 R.T. Howe, R.S. Muller: Stress in polysilicon and amorphous silicon thin films, J. Appl. Phys. **54**, 4674–4675 (1983)
7.9 X. Zhang, T.Y. Zhang, M. Wong, Y. Zohar: Rapid thermal annealing of polysilicon thin films, J. Microelectromech. Syst. **7**, 356–364 (1998)
7.10 J. Yang, H. Kahn, A.-Q. He, S.M. Phillips, A.H. Heuer: A new technique for producing large-area as-deposited zero-stress LPCVD polysilicon films: The multipoly process, J. Microelectromech. Syst. **9**, 485–494 (2000)
7.11 T.J. Kang, H.Y. Lee, Y.H. Kim: Reduction of sheet resistance and low-thermal budget relaxation of stress gradients in polysilicon microcantilever beams using nickel-silicides, J. Microelectromech. Syst. **16**, 279–288 (2007)
7.12 P. Gennissen, M. Bartek, P.J. French, P.M. Sarro: Bipolar-compatible epitaxial poly for smart sensors: Stress minimization and applications, Sens. Actuat. **A62**, 636–645 (1997)
7.13 P. Lange, M. Kirsten, W. Riethmuller, B. Wenk, G. Zwicker, J.R. Morante, F. Ericson, J.A. Schweitz: Thick polycrystalline silicon for surface-micromechanical applications: Deposition, structuring, and mechanical characterization, Sens. Actuat. **A54**, 674–678 (1996)
7.14 S. Greek, F. Ericson, S. Johansson, M. Furtsch, A. Rump: Mechanical characterization of thick polysilicon films: Young's modulus and fracture strength evaluated with microstructures, J. Micromech. Microeng. **9**, 245–251 (1999)
7.15 K. Funk, H. Emmerich, A. Schilp, M. Offenberg, R. Neul, F. Larmer: A surface micromachined silicon gyroscope using a thick polysilicon layer. In: *Proc. 12th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 1999) pp. 57–60
7.16 T. Abe, M.L. Reed: Low strain sputtered polysilicon for micromechanical structures. In: *Proc. 9th Int. Workshop Microelectromech. Syst.* (IEEE, Piscataway 1996) pp. 258–262
7.17 K. Honer, G.T.A. Kovacs: Integration of sputtered silicon microstructures with pre-fabricated CMOS circuitry, Sens. Actuat. A **91**, 392–403 (2001)
7.18 J. Gaspar, T. Adrega, V. Chu, J.P. Conde: Thin-film paddle microresonators with high quality factors fabricated at temperatures below 110°C. In: *Proc. 18th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 2005) pp. 125–128
7.19 S.B. Patil, V. Chu, J.P. Conde: Surface micromachining of a thin film microresonator using dry decomposition of a polymer sacrificial layer, J. Vacuum Sci. Technol. B **25**, 455–458 (2007)
7.20 R. Anderson, R.S. Muller, C.W. Tobias: Porous polycrystalline silicon: A new material for MEMS, J. Microelectromech. Syst. **3**, 10–18 (1994)
7.21 W. Lang, P. Steiner, H. Sandmaier: Porous silicon: A novel material for microsystems, Sens. Actuat. A **51**, 31–36 (1995)
7.22 R. He, C.J. Kim: On-chip hermetic packaging enabled by post-deposition electrochemical etching of polysilicon. In: *Proc. 18th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 2005) pp. 544–547
7.23 R. He, C.J. Kim: On-wafer monolithic encapsulation by surface micromachining with porous polysilicon shell, J. Microelectromech. Syst. **16**, 462–472 (2007)
7.24 S.K. Ghandhi: *VLSI Fabrication Principles – Silicon and Gallium Arsenide* (Wiley, New York 1983)
7.25 W.A. Pilskin: Comparison of properties of dielectric films deposited by various methods, J. Vacuum Sci. Technol. **21**, 1064–1081 (1977)
7.26 J.S. Danel, F. Michel, G. Delapierre: Micromachining of quartz and its application to an acceleration sensor, Sens. Actuat. A **21/23**, 971–977 (1990)
7.27 A. Yasseen, J.D. Cawley, M. Mehregany: Thick glass film technology for polysilicon surface micromachining, J. Microelectromech. Syst. **8**, 172–179 (1999)
7.28 R. Liu, M.J. Vasile, D.J. Beebe: The fabrication of nonplanar spin-on glass microstructures, J. Microelectromech. Syst. **8**, 146–151 (1999)
7.29 B.A. Walmsley, Y.L. Liu, X.Z. Hu, M.B. Bush, J.M. Dell, L. Faraone: Poisson's ratio of low-temperature PECVD silicon nitride thin films, J. Microelectromech. Syst. **16**, 622–627 (2007)
7.30 B. Folkmer, P. Steiner, W. Lang: Silicon nitride membrane sensors with monocrystalline transducers, Sens. Actuat. A **51**, 71–75 (1995)
7.31 M. Sekimoto, H. Yoshihara, T. Ohkubo: Silicon nitride single-layer X-ray mask, J. Vacuum Sci. Technol. **21**, 1017–1021 (1982)
7.32 P.P. Tsai, I.-C. Chen, C.J. Ho: Ultralow power carbon monoxide microsensor by micromachining techniques, Sens. Actuat. B **76**, 380–387 (2001)
7.33 P.J. French, P.M. Sarro, R. Mallee, E.J.M. Fakkeldij, R.F. Wolffenbuttel: Optimization of a low-stress silicon nitride process for surface micromachining applications, Sens. Actuat. A **58**, 149–157 (1997)
7.34 B. Li, B. Xiong, L. Jiang, Y. Zohar, M. Wong: Germanium as a versatile material for low-temperature micromachining, J. Microelectromech. Syst. **8**, 366–372 (1999)
7.35 A. Franke, D. Bilic, D.T. Chang, P.T. Jones, T.J. King, R.T. Howe, C.G. Johnson: Post-CMOS integration of germanium microstructures. In: *Proc. 12th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 1999) pp. 630–637
7.36 A.E. Franke, Y. Jiao, M.T. Wu, T.J. King, R.T. Howe: Post-CMOS modular integration of poly-SiGe microstructures using poly-Ge sacraficial layers. In: *Tech. Digest – Solid State Sens. Actuat. Workshop* (Transducers Research Foundation, Cleveland 2000) pp. 18–21
7.37 S. Sedky, P. Fiorini, M. Caymax, S. Loreti, K. Baert, L. Hermans, R. Mertens: Structural and mechanical properties of polycrystalline silicon germanium for micromachining applications, J. Microelectromech. Syst. **7**, 365–372 (1998)
7.38 S. Sedky, A. Witvrouw, K. Baert: Poly SiGe, a promising material for mems monolithic integration with the driving electronics, Sens. Actuat. A **97/98**, 503–511 (2002)
7.39 C.W. Low, T.J.K. Liu, R. Howe: Characterization of polycrystalline silicon-germanium film deposition for modularly integrated MEMS applications, J. Mi-

croelectromech. Syst. **16**, 68–77 (2007)
7.40 S. Sedky, A. Bayoumy, A. Alaa, A. Nagy, A. Witvrouw: Optimal conditions for micromachining $Si_{1-x}Ge_x$ at 210°C, J. Microelectromech. Syst. **16**, 581–588 (2007)
7.41 J.M. Heck, C.G. Keller, A.E. Franke, L. Muller, T.-J. King, R.T. Howe: High aspect ratio polysilicon-germanium microstructures. In: *Proc. 10th Int. Conf. Solid State Sens. Actuat.* (Institute of Electrical Engineers of Japan, Tokyo 1999) pp. 328–334
7.42 P. Van Gerwen, T. Slater, J.B. Chevrier, K. Baert, R. Mertens: Thin-film boron-doped polycrystalline $silicon_{70\%}$-$Germanium_{30\%}$ for Thermopiles, Sens. Actuat. A **53**, 325–329 (1996)
7.43 D. Hyman, J. Lam, B. Warneke, A. Schmitz, T.Y. Hsu, J. Brown, J. Shaffner, A. Walson, R.Y. Loo, M. Mehregany, J. Lee: Surface micromachined RF MEMS switches on GaAs substrates, Int. J. Radio Freq. Microw. Commun. Eng. **9**, 348–361 (1999)
7.44 C. Chang, P. Chang: Innovative micromachined microwave switch with very low insertion loss, Sens. Actuat. **79**, 71–75 (2000)
7.45 A. Reddy, H. Kahn, A.H. Heuer: A MEMS-based evaluation of the mechanical properties of metallic thin films, J. Microelectromech. Syst. **16**, 650–658 (2007)
7.46 M.F. Aimi, M.P. Rao, N.C. MacDonald, A.S. Zuruzi, D.P. Bothman: High-aspect-ratio bulk micromachining of Ti, Nat. Mater. **3**, 103–105 (2004)
7.47 E.R. Parker, M.P. Rao, K.L. Turner, C.D. Meinhart, N.C. MacDonald: Bulk micromachined titanium microneedles, J. Microelectromech. Syst. **16**, 289–295 (2007)
7.48 C.L. Shih, B.K. Lai, H. Kahn, S.M. Phillips, A.H. Heuer: A robust co-sputtering fabrication procedure for TiNi shape memory alloys for MEMS, J. Microelectromech. Syst. **10**, 69–79 (2001)
7.49 G. Hahm, H. Kahn, S.M. Phillips, A.H. Heuer: Fully microfabricated silicon spring biased shape memory actuated microvalve. In: *Proc. Solid State Sens. Actuat. Workshop* (Transducers Research Foundation, San Diego 2000) pp. 230–233
7.50 S.D. Leith, D.T. Schwartz: High-rate through-mold electrodeposition of thick (> 200 micron) NiFe MEMS components with uniform composition, J. Microelectromech. Syst. **8**, 384–392 (1999)
7.51 N. Rajan, M. Mehregany, C.A. Zorman, S. Stefanescu, T. Kicher: Fabrication and testing of micromachined silicon carbide and nickel fuel atomizers for gas turbine engines, J. Microelectromech. Syst. **8**, 251–257 (1999)
7.52 T. Pornsin-Sirirak, Y.C. Tai, H. Nassef, C.M. Ho: Titanium-alloy MEMS wing technology for a microaerial vehicle application, Sens. Actuat. A **89**, 95–103 (2001)
7.53 C.R. Stoldt, C. Carraro, W.R. Ashurst, D. Gao, R.T. Howe, R. Maboudian: A low temperature CVD process for silicon carbide MEMS, Sens. Actuat. A **97/98**, 410–415 (2002)
7.54 M. Eickhoff, H. Moller, G. Kroetz, J. von Berg, R. Ziermann: A high temperature pressure sensor prepared by selective deposition of cubic silicon carbide on SOI substrates, Sens. Actuat. **74**, 56–59 (1999)
7.55 Y.T. Yang, K.L. Ekinci, X.M.H. Huang, L.M. Schiavone, M.L. Roukes, C.A. Zorman, M. Mehregany: Monocrystalline silicon carbide nanoelectromechanical systems, Appl. Phys. Lett. **78**, 162–164 (2001)
7.56 D. Young, J.D. Du, C.A. Zorman, W.H. Ko: High-temperature single-crystal 3C-SiC capacitive pressure sensor, IEEE Sens. J. **4**, 464–470 (2004)
7.57 C.A. Zorman, S. Rajgolpal, X.A. Fu, R. Jezeski, J. Melzak, M. Mehregany: Deposition of polycrystalline 3C-SiC films on 100 mm-diameter (100) Si wafers in a large-volume LPCVD furnace, Electrochem. Solid State Lett. **5**, G99–G101 (2002)
7.58 L. Behrens, E. Peiner, A.S. Bakin, A. Schlachetzski: Micromachining of silicon carbide on silicon fabricated by low-pressure chemical vapor deposition, J. Micromech. Microeng. **12**, 380–384 (2002)
7.59 C.A. Zorman, S. Roy, C.H. Wu, A.J. Fleischman, M. Mehregany: Characterization of polycrystalline silicon carbide films grown by atmospheric pressure chemical vapor deposition on polycrystalline silicon, J. Mater. Res. **13**, 406–412 (1996)
7.60 C.H. Wu, C.A. Zorman, M. Mehregany: Growth of polycrystalline SiC films on SiO_2 and Si_3N_4 by APCVD, Thin Solid Films **355/356**, 179–183 (1999)
7.61 P. Sarro: Silicon carbide as a new MEMS technology, Sens. Actuat. **82**, 210–218 (2000)
7.62 N. Ledermann, J. Baborowski, P. Muralt, N. Xantopoulos, J.M. Tellenbach: Sputtered silicon carbide thin films as protective coatings for MEMS applications, Surf. Coatings Technol. **125**, 246–250 (2000)
7.63 X.A. Fu, R. Jezeski, C.A. Zorman, M. Mehregany: Use of deposition pressure to control the residual stress in polycrystalline SiC films, Appl. Phys. Lett. **84**, 341–343 (2004)
7.64 J. Trevino, X.A. Fu, M. Mehregany, C. Zorman: Low-stress, heavily-doped polycrystalline silicon carbide for MEMS applications. In: *Proc. 18th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 2005) pp. 451–454
7.65 R.S. Okojie, A.A. Ned, A.D. Kurtz: Operation of a 6H-SiC pressure sensor at 500°C, Sens. Actuat. A **66**, 200–204 (1998)
7.66 K. Lohner, K.S. Chen, A.A. Ayon, M.S. Spearing: Microfabricated silicon carbide microengine structures, Mater. Res. Soc. Symp. Proc. **546**, 85–90 (1999)
7.67 K.O. Min, S. Tanaka, M. Esashi: Micro/nano glass press molding using silicon carbide molds fabricated by silicon lost molding. In: *Proc. 18th Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 2005) pp. 475–478
7.68 S. Tanaka, S. Sugimoto, J.-F. Li, R. Watanabe, M. Esashi: Silicon carbide micro-reaction-sintering using micromachined silicon molds, J. Microelectromech. Syst. **10**, 55–61 (2001)
7.69 L.A. Liew, W. Zhang, V.M. Bright, A. Linan, M.L. Dunn, R. Raj: Fabrication of SiCN ceramic MEMS using injectable polymer-precursor technique, Sens. Actuat. A **89**, 64–70 (2001)

7.70 A.J. Fleischman, S. Roy, C.A. Zorman, M. Mehregany: Polycrystalline silicon carbide for surface micromachining. In: *Proc. 9th Int. Workshop Microelectromech. Syst.* (IEEE, Piscataway 1996) pp. 234–238

7.71 A.J. Fleischman, X. Wei, C.A. Zorman, M. Mehregany: Surface micromachining of polycrystalline SiC deposited on SiO_2 by APCVD, Mater. Sci. Forum **264–268**, 885–888 (1998)

7.72 G. Beheim, C.S. Salupo: Deep RIE process for silicon carbide power electronics and MEMS, Mater. Res. Soc. Symp. Proc. **622**, T8.8.1–T8.8.6 (2000)

7.73 W.N. Sharpe, G.M. Beheim, L.J. Evans, N.N. Nemeth, O.M. Jadaan: Fracture strength of single-crystal silicon carbide microspecimens at 24°C and 1000°C, J. Microelectromech. Syst. **17**, 244–254 (2008)

7.74 A. Yasseen, C.H. Wu, C.A. Zorman, M. Mehregany: Fabrication and testing of surface micromachined polycrystalline SiC micromotors, Electron Dev. Lett. **21**, 164–166 (2000)

7.75 X. Song, S. Rajgolpal, J.M. Melzak, C.A. Zorman, M. Mehregany: Development of a multilayer sic surface micromachining process with capabilities and design rules comparable with conventional polysilicon surface micromachining, Mater. Sci. Forum **389–393**, 755–758 (2001)

7.76 D. Gao, R.T. Howe, R. Maboudian: High-selectivity etching of polycrystalline 3C-SiC films using HBr-based transformer coupled plasma, Appl. Phys. Lett. **82**, 1742–1744 (2004)

7.77 D. Gao, M.B. Wijesundara, C. Carraro, R.T. Howe, R. Maboudian: Recent progress toward and manufacturable polycrystalline SiC surface micromachining technology, IEEE Sens. J. **4**, 441–448 (2004)

7.78 X.M.H. Huang, C.A. Zorman, M. Mehregany, M.L. Roukes: Nanodevice motion at microwave frequenies, Nature **421**, 496 (2003)

7.79 T. Shibata, Y. Kitamoto, K. Unno, E. Makino: Micromachining of diamond film for MEMS applications, J. Microelectromech. Syst. **9**, 47–51 (2000)

7.80 H. Bjorkman, P. Rangsten, P. Hollman, K. Hjort: Diamond replicas from microstructured silicon masters, Sens. Actuat. **73**, 24–29 (1999)

7.81 P. Rangsten, H. Bjorkman, K. Hjort: Microfluidic components in diamond. In: *Proc. 10th Int. Conf. Solid State Sens. Actuat.* (Institute of Electrical Engineers of Japan, Tokyo 1999) pp. 190–193

7.82 H. Bjorkman, P. Rangsten, K. Hjort: Diamond microstructures for optical microelectromechanical systems, Sens. Actuat. **78**, 41–47 (1999)

7.83 M. Aslam, D. Schulz: Technology of diamond microelectromechanical systems. In: *Proc. 8th Int. Conf. Solid State Sens. Actuat.* (IEEE, Piscataway 1995)

7.84 R. Ramesham: Fabrication of diamond microstructures for microelectromechanical systems (MEMS) by a surface micromachining process, Thin Solid Films **340**, 1–6 (1999)

7.85 Y. Yang, X. Wang, C. Ren, J. Xie, P. Lu, W. Wang: Diamond surface micromachining technology, Diam. Relat. Mater. **8**, 1834–1837 (1999)

7.86 X.D. Wang, G.D. Hong, J. Zhang, B.L. Lin, H.Q. Gong, W.Y. Wang: Precise patterning of diamond films for MEMS application, J. Mater. Process. Technol. **127**, 230–233 (2002)

7.87 J. Wang, J.E. Butler, D.S.Y. Hsu, C.T.C. Nguyen: CVD polycrystalline diamond high-Q micromechanical resonators. In: *Proc. 15th Int. Conf. Micrelectromech. Syst.* (IEEE, Piscataway 2001) pp. 657–660

7.88 J. Wang, J.E. Butler, T. Feygelson, C.T.C. Nguyen: 1.51 GHz nanocrystalline diamond micromechanical disk resonator with material mismatched isolating support. In: *Proc. 17th IEEE Int. Conf. Microelectromech. Syst.* (IEEE, Piscataway 2004) pp. 641–644

7.89 N. Sepulveda, D. Aslam, J.P. Sullivan: Polycrystalline diamond MEMS resonator technology for sensor applications, Diam. Relat. Mater. **15**, 398–403 (2006)

7.90 L. Sekaric, J.M. Parpia, H.G. Craighead, T. Feygelson, B.H. Houston, J.E. Butler: Nanomechanical resonant structures in nanocrystalline diamond, Appl. Phys. Lett. **81**, 4455–4457 (2002)

7.91 A.R. Krauss, O. Auciello, D.M. Gruen, A. Jayatissa, A. Sumant, J. Tucek, D.C. Mancini, N. Moldovan, A. Erdemire, D. Ersoy, M.N. Gardos, H.G. Busmann, E.M. Meyer, M.Q. Ding: Ultrananocrystalline diamond thin films for MEMS and moving mechanical assembly devices, Diam. Relat. Mater. **10**, 1952–1961 (2001)

7.92 X. Xiao, J. Birrell, J.E. Gerbi, O. Auciello, J.A. Carlisle: Low temperature growth of ultrananocrystalline diamond, J. Appl. Phys. **96**, 2232–2239 (2004)

7.93 S. Srinivasan, J. Hiller, B. Kabius, O. Auciello: Piezoelectric/untrananocrytalline diamond heterostructures for high-performance multifunctional micro/nanoelectromechanical systems, Appl. Phys. Lett. **90**, 134101-1–134101-3 (2007)

7.94 H.D. Espinosa, B. Peng, N. Moldovan, T.A. Friedmann, X. Xiao, D.C. Mancini, O. Auciello, J. Carlisle, C.A. Zorman, M. Mehregany: Elasticity, strength and toughness of single-crystal silicon carbide, ultrananocrystalline diamond, and hydrogen-free tetrahedral amporphous carbon, Appl. Phys. Lett. **89**, 073111-1–073111-3 (2006)

7.95 X. Xiao, J. Wang, C. Liu, J.A. Carlisle, B. Mech, R. Greenberg, D. Guven, R. Freda, M.S. Humayun, J. Weiland, O. Auciello: In vitro and in vivo evaluation of ultrananocrystalline diamond for coating of implantable retinal microchips, J. Biomed. Mater. Res. B **77B**, 273–281 (2006)

7.96 F.J. Hernandez-Guillen, K. Janischowsky, W. Ebert, E. Kohn: Nanocrystalline diamond films for mechanical applications, Phys. Stat. Solidi (a) **201**, 2553–2557 (2004)

7.97 T.A. Friedmann, J.P. Sullivan, J.A. Knapp, D.R. Tallant, D.M. Follstaedt, D.L. Medlin, P.B. Mirkarimi: Thick stress-free amorphous-tetrahedral carbon films with hardness near that of diamond, Appl. Phys. Lett. **71**, 3820–3822 (1997)

7.98 J.P. Sullivan, T.A. Friedmann, K. Hjort: Diamond and amorphous carbon MEMS, MRS Bull. **26**, 309–311 (2001)

7.99 J.R. Webster, C.W. Dyck, J.P. Sullivan, T.A. Friedmann, A.J. Carton: Performance of amorphous diamond RF MEMS capacitive switch, Electron. Lett.

40, 43–44 (2004)
7.100 K. Hjort, J. Soderkvist, J.-A. Schweitz: Galium arsenide as a mechanical material, J. Micromech. Microeng. **4**, 1–13 (1994)
7.101 K. Hjort: Sacrificial etching of III-V compounds for micromechanical devices, J. Micromech. Microeng. **6**, 370–365 (1996)
7.102 K. Fobelets, R. Vounckx, G. Borghs: A GaAs pressure sensor based on resonant tunnelling diodes, J. Micromech. Microeng. **4**, 123–128 (1994)
7.103 A. Dehe, K. Fricke, K. Mutamba, H.L. Hartnagel: A piezoresistive GaAs pressure sensor with GaAs/AlGaAs membrane technology, J. Micromech. Microeng. **5**, 139–142 (1995)
7.104 A. Dehe, K. Fricke, H.L. Hartnagel: Infrared thermopile sensor based on AlGaAs-GaAs micromachining, Sens. Actuat. A **46/47**, 432–436 (1995)
7.105 A. Dehe, J. Peerlings, J. Pfeiffer, R. Riemenschneider, A. Vogt, K. Streubel, H. Kunzel, P. Meissner, H.L. Hartnagel: III-V compound semiconductor micromachined actuators for long resonator tunable fabry-perot detectors, Sens. Actuat. A **68**, 365–371 (1998)
7.106 T. Lalinsky, S. Hascik, Z. Mozolova, E. Burian, M. Drzik: The improved performance of GaAs micromachined power sensor microsystem, Sens. Actuat. **76**, 241–246 (1999)
7.107 T. Lalinsky, E. Burian, M. Drzik, S. Hascik, Z. Mozolova, J. Kuzmik, Z. Hatzopoulos: Performance of GaAs micromachined microactuator, Sens. Actuat. **85**, 365–370 (2000)
7.108 H.X. Tang, X.M.H. Huang, M.L. Roukes, M. Bichler, W. Wegscheider: Two-dimensional electron-gas actuation and transduction for GaAs nanoelectromechanical systems, Appl. Phys. Lett. **81**, 3879–3881 (2002)
7.109 T.S. Tighe, J.M. Worlock, M.L. Roukes: Direct thermal conductance measurements on suspended monocrystalline nanostructures, Appl. Phys. Lett. **70**, 2687–2689 (1997)
7.110 J. Miao, B.L. Weiss, H.L. Hartnagel: Micromachining of three-dimensional GaAs membrane structures using high-energy nitrogen implantation, J. Micromech. Microeng. **13**, 35–39 (2003)
7.111 C. Seassal, J.L. Leclercq, P. Viktorovitch: Fabrication of InP-based freestanding microstructures by selective surface micromachining, J. Micromech. Microeng. **6**, 261–265 (1996)
7.112 J. Leclerq, R.P. Ribas, J.M. Karam, P. Viktorovitch: III-V micromachined devices for microsystems, Microelectron. J. **29**, 613–619 (1998)
7.113 H. Yamaguchi, R. Dreyfus, S. Miyashita, Y. Hirayama: Fabrication and elastic properties of InAs freestanding structures based on InAs/GaAs(111)A heteroepitaxial systems, Physica E **13**, 1163–1167 (2002)
7.114 K. Deng, P. Kumar, L. Li, D.L. Devoe: Piezoelectric disk resonators based on epitaxial AlGaAs films, J. Microelectromech. Syst. **16**, 155–162 (2007)
7.115 C. Lee, T. Itoh, T. Suga: Micromachined piezoelectric force sensors based on PZT thin films, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 553–559 (1996)
7.116 B. Xu, L.E. Cross, J.J. Bernstein: Ferroelectric and antiferroelectric films for microelectromechanical systems applications, Thin Solid Films **377/378**, 712–718 (2000)
7.117 E. Hong, S. Trolier-McKinstry, R.L. Smith, S.V. Krishnaswamy, C.B. Freidhoff: Design of MEMS PZT circular diaphragm actuators to generate large deflections, J. Microelectromech. Syst. **15**, 832–839 (2006)
7.118 S.P. Beeby, A. Blackburn, N.M. White: Processing of PZT piezoelectric thick films on silicon for microelectromechanical systems, J. Micromech. Microeng. **9**, 218–229 (1999)
7.119 K. Tonisch, C. Buchheim, F. Niebelschütz, M. Donahue, R. Goldhahn, V. Cimalla, O. Ambacher: Piezoelectric actuation of all-nitride MEMS, Phys. Stat. Solidi (c) **5**, 1910–1913 (2008)
7.120 C. Giordano, I. Ingrosso, M.T. Todaro, G. Maruccio, S. De Guido, R. Cingolani, A. Passaseo, M. De Vittorio: AlN on polysilicon piezoelectric cantilevers for sensors/actuators, Microelectron. Eng. **86**, 1204–1207 (2009)
7.121 M. Schneider, A. Bittner, U. Schmid: Thickness dependence of Young's modulus and residual stress of sputtered aluminum nitride thin films, Appl. Phys. Lett. **105**, 201912 (2014)
7.122 G. Piazza, P.J. Stephanou, A.P. Pisano: Piezoelectric aluminum nitride vibrating contour-mode MEMS resonators, J. Microelectromech. Syst. **15**, 1406–1418 (2006)
7.123 N. Sinha, G.E. Wabiszewski, R. Mahameed, V.V. Felmetsger, S.M. Tanner, R.W. Carpick, G. Piazza: Piezoelectric aluminum nitride nanoelectromechanical actuators, Appl. Phys. Lett. **95**, 053106 (2009)
7.124 R.B. Karabalin, M.H. Matheny, X.L. Feng, E. Defaÿ, G. Le Rhun, C. Marcoux, S. Hentz, P. Andreucci, M.L. Roukes: Piezoelectric nanoelectromechanical resonators based on aluminum nitride thin films, Appl. Phys. Lett. **95**, 103111 (2009)
7.125 P. Ramesh, S. Krishnamoorthy, S. Rajan, G.N. Washington: Fabrication and characterization of a piezoelectric gallium nitride switch for optical MEMS applications, Smart Mater. Struct. **21**, 094003 (2012)
7.126 J. Lv, Z. Yang, G. Member: Yan, W. Lin, Y. Cai, B. Zhang, K.J. Chen: Fabrication of large-area suspended MEMS structures using GaN-on-Si platform, IEEE Electr. Dev. Lett. **30**, 1045–1047 (2009)
7.127 M. Rais-Zadeh, V.J. Gokhale, A. Ansari, M. Faucher, D. Théron, Y. Cordier, L. Buchaillot: Gallium nitride as an electromechanical material, J. Microelectromech. Syst. **23**, 1252–1271 (2014)
7.128 A.B. Amar, M. Faucher, V. Brandli, Y. Cordier, D. Théron: Young's modulus extraction of epitaxial heterostructure AlGaN/GaN for MEMS application, Phys. Stat. Solidi (a) **211**, 1655–1659 (2014)
7.129 T. Zimmermann, M. Neuburger, P. Benkart, F.J. Hernández-Guillén, C. Pietzka, M. Kunze, I. Daumiller, A. Dadgar, A. Krost, E. Kohn: Piezoelectric GaN sensor structures, IEEE Electron Dev. Lett. **27**, 309–312 (2006)

7.130 C. Shearwood, M.A. Harradine, T.S. Birch, J.C. Stevens: Applications of polyimide membranes to MEMS technology, Microelectron. Eng. **30**, 547–550 (1996)

7.131 F. Jiang, G.B. Lee, Y.C. Tai, C.M. Ho: A flexible micromachine-based shear-stress sensor array and its application to separation-point detection, Sens. Actuat. **79**, 194–203 (2000)

7.132 H. Yousef, K. Hjort, M. Lindberg: Vertical thermopiles embedded in a polyimide-based flexible printed circuit board, J. Microelectromech. Syst. **16**, 1341–1348 (2007)

7.133 D. Memmi, V. Foglietti, E. Cianci, G. Caliano, M. Pappalardo: Fabrication of capacitive micromechanical ultrasonic transducers by low-temperature process, Sens. Actuat. A **99**, 85–91 (2002)

7.134 A. Bagolini, L. Pakula, T.L.M. Scholtes, H.T.M. Pham, P.J. French, P.M. Sarro: Polyimide sacrificial layer and novel materials for post-processing surface micromachining, J. Micromech. Microeng. **12**, 385–389 (2002)

7.135 T. Stieglitz: Flexible biomedical microdevices with double-sided electrode arrangements for neural applications, Sens. Actuat. A **90**, 203–211 (2001)

7.136 T. Stieglitz, G. Matthias: Flexible BioMEMS with electrode arrangements on front and back side as key component in neural prostheses and biohybrid systems, Sens. Actuat. B **83**, 8–14 (2002)

7.137 T. Stieglitz, M. Schuettler, K.P. Koch: Implantable Biomedical Microsystems for neural prostheses, IEEE Eng. Med. Biol. **24**, 58–65 (2005)

7.138 H. Lorenz, M. Despont, N. Fahrni, J. Brugger, P. Vettiger, P. Renaud: High-aspect-ratio, ultrathick, negative-tone-near-UV photoresist and its applications in MEMS, Sens. Actuat. A **64**, 33–39 (1998)

7.139 H. Lorenz, M. Despont, N. Fahrni, N. LaBianca, P. Renaud, P. Vettiger: SU-8: A low-cost negative resist for MEMS, J. Micromech. Microeng. **7**, 121–124 (1997)

7.140 E.H. Conradie, D.F. Moore: SU-8 thick photoresist processing as a functional material for MEMS applications, J. Micromech. Microeng. **12**, 368–374 (2002)

7.141 C.T. Pan, H. Yang, S.C. Shen, M.C. Chou, H.P. Chou: A low-temperature wafer bonding technique using patternable materials, J. Micromech. Microeng. **12**, 611–615 (2002)

7.142 P.A. Stupar, A.P. Pisano: Silicon, parylene, and silicon/parylene micro-needles for strength and toughness. In: *Tech. Digest 11th Int. Conf. Solid State Sens. Actuat.* (IEEE, Piscataway 2001) pp. 1368–1389

7.143 X. Yang, J.M. Yang, Y.C. Tai, C.M. Ho: Micromachined membrane particle filters, Sens. Actuat. **73**, 184–191 (1999)

7.144 J.M. Zara, S.W. Smith: Optical scanner using a MEMS actuator, Sens. Actuat. A **102**, 176–184 (2002)

7.145 H.S. Noh, P.J. Hesketh, G.C. Frye-Mason: Parylene gas chromatographic column for rapid thermal cycling, J. Microelectromech. Syst. **11**, 718–725 (2002)

7.146 Y. Suzuki, Y.C. Tai: Micromachined high aspect ratio parylene spring and its application to low frequency accelerometers, J. Microelectromech. Syst. **15**, 1364–1370 (2006)

7.147 T.J. Yao, X. Yang, Y.C. Tai: BrF_3 dry release technology for large freestanding parylene microstructures and electrostatic actuators, Sens. Actuat. A **97/98**, 771–775 (2002)

7.148 P.-J. Chen, D.C. Rodger, E.M. Meng, M.S. Humayun, Y.C. Tai: Surface-micromachined parylene dual valves for on-chip unpowered microflow regulation, J. Microelectromech. Syst. **16**, 223–231 (2007)

7.149 D.C.Y.C. Rodger: Tai: Microelectronic packaging for retinal prosthesis, IEEE Eng. Med. Biol. **24**, 52–57 (2005)

7.150 D. Ziegler, T. Suzuki, S. Takeuchi: Fabrication of flexible neural probes with built-in microfluidic channels by thermal bonding of parylene, J. Microelectromech. Syst. **15**, 1477–1482 (2006)

7.151 X. Wang, J. Engel, C. Liu: Liquid crystal polymer (LCP) for MEMS: Processing and applications, J. Micromech. Microeng. **13**, 628–633 (2003)

7.152 C.J. Lee, S.J. Oh, J.K. Song, S.J. Kim: Neural signal recording using microelectrode arrays fabricated on liquid crystal polymer material, Mater. Sci. Eng. C **4**, 265–268 (2004)

7.153 F.F. Faheem, K.C. Gupta, Y.C. Lee: Flip-chip assembly and liquid crystal polymer encapsulation for variable MEMS capacitors, IEEE Trans. Microw. Theory Tech. **51**, 2562–2567 (2003)

7.154 J.N. Palasagaram, R. Ramadoss: MEMS-capacitive pressure sensor fabricated using printed-circuit-processing techniques, IEEE Sens. J. **6**, 1374–1375 (2006)

Part B Nanomat

Part B Nanomaterial and Nanostructures

8 Carbon Nanotubes
Marc Monthioux, Toulouse, France
Philippe Serp, Toulouse, France
Brigitte Caussat, Toulouse, France
Emmanuel Flahaut, Toulouse, France
Manitra Razafinimanana, Toulouse, France
Flavien Valensi, Toulouse, France
Christophe Laurent, Toulouse, France
Alain Peigney, Toulouse, France
David Mesguich, Toulouse, France
Alicia Weibel, Toulouse, France
Wolfgang Bacsa, Toulouse, France
Jean-Marc Broto, Toulouse, France

9 Nanowires
Mildred S. Dresselhaus
Marcie R. Black, Lincoln, USA
Vincent Meunier, Troy, USA
Oded Rabin, College Park, USA

10 Nanoribbons
Toshiaki Enoki, Tokyo, Japan
Shintaro Sato, Atsugi, Kanagawa, Japan

11 Nanoparticles and Their Applications
Seyedsina Moeinzadeh, Columbia, USA
Esmaiel Jabbari, Columbia, USA

12 Graphene
Aravind Vijayaraghavan, Manchester, UK
Maria Iliut, Manchester, UK

8. Carbon Nanotubes

Marc Monthioux, Philippe Serp, Brigitte Caussat, Emmanuel Flahaut, Manitra Razafinimanana, Flavien Valensi, Christophe Laurent, Alain Peigney, David Mesguich, Alicia Weibel, Wolfgang Bacsa, Jean-Marc Broto

Carbon nanotubes (CNTs) are remarkable objects that once looked set to revolutionize the technological landscape in the near future. Since the 1990s and for twenty years thereafter, it was repeatedly claimed that tomorrow's society would be shaped by nanotube applications, just as silicon-based technologies dominate society today. Space elevators tethered by the strongest of cables, hydrogen-powered vehicles, artificial muscles: these were just a few of the technological marvels that we were told would be made possible by the science of carbon nanotubes.

Of course, this prediction is still some way from becoming reality; most often the possibilities and potential have been evaluated, but actual technological development is facing the unforgiving rule that drives the transfer of a new material or a new device to market: profitability. New materials, even more so for nanomaterials, no matter how wonderful they are, have to be cheap to produce, constant in quality, easy to handle, and nontoxic. Those are the conditions for an industry to accept a change in its production lines to make them nanocompatible. Consider the example of fullerenes – molecules closely related to nanotubes. The anticipation that surrounded these molecules, first reported in 1985, resulted in the bestowment of a Nobel Prize for their discovery in 1996. However, two decades later, very few fullerene applications have reached the market, suggesting that similarly enthusiastic predictions about nanotubes should be approached with caution, and so should it be with graphene, another member of the carbon nanoform family which joined the game in 2004, again acknowledged by a Nobel Prize in 2010.

There is no denying, however, that the expectations surrounding carbon nanotubes are still high, because of specificities that make them special compared to fullerenes and graphene: their easiness of production, their dual molecule/nano-object nature, their unique aspect ratio, their robustness, the ability of their electronic structure to be given a gap, and their wide typology etc. Therefore, carbon nanotubes may provide the building blocks for further technological progress, enhancing our standard of living.

In this chapter, we first describe the structures, syntheses, growth mechanisms, and properties of carbon nanotubes. Then we introduce nanotube-based materials, which comprise on the one hand those formed by reactions and associations of all-carbon nanotubes with foreign atoms, molecules and compounds, and on the other hand, composites, obtained by incorporating carbon nanotubes in various matrices. Finally, we will provide a list of applications currently on the market, while skipping the potentially endless and speculative list of possible applications.

8.1 **Structure of Carbon Nanotubes** 194
8.1.1 Single-Wall Nanotubes 194
8.1.2 Multiwall Nanotubes 197

8.2 **Synthesis of Carbon Nanotubes** 199
8.2.1 Solid Carbon Source-Based Synthesis Techniques: The DC Electric Arc 199
8.2.2 Gaseous Carbon Source-Based Synthesis Techniques 202
8.2.3 Miscellaneous Techniques 207
8.2.4 Synthesis with Controlled Orientation 208

8.3 **Growth Mechanisms of Carbon Nanotubes** 210
8.3.1 Catalyst-Free Growth 210
8.3.2 Catalytically Activated Growth 210

8.4 **Properties of Carbon Nanotubes** 213
8.4.1 Overall Properties of SWNTs 213
8.4.2 Adsorption Properties 213
8.4.3 Electronic and Optical Properties 215
8.4.4 Mechanical Properties 216
8.4.5 Reactivity 216

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_8

8.5 **Carbon Nanotube-Based Nano-Objects (Carbon Meta-Nanotubes)** 217
8.5.1 Heteronanotubes 217
8.5.2 Filled Carbon Nanotubes 217
8.5.3 Functionalized Nanotubes 220
8.5.4 Coated/Decorated Nanotubes 221
8.6 **Carbon-Nanotube-Containing Materials (Composites)** 223
8.6.1 Metal Matrix Composites 223
8.6.2 Ceramic Matrix Composites 223
8.6.3 Polymer-Matrix Composites 224
8.6.4 Composites as Multifunctional Materials 225
8.7 **Current Applications of Carbon Nanotubes (on the Market)** 227
8.7.1 Carbon Nanotubes and Master Batches 227
8.7.2 Near-Field Microscopy Probes 227
8.7.3 Electron Emitter 229
8.7.4 Flexible and Touch-Screen Displays 229
8.7.5 Nonvolatile Random Access Memory 229
8.7.6 Light Absorbants 229
8.7.7 Automotive and Aeronautic Industry 230
8.7.8 High-Tech Goods and Clothes 230
8.7.9 Anodes for Li-Ion Batteries 230
8.7.10 Chemical Sensors 230
8.7.11 Catalyst Support 231
8.8 **Toxicity and Environmental Impact of Carbon Nanotubes** 231
8.9 **Concluding Remarks** 233
References 233

Carbon nanotubes (CNTs) have long been synthesized as products of the action of a catalyst on the gaseous species originating from the thermal decomposition of hydrocarbons (Sect. 8.2) [8.1]. The first evidence that the nanofilaments produced in this way were actually nanotubes – that they exhibited an inner cavity – can be found in the transmission electron microscope micrographs published by *Radushkevich* and *Lukyanovich* in 1952 [8.2]. Since then, the interest in carbon nanofilaments/nanotubes was recurrent, though within a scientific area almost limited to the carbon material scientist community [8.3]. Worldwide enthusiasm began unexpectedly in 1991, after the catalyst-free formation of nearly perfect concentric multiwall carbon nanotubes (c-MWNTs, Sect. 8.1) was reported [8.4] as by-products of the formation of fullerenes via the electric-arc technique (Sect. 8.3). But the real breakthrough occurred two years later, when attempts to fill the nanotubes in situ with various metals (Sect. 8.5) led to the discovery – again unexpected – of single-wall carbon nanotubes (SWNTs) (Sect. 8.1) simultaneously by *Iijima* and *Ichihashi* [8.5] and *Bethune* et al. [8.6]. SWNTs were really new nano-objects with properties and behaviors that are often quite specific (Sect. 8.4). They are also beautiful objects for fundamental physics as well as unique molecules for experimental chemistry. Potential applications seem countless for such CNTs or CNTs combined with a matrix (Sect. 8.6), and some have reached the market (Sect. 8.7) yet toxicity and environmental issues should not be ignored (Sect. 8.8). Consequently, it still remains an extraordinarily active – and highly competitive – field of research, as that of fullerenes was and that of graphene now is, which, by the way, are again carbon nano-objects structurally closely related to nanotubes.

8.1 Structure of Carbon Nanotubes

It is relatively easy to imagine a single-wall carbon nanotube (SWNT). Ideally, it is enough to consider a perfect graphene sheet (graphene is a polyaromatic monoatomic layer consisting of sp^2-hybridized carbon atoms arranged in hexagons; genuine graphite consists of layers of this graphene) and to roll it into a cylinder (Fig. 8.1), making sure that the hexagonal rings placed in contact join coherently.

Then the tips of the tube are sealed by two caps, each cap being a hemifullerene with the appropriate diameter (Fig. 8.2a–c).

8.1.1 Single-Wall Nanotubes

Calculations have shown that collapsing the single-wall tube into a flattened two-layer ribbon is energetically more favorable than maintaining the tubular morphology beyond a diameter value of ≈ 2.5 nm [8.8]. Conversely, the shorter the radius of curvature, the higher the stress and the energetic cost, although SWNTs with diameters as low as 0.4 nm have been synthesized successfully [8.9]. A suitable energetic compromise is therefore reached for ≈ 1.35 nm, the most frequent di-

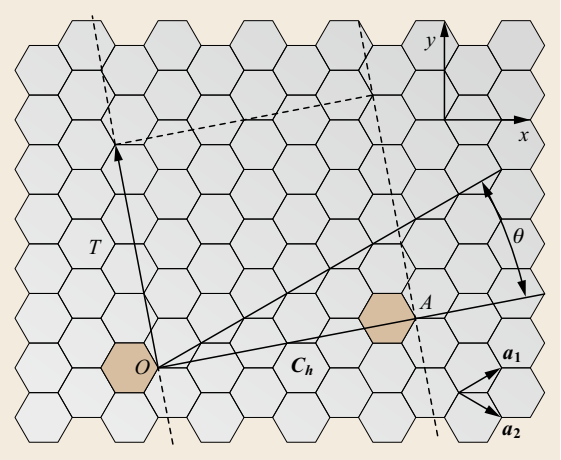


Fig. 8.1 Making an SWNT, starting from a graphene sheet (after [8.7])

ameter encountered when the growth is free from the influence of the catalytic particle size (as for the synthesis technique based on solid carbon sources) and when conditions ensuring high SWNT yields are used. Nanotube length only depends on the specific synthesis conditions (thermal gradients, residence time, and so on). Experimental data are consistent with these statements, since SWNTs wider than 2.5 nm are only rarely reported in the literature, whatever the preparation method, while the length of the SWNTs can be in the micrometer or the millimeter range. These features make SWNTs a unique example of single molecules with huge aspect ratios.

Two important consequences derive from the SWNT structure as described above:

1. All carbon atoms are involved in hexagonal aromatic rings only and are therefore in equivalent positions, except at each nanotube tip, where $6 \times 5 = 30$ atoms are involved in pentagonal rings (considering that adjacent pentagons are unlikely) as a consequence of Euler's rule that also governs the fullerene structure. For ideal SWNTs, chemical reactivity will therefore be highly favored at the tube tips, at the locations of the pentagonal rings.
2. Although carbon atoms are involved in aromatic rings, the C=C bond angles are not planar. This means that the hybridization of carbon atoms is not pure sp^2; it has some degree of sp^3 character, in a proportion that increases as the tube radius of curvature decreases. The effect is the same as for the C_{60} fullerene molecules, whose radius of curvature is 0.35 nm, and whose bonds therefore have 10% sp^3 character [8.11]. This has consequences for SWNT reactivity and electronic behavior (Sect. 8.4).

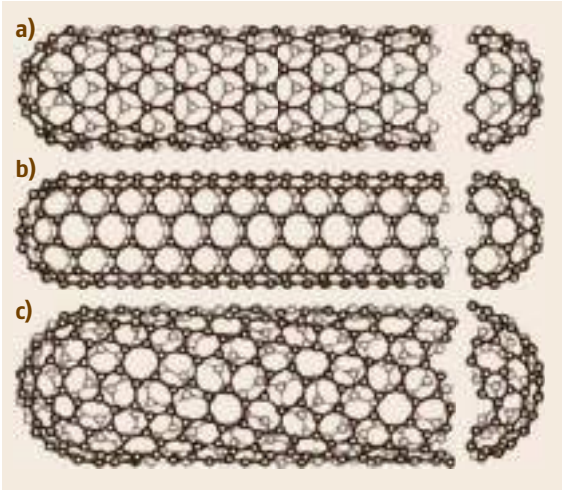


Fig. 8.2a–c Sketches of three different SWNT structures that are examples of **(a)** a zigzag-type nanotube, **(b)** an armchair-type nanotube, **(c)** a helical nanotube (after [8.10])

As illustrated by Fig. 8.2, there are many ways to roll a graphene into a single-walled nanotube, with some of the resulting nanotubes possessing planes of symmetry both parallel and perpendicular to the nanotube axis (such as the SWNTs from Fig. 8.2a,b), while others do not (such as the SWNT from Fig. 8.2c). Similar to the terms used for molecules, the latter are commonly called *chiral* nanotubes, since one cannot superimpose them on their own image in a mirror. *Helical* should however be preferred (see below). The various ways to roll graphene into tubes are therefore mathematically defined by the vector of helicity $\boldsymbol{C}_\mathrm{h}$, and the angle of helicity θ, as follows (referring to Fig. 8.1)

$$\boldsymbol{OA} = \boldsymbol{C}_\mathrm{h} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$$

with

$$\boldsymbol{a}_1 = \frac{a\sqrt{3}}{2}\boldsymbol{x} + \frac{a}{2}\boldsymbol{y} \quad \text{and} \quad \boldsymbol{a}_2 = \frac{a\sqrt{3}}{2}\boldsymbol{x} + \frac{a}{2} - \boldsymbol{y}\,,$$

where $a = 2.46$ Å, and

$$\cos\theta = \frac{2n+m}{2\sqrt{n^2+m^2+mn}}\,,$$

where n and m are the integers of the vector $\boldsymbol{OA}$ considering the unit vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.

The vector of helicity $\boldsymbol{C}_\mathrm{h}$ $(= \boldsymbol{OA})$ is perpendicular to the tube axis, while the angle of helicity θ is taken with respect to the so-called zigzag axis: the vector of helicity that results in nanotubes of the *zigzag* type (see below). The diameter D of the corresponding nanotube

is related to $\boldsymbol{C}_\mathrm{h}$ by the relation

$$D = \frac{|\boldsymbol{C}_\mathrm{h}|}{\pi} = \frac{a_{\mathrm{CC}}\sqrt{3(n^2+m^2+nm)}}{\pi} ,$$

where

$$\underset{\text{(graphite)}}{1.41\,\text{Å}} \leq a_{\mathrm{C=C}} \leq \underset{(\mathrm{C_{60}})}{1.44\,\text{Å}} .$$

The C−C bond length is actually elongated by the curvature imposed by the structure; the average bond length in the C_{60} fullerene molecule is a reasonable upper limit, while the bond length in flat graphene in genuine graphite is the lower limit (corresponding to an infinite radius of curvature). Since $\boldsymbol{C}_\mathrm{h}$, θ, and D are all expressed as a function of the integers n and m, they are sufficient to define any particular SWNT by denoting them (n, m). The values of n and m for a given SWNT can be simply obtained by counting the number of hexagons that separate the extremities of the $\boldsymbol{C}_\mathrm{h}$ vector following the unit vector $\boldsymbol{a}_1$ first and then $\boldsymbol{a}_2$ [8.7]. In the example of Fig. 8.1, the SWNT that is obtained by rolling the graphene so that the two shaded aromatic cycles can be superimposed exactly is a (4,2) chiral nanotube. Similarly, SWNTs from Fig. 8.2a–c are (9,0), (5,5), and (10,5) nanotubes respectively, thereby providing examples of zigzag-type SWNT (with an angle of helicity = 0°), armchair-type SWNT (with an angle of helicity of 30°), and a chiral SWNT, respectively. This also illustrates why the term *chiral* is sometimes inappropriate and should preferably be replaced with *helical*. Armchair (n, n) nanotubes, although definitely achiral from the standpoint of symmetry, exhibit a nonzero *chiral angle*. *Zigzag* and *armchair* qualifications for achiral nanotubes refer to the way that the carbon atoms are displayed at the edge of the nanotube cross section (Fig. 8.2a,b). Generally speaking, it is clear from Figs. 8.1 and 8.2a that having the vector of helicity perpendicular to any of the three overall C=C bond directions will provide zigzag-type SWNTs, denoted $(n, 0)$, while having the vector of helicity parallel to one of the three C=C bond directions will provide armchair-type SWNTs, denoted (n, n). Conversely, because of the sixfold symmetry of the graphene sheet, the angle of helicity θ for the chiral (n, m) nanotubes is such that $0 < \theta < 30°$. Figure 8.3 provides two examples of what chiral SWNTs look like, as seen via atomic force microscopy.

The graphenes in graphite have π electrons which are accommodated by the stacking of graphenes, allowing van der Waals forces to develop. Similar reasons make fullerenes gather and order into fullerite crystals and SWNTs into SWNT ropes (Fig. 8.4a). Provided the SWNT diameter distribution is narrow, the SWNTs in ropes tend to spontaneously arrange into hexagonal arrays, which correspond to the highest compactness achievable (Fig. 8.4b). This feature brings new periodicities with respect to graphite or turbostratic polyaromatic carbon crystals. Turbostratic structure corresponds to graphenes that are stacked with random rotations or translations instead of being piled up following sequential *ABAB* positions, as in graphite structure. This implies that no lattice atom plane exists other than the graphene planes themselves (corresponding to the (001) atom plane family). These new periodicities give specific diffraction patterns that are quite

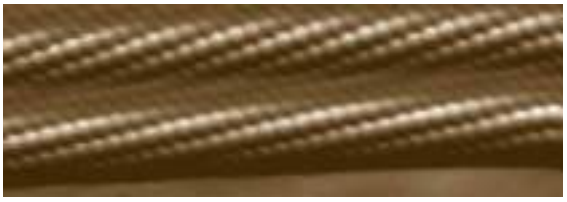


Fig. 8.3 Image of two neighboring chiral SWNTs within an SWNT bundle as seen using high-resolution scanning tunneling microscopy (courtesy of Prof. Yazdani, University of Illinois at Urbana, USA)

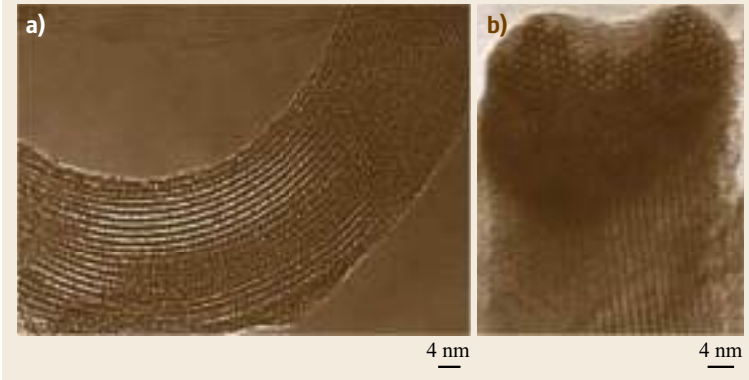


Fig. 8.4a,b High-resolution transmission microscopy (TEM) images of an SWNT rope. **(a)** Longitudinal view. An isolated single SWNT also appears at the top of the image. **(b)** Cross-sectional view (after [8.12])

different to those of other sp^2-carbon-based crystals, although *hk* reflections, which account for the hexagonal symmetry of the graphene plane, are still present. Conversely, 00*l* reflections, which account for the stacking sequence of graphenes in regular, *multilayered* polyaromatic crystals (which do not exist in SWNT ropes) are absent. This hexagonal packing of SWNTs within the ropes requires that SWNTs exhibit similar diameters, which is the usual case for SWNTs prepared by electric arc or laser vaporization processes. SWNTs prepared using these methods are actually about 1.35 nm wide (diameter of a (10,10) tube, among others), for reasons related to the energetic compromise mentioned earlier that the latter synthesis techniques allow to occur during the specific growth mechanisms taking place (Sect. 8.3).

8.1.2 Multiwall Nanotubes

Building MWNTs (multiwall carbon nanotubes) is a little bit more complex, since it involves the various ways graphenes can be displayed and mutually arranged within filamentary morphology. A similar versatility can be expected to the usual textural versatility of polyaromatic solids. Likewise, their diffraction patterns are barely differentiable from those of anisotropic polyaromatic solids. The easiest MWNT to imagine is the concentric type (c-MWNT), in which SWNTs with regularly increasing diameters are coaxially arranged (according to a Russian-doll model) into a multiwall nanotube (Fig. 8.5). Such nanotubes are generally formed either by the electric-arc technique (without the need for a catalyst), by catalyst-enhanced thermal cracking of gaseous hydrocarbons, or by CO disproportionation (Sect. 8.2). There can be any number of walls (or coaxial tubes), from two upwards. The intertube distance is approximately the same as the intergraphene distance in turbostratic, polyaromatic solids, 0.34 nm (as opposed to 0.335 nm in genuine graphite), since the increasing radius of curvature imposed on the concentric graphenes prevents the carbon atoms from being arranged according to the Bernal sequence as in graphite. However, two cases allow a nanotube to reach – totally or partially – the 3-D crystal periodicity of graphite. One is to consider a high number of concentric graphenes, where the more external concentric graphenes may have a long radius of curvature. In this case, the shift in the relative positions of carbon atoms from superimposed graphenes is so small with respect to that in graphite that some commensurability is possible over distances long enough to make Bernal stacking possible over a large enough area. This may result in MWNTs where both structures are associated; in other words they have turbostratic cores

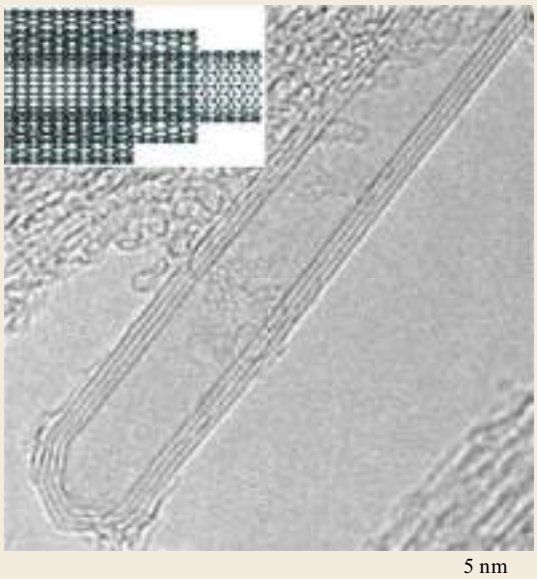


Fig. 8.5 High-resolution TEM image (longitudinal view) of a concentric MWNT (c-MWNT prepared using an electric arc). The *insert* shows a model of the related Russian doll-like arrangement of graphenes (image by M. Monthioux, CNRS. Model by I. Suarez-Martinez, Curtin University)

and graphitic outer parts [8.13]. The other case occurs for c-MWNTs exhibiting faceted morphologies, usually originating from subsequent heat treatment at high temperature (e.g., 2000 °C) in an inert atmosphere. Facets allow the graphenes to resume a flat arrangement of atoms (except at the junction between neighboring facets) which allows the specific stacking sequence of graphite to develop.

Another frequent inner texture for MWNTs is the so-called herringbone texture (h-MWNTs), in which the graphenes make an angle with respect to the nanotube axis (Fig. 8.6). The value of this angle varies with the processing conditions (such as catalyst morphology or composition of the atmosphere), from 0 (in which case the texture becomes that of a c-MWNT) to 90° (in which case the filament is no longer a tube, see below), and the inner diameter varies so that the tubular arrangement can be lost [8.15], meaning that the latter are more accurately called nanofibers rather than nanotubes. h-MWNTs are exclusively obtained by processes involving catalysts, generally catalyst-enhanced thermal cracking of hydrocarbons or CO disproportionation. One long-debated question was whether the herringbone texture, which actually describes the texture projection rather than the overall three-dimensional texture, originates from the scroll-like spiral arrange-

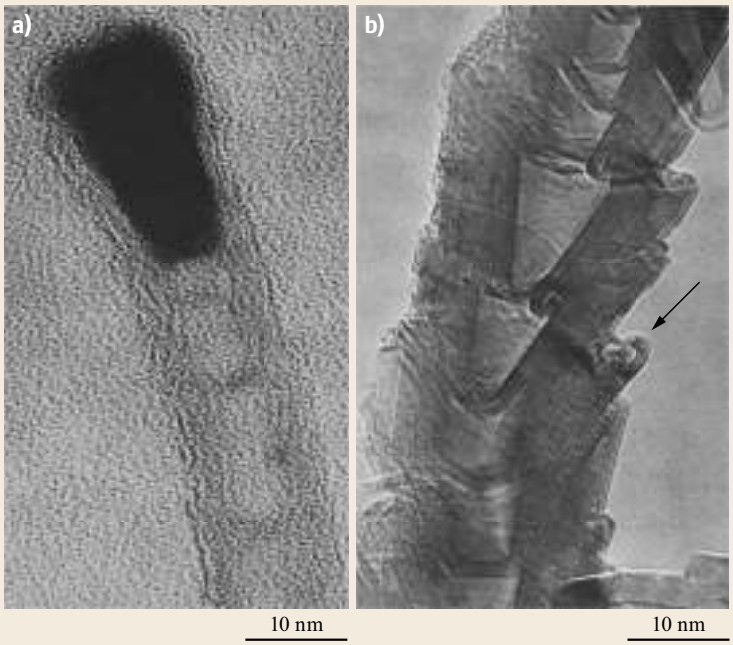


Fig. 8.6a,b Some of the earliest high-resolution TEM images of a herringbone (and bamboo) multiwall nanotube (bh-MWNT, longitudinal view) prepared by CO disproportionation on an Fe-Co catalyst. **(a)** As-grown. The nanotube surface is made of free graphene edges. **(b)** After 2900 °C heat treatment. Both the herringbone and the bamboo textures have become obvious. Graphene edges from the surface have buckled with their neighbors (*arrow*), closing off access to the intergraphene space (after [8.14])

ment of a single graphene ribbon or from the stacking of independent truncated cone-like graphenes in what is also called a *cup-stack* texture. It has now been demonstrated that both exist [8.16, 17].

Another common feature is the occurrence, to some degree, of a limited amount of graphenes oriented perpendicular to the nanotube axis, thus forming a *bamboo* texture. This is not a texture that can exist on its own; it affects either the c-MWNT (bc-MWNT) or the h-MWNT (bh-MWNT) textures (Figs. 8.6 and 8.7). The question is whether such filaments, although hollow, should still be called nanotubes, since the inner cavity is no longer open all the way along the filament as it is for a genuine tube. These are therefore sometimes also referred to as *nanofibers* in the literature.

One nanofilament that definitely cannot be called a nanotube is built from graphenes oriented perpendicular to the filament axis and stacked as piled-up plates. Such filaments are referred to as *platelet nanofibers* in the literature [8.15].

Unlike SWNTs, whose aspect ratios are so high that it is almost impossible to find the tube tips, the aspect ratios for MWNTs (and carbon nanofibers) are generally lower and often allow one to image tube ends by TEM. Aside from c-MWNTs derived from an electric arc (Fig. 8.5), which grow in a catalyst-free process, nanotube tips are frequently found to be associated with the catalyst crystals from which they were formed.

The properties of the MWNT (Sect. 8.4) will obviously largely depend on the perfection and the orientation of the graphenes in the tube (for example, the helicity angles of the nanotubes constituting c-MWNTs has little importance). Graphene orientation is a matter of texture, whereas graphene perfection is a matter of nanotexture, which both are commonly used to describe graphene-based materials [8.20], and depend on the processing conditions. While texture is a permanent fea-



Fig. 8.7a,b TEM images from bamboo multiwall nanotubes (longitudinal views). **(a)** Low magnification of a bamboo-herringbone multiwall nanotube (bh-MWNT) showing the nearly periodic nature of the texture, which occurs very frequently (after [8.18]); **(b)** high-resolution image of a bamboo-concentric multiwall nanotube (bc-MWNT) (after [8.19])

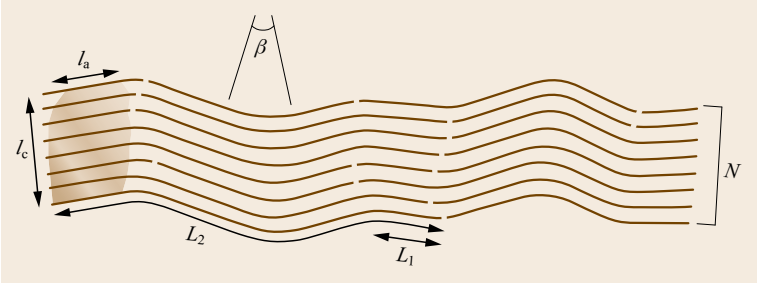


Fig. 8.8 Sketch explaining the various parameters obtained from high-resolution (lattice fringe mode) TEM, used to quantify nanotexture: L_1 is the average length of perfect (distortion-free) graphenes of coherent areas; N is the number of piled-up graphenes in coherent (distortion-free) areas; L_2 is the average length of continuous though distorted graphenes within graphene stacks; β is the average distortion angle. L_1 and N are related to the l_a and l_c values obtained from X-ray diffraction

ture (unless subjected to severe degradation), nanotexture can be quantified (Fig. 8.8), as it can be improved by subsequent thermal treatments at high temperatures (such as $> 2000\,^{\circ}\mathrm{C}$) and potentially degraded by chemical treatments (such as slightly oxidizing conditions).

To summarize, because work in the field started more than a century ago, the names of filamentous carbons have changed with time from vapor-grown carbon fibers, to nanofilaments, nanofibers, and eventually nanotubes. For all filamentous carbons, the generic name should be vapor-grown carbon nanofilaments (VGCNF). Whether or not the filaments are tubular is then a matter of inner morphology description [8.20], which can be further refined by considering textural features such as bamboo, herringbone, and concentric (see above). In the following, we will therefore use MWNTs for any hollowed nanofilament, whether they contain graphene walls oriented transversally or not. Any other nanofilament (i. e., not hollowed) will be termed a *nanofiber*.

8.2 Synthesis of Carbon Nanotubes

8.2.1 Solid Carbon Source–Based Synthesis Techniques: The DC Electric Arc

Principle

One route to produce CNTs is to vaporize a solid carbon source (usually graphite), which requires processes able to generate high temperatures ($T > 5000\,\mathrm{K}$). Various techniques have been employed for this (e.g., laser ablation, DC electric arc, solar furnace). We will focus here on the DC electric-arc technique, as the other methods have been almost abandoned for being less robust and versatile. Some information on the other techniques will be provided in Sect. 8.2.3. Further information can be found in the previous edition of this chapter [8.21].

The DC electric-arc process is based on carbon vaporization in an inert gas atmosphere at reduced pressure. The morphologies of the carbon nanostructures and the SWNT yields can differ notably with respect to the experimental conditions. It was (and still is) the first method of producing fullerenes in relatively large quantities [8.22–24]. In the course of investigating other carbon nanostructures formed along with the fullerenes, and more particularly the solid carbon deposit that formed on the cathode, *Iijima* [8.4] discovered the catalyst-free formation of perfect c-MWNT-type CNTs. Then, as mentioned above, the catalyst-promoted formation of SWNTs was accidentally discovered after amounts of transition metal were introduced into the anode in an attempt to fill the c-MWNTs with metals during growth [8.5, 6]. Since then, a lot of work has been carried out by many groups using this technique in order to understand the mechanisms of nanotube growth in the synthesis of MWNTs and/or SWNTs [8.25–36].

The different mechanisms (such as carbon molecule dissociation and atom-recombination processes) involved in this high-temperature technique take place at different time scales, from nanoseconds to microseconds and even milliseconds. The formation of nanotubes and other graphene-based products occurs afterward with a relatively long delay.

The electric-arc method is based on the energy transfer resulting from the interaction between the electrode and the plasma. This interaction causes anode erosion, leading to the formation of a plasma, i. e., an electrically neutral ionized gas, composed of neutral atoms, charged particles (molecules and ionized

species), and electrons. The ionization rate of this plasma, defined by the ratio $(n^e/(n^e + n^o))$, where n^e and n^o are the electron and the neutral atom densities respectively, highlights the importance of energy transfer between the plasma and the material. The characteristics of this plasma and notably the ranges in temperature and concentrations of the various species present in the plasma thereby depend not only on the nature and composition of the anode but also on the energy transferred.

One of the advantages of this synthesis technique is the ability to vary a large variety of parameters which modify the composition of the high-temperature medium and consequently allow the most relevant plasma characteristics to be determined so that optimal conditions for the control of CNT formation can be obtained. However, a major drawback of this technique – and of any other technique used to produce SWNTs – is that the SWNTs formed are not pure: they are associated with other carbon phases and remnants of the catalyst. Although purification processes exist, they are all based on oxidation processes (e.g., HNO_3, or heat treatments in air) that are likely to significantly affect the SWNT structure [8.12]. Subsequent thermal treatments at $\approx 1200\,^{\circ}C$ under inert atmosphere, however, succeed in somewhat recovering structural quality [8.37].

Experimental Device

Several elements such as iron, nickel, cobalt, yttrium, boron, gadolinium, or cerium can be used as catalysts, however highest yields were obtained using a bimetallic catalyst, including a transition metal (typically Ni) and a rare earth (typically yttrium); the inert gas atmosphere can be composed of argon or (preferably) helium. Some work was also performed with liquid media, such as liquid nitrogen [8.38] or water [8.39, 40], but this adaptation has remained prospective. The arc burns between the catalyst-doped graphite anode and a graphite (or copper, for instance) cathode: the plasma formed consists of the mixture of carbon vapor, inert gas, and catalyst vapors. The vaporization is the consequence of energy transfer from the arc to the anode. The importance of the anode erosion rate depends on the power of the arc and also on other experimental conditions. It is worth noting that a high anode erosion rate does not necessarily lead to a high CNT production.

An example of a reactor layout is shown in Fig. 8.9. It consists of a cylinder about 30 cm in diameter and about 1 m in height with two valves, one for performing the primary evacuation (0.1 Pa) of the chamber, the other for filling it with a rare gas up to the desired working pressure. Arc observation is possible through diametrically opposed sapphire windows facing the plasma zone. Carbon nanoforms are found in different regions of the reactor:

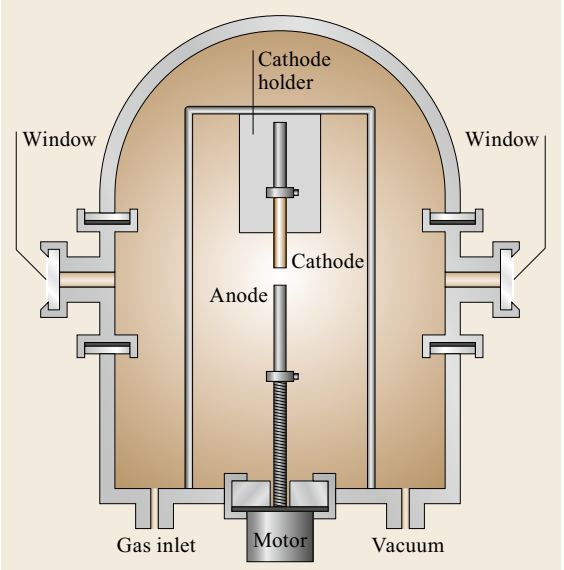


Fig. 8.9 Sketch of an electric-arc reactor

1. The collaret, which forms around the cathode
2. The weblike deposits found above the cathode
3. The soot deposited all around the reactor walls and the bottom
4. A hard deposit adherent to the cathode, whether catalysts are used or not.

SWNTs are found in the collaret and the web, yet their formation is not systematic and depends on the experimental conditions (for instance they do not form with a pressure of 40 kPa) while the cathode deposit and soot are obtained systematically. The catalyst-free growth of MWNTs occurs in the cathode deposit.

The electrode diameter is a few millimeters but varies from one author to another. In some cases, the cathode has a larger diameter than the anode in order to facilitate their alignment [8.28]. Other authors utilize electrodes of the same diameter [8.37]. The whole device can be designed horizontally [8.29, 36] or vertically [8.31–33]. The advantage of the latter is the symmetry with respect to gravity, which facilitates computer modeling (regarding convection flows, for instance).

Two types of anode can be used, a graphite rod with a coaxial hole (several centimeters in length) filled with a mixture of powdered catalyst(s), graphite, and possibly other dopants as needed, or an anode within which the catalyst(s) are homogeneously dispersed [8.41]. The former is by far the most popular, due to their ease of fabrication.

The role of synthesis parameters is studied for process optimization, in terms of nanotube yield and quality. They include the type of doped anode, the nature and concentration of the catalyst(s), the pressure and nature of the plasmagen gas, the arc current intensity, or the distance between electrodes. Experiment duration, however, is preferably short (in the range of a few minutes) in order for the synthesis not to be significantly affected by the increasing pressure rise resulting from anode vaporization. The plasma characteristics (species concentrations and temperature) constitute the link between input parameters and obtained products. Their study as a function of operating parameters is important for better understanding of the phenomena involved during nanotube formation. However, few works are dedicated to this; they are based on atomic and molecular optical emission spectroscopy [8.31–34, 36].

Results

The numerous results available in the literature regarding the synthesis of CNTs by the DC arc process show that both the morphology and yield strongly depend on the experimental conditions. It is also worth noting that the products obtained do not consist solely of CNTs. In addition to catalyst remnants found in various concentrations in all deposit areas, nontubular forms of carbon can be obtained [8.30, 32, 33]. They include nanoparticles (NPs), fullerene-like structures including C_{60}, poorly organized polyaromatic carbons, nearly amorphous nanofibers, multiwall shells, single-wall nanocapsules, and amorphous carbon. The nature, proportion, and distribution of the catalysts in the anode are particularly important. Optimized conditions for high SWNT yields (with lengths in the micrometer range and diameters typically around 1.4 nm, mostly found in the collaret (50–70%) and the web ($\approx$ 50% or less)) may correspond to a helium pressure of about 60 kPa, an arc current of 80 A, an electrode gap of 1 mm, the use of a heterogeneous anode and Ni/Y as coupled catalysts [8.5, 29, 42], and graphite powder grain size in the 1 μm range [8.43, 44]. Deviations from those conditions can have significant consequences: using Ni/Co instead of Ni/Y as catalysts may prevent the formation of SWNTs; Ni/Co catalysts in a homogeneous instead of heterogeneous anode mostly generate double-wall nanotubes (DWNTs), while decreasing the ambient pressure from 60 to 40 kPa also suppresses nanotube formation. It is worth noting that, unlike SWNTs, MWNTs are formed as a cathode deposit almost systematically, i. e., it is less sensitive to parameter variations. However, although using a rare earth as a single catalyst does not allow one to grow SWNTs, it may affect MWNT growth by promoting the closure of graphene edges, the preferred formation of telescope-like morphology, and the occurrence of nested crystals (Gd, in the paper cited) [8.31, 33]. Generally speaking, the optimized result achieved using bimetallic catalysts such as Ni+Y regarding SWNT yield was proposed once to be due to the transitory formation of an yttrium carbide coating over the nickel particles whose lattice constants were somewhat commensurable with that of graphene, thereby resulting in SWNT growth [8.45]. However, another explanation, more plasma-related, could be that the presence of yttrium increases the anode erosion rate and leads to higher plasma temperatures, which both are plasma features favorable to SWNT growth.

Surprisingly, loading heterogeneous anodes with diamond instead of graphite (with similar 1 μm range grain size) does not affect the plasma characteristics (both graphite and diamond lead to the same plasma composition once carbon is vaporized at high temperature, > 4000 K), nor the type and yield of the products formed [8.34, 35], while changing graphite grain size from 1 to 100 μm does ([8.43, 44] and Fig. 8.10a). This shows that thermal conductivity prevails over electrical conductivity as a leading parameter, and is driven by the porosity amount left when compacting the filler materials into the anode cavity. Indeed, lower grain size led to better compaction and powder homogeneity, hence to a steadier erosion of the anode, a feeding of the plasma with more constant carbon and catalyst proportions which goes along with smoother radial temperature profiles (see the previous version of this chapter in [8.21]). The consequence on the product is a lower variability in the carbon phases formed, finally resulting in an enhanced purity and yield of the SWNT phase.

Finally, the chamber volume was also shown to be an important parameter for SWNT growth [8.46]. The volume must be limited to keep convection cooling efficient and sustain a high thermal radial gradient which is favorable to SWNT growth. However, the volume has to remain sufficient to ensure recirculation of catalyst particles responsible for growing SWNTs. Therefore, Fig. 8.10 shows that lower chamber volume and smaller grain size both generate hotter plasma and higher C_2 concentration, a set of conditions found to correspond to optimized SWNT growth.

The analysis of the plasma from the anode to the cathode along the discharge axis shows a huge vertical gradient (≈ 500 K mm) close to it (probably due to convection) along with a dramatic increase of the C_2 concentration [8.21]. The latter demonstrates that C_2 moieties are secondary products resulting from the recombination of primary species formed from the anode. It also suggests that C_2 moieties may also be the building blocks for MWNTs (formed at the cathode) [8.33, 35]. A plausible reason is that the zone of

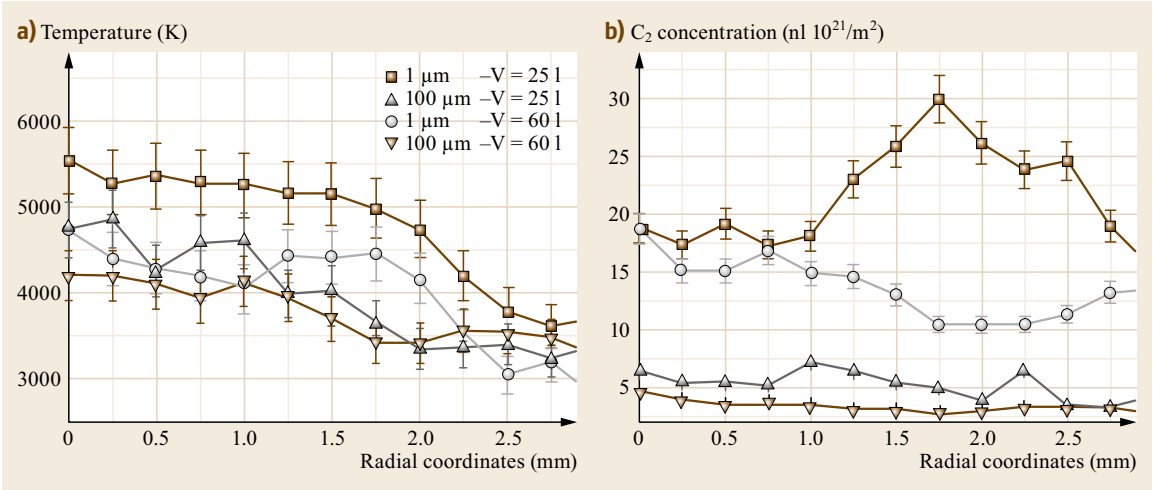


Fig. 8.10a,b Emission spectroscopy data obtained from experiments involving heterogeneous-type anodes, with two different graphite grain sizes (1 and 100 µm) and two different chamber volumes (25 and 60 l). **(a)** Radial temperature profiles; **(b)** C_2 radial concentration ratio. Smaller grain size results in better anode compaction, hence in higher anode thermal conductivity, and lower chamber volume enables more efficient convection cooling (after [8.46])

actual SWNT formation corresponds to colder areas, outside the plasma. Figure 8.11 shows an example of the evolution with time of the gas temperature in such an area.

Overall, the importance of the physical phenomena (charge and heat transfers) that occur in the anode during the arc clearly makes the characterization of the plasma indispensable, an aspect which is still neglected in many works.

Although many aspects still need to be understood, the electric-arc method is one of three methods currently used to produce SWNTs as commercial products, although it is progressively being abandoned due to the difficulty to reduce production costs down to that of chemical vapor deposition (CVD)-based processes, because of the subsequent extensive purification needed and the drawback of being a discontinuous process. For instance, Nano-lab (MA, USA) was proposing until recently (2016) 40–90%-purified SWNTs at prices ranging from 225 to 2500 US $/g.

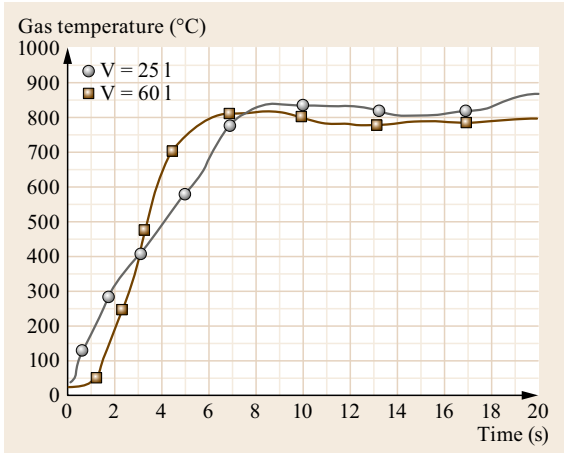


Fig. 8.11 Evolution with time of the gas temperature at the limit of the growth zone (1 cm away from the electrode axis and 1 cm above the arc center) for a graphite grain size of 1 µm and two chamber volumes. A smaller volume induces a lower temperature ramp thanks to an enhanced convection cooling, which seems to be favorable to SWNT formation

8.2.2 Gaseous Carbon Source–Based Synthesis Techniques

As mentioned in [8.1], the catalyst-enhanced thermal cracking of a gaseous carbon source (hydrocarbons, CO) – commonly referred to as catalytic chemical vapor deposition (CCVD) – has been known to produce carbon nanofilaments for over a century. However, in early works, carbon nanofilaments were mainly produced to act as a core substrate for the subsequent growth of larger (micrometric) carbon fibers – so-called vapor-grown carbon fibers – via thickening in catalyst-free CVD processes [8.47, 48].

Instead, later attempts in the 1990s focused on the synthesis of genuine CNTs (either single- or multi-walled) by CCVD. This method involves the catalytic decomposition of a carbon-containing source on small metallic particles or clusters, either through a heterogeneous (if a solid substrate is involved) or homogeneous

process (if everything takes place in the gas phase). The metals generally used for these reactions are transition metals, such as Fe, Co, and Ni. CNTs prepared by this method are generally much longer (a few tens to hundreds of micrometers) than those obtained by arc discharge (a few micrometers). It is also possible to grow dense arrays of nanotubes. CCVD processes appear very sensitive to the nature and the structure of the catalyst used, as well as to the operating conditions [8.49]. This is a rather low-temperature process typically between 600–1000 °C, which enables high selectivity for the production of MWNTs preferably to graphitic particles and amorphous-like carbon, compared to the solid carbon source-based methods such as arc discharge. In turn, this lower reaction temperature does not allow any structural rearrangements therefore leading to more structural defects (lower nanotexture) for MWNTs produced by CCVD compared to arc discharge. These defects can be removed by subsequent heat treatments in vacuum or inert atmosphere. Whether such a discrepancy is also true for SWNTs remains questionable as CCVD SWNTs are generally gathered into bundles of smaller diameter (a few tens of nm) than their arc discharge and laser ablation counterparts (around 100 nm in diameter). CCVD, specifically when performed in a fluidized bed reactor [8.50] (see next section), provides reasonably good prospects for use in large-scale and low-cost processes for the mass production of CNTs, a key point for their application at the industrial scale.

Heterogeneous Processes

Heterogeneous CCVD processes simply involve passing a gaseous flow containing a given proportion of a hydrocarbon (mainly CH_4, C_2H_2, C_2H_4, or C_6H_6, usually as a mixture with either H_2 or an inert gas such as Ar) over small transition metal particles (Fe, Co, Ni) in a furnace. The particles are deposited onto an inert substrate, by spraying a suspension of the metal particles on it or by another method. The reaction

$$C_xH_y \rightarrow xC + \frac{y}{2}H_2$$

is chemically defined as catalyst-enhanced thermal cracking and has been used since the late nineteenth century. Extensive publications on this topic include those by *Baker* et al. [8.3, 51], or *Endo* et al. [8.52, 53] as well as several review papers published after 1990, such as [8.54]. However, CO can be used instead of hydrocarbons; the reaction is then chemically defined as catalyst-enhanced disproportionation (the so-called Boudouard equilibrium)

$$2CO \rightleftharpoons C + CO_2 \,.$$

The ability of catalyst-enhanced CO disproportionation to make carbon nanofilaments was reported by *Davis* et al. [8.55] as early as 1953, probably for the first time. Extensive follow-up work was performed by *Boehm* [8.56], *Audier* et al. [8.14, 57, 58], and *Gadelle* et al. [8.59, 60].

Experimental Devices. Although formation mechanisms for SWNTs and MWNTs can be quite different (Sect. 8.3, or refer to a review article such as [8.61]), many of the catalytic process parameters play similar roles in the type of nanotubes formed: the temperature, the duration of the treatment, the gas composition and flow rate, and of course the catalyst nature and size. At a given temperature, depending mainly on the nature of the catalyst and the carbon-containing gas, catalytic decomposition will take place at the surface of the metal particles, followed by mass transport of the freshly produced carbon by surface or volume diffusion until carbon concentration exceeds the solubility limit causing precipitation to start.

CCVD CNTs form on very small metal particles, typically in the nanometer range [8.61]. These catalytic metal particles are prepared mainly by reducing transition metal compounds (salts, oxides) by H_2 prior to the nanotube formation step (where the carbon-containing gas is required). It is possible to produce these catalytic metal particles in situ in the presence of the carbon source, allowing for a one-step process [8.62]. As controlling the metal particle size is the key issue, coalescence is generally avoided by using low concentrations of catalytic metal precursors or by placing metal particles on an inert support such as an oxide (Al_2O_3, SiO_2, zeolites, $MgAl_2O_4$, MgO) or more rarely on graphite. The easiest hence most frequent route is to use the supported catalysts as a static phase placed within the gas flow. This is what most laboratories do, and also companies such as Nanocyl, the only CNT manufacturer in Belgium, which sells MWNTs (research grade NC3100, 95% C, $\approx$ 10 nm average diameter) at 5 €/g for 500 g order (2016 data). The CCVD process can also be coupled to a fluidization contactor by using catalytic porous powders to maximize the catalyst specific surface area.

In the fluidized bed (FB) reactor, gas-solid fluidization consists in flowing a gas upwards through a vertical bed of powders which form a suspension in the gas (Fig. 8.12). Gaseous bubbles appear and rise through the powders, generating intense gas–solid mixing and then high thermal and mass transfer rates between the gas, the powders, and the heated reactor walls [8.63]. As a consequence, a fluidized bed is fully isothermal. When coupled to CCVD, it maximizes the conversion rate of the gaseous carbon precursor, the catalyst

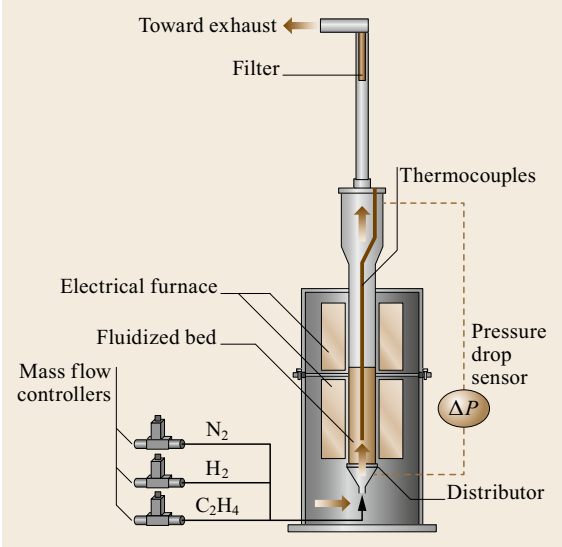


Fig. 8.12 Sketch of a typical lab-scale fluidized bed CCVD reactor as in use in the LGC-CNRS laboratory in collaboration with the LCC-CNRS laboratory, Toulouse (France). The *distributor* is a perforated plate supporting the powders when at rest

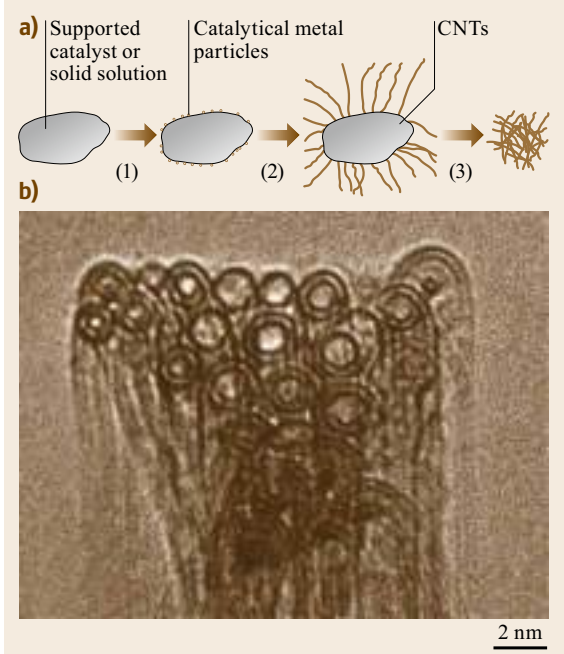


Fig. 8.13 **(a)** Formation of nanotubes via the CCVD-based impregnation technique. (1) Formation of catalytic metal particles by reduction of a precursor; (2) catalytic decomposition of a carbon-containing gas, leading to the growth of CNTs; (3) removal of the catalyst to recover the nanotubes (after [8.65]). **(b)** Example of a bundle of double-wall nanotubes (DWNTs) prepared this way (after [8.66])

selectivity, and the product uniformity. As the FB-CCVD process allows kilograms to tons of MWNTs to be produced from the lab scale to the industrial scale [8.64], it was selected by ARKEMA, the only CNT manufacturer in France, which is able to sell MWNTs (Graphistrength C100 grade) at $\approx 50\,€/\mathrm{kg}$ (2016 data).

There are two main ways to prepare the catalyst:

1. Impregnation of a substrate with a salt solution of the desired metal catalyst
2. Preparation of a solid solution of an oxide of the chosen catalytic metal in a chemically inert and thermally stable host oxide.

The catalyst is then reduced to form the metal particles on which the catalytic decomposition of the carbon source will lead to CNT growth. In most cases, the nanotubes can then be separated from the catalyst (Fig. 8.13).

Results with CCVD Involving Impregnated Catalysts. The impregnation-based process generally involves four successive steps:

1. Support impregnation by a salt solution (nitrate, chloride) of the metal catalyst
2. Drying and calcination of the supported catalyst to form the oxide of the catalytic metal
3. Reduction in a H_2-containing atmosphere to form the catalytic metal particles
4. Decomposition of a carbon-containing gas over the freshly prepared metal particles leading to nanotube growth.

For example, *Ivanov* et al. [8.67] prepared nanotubes through the decomposition of C_2H_2 (pure or mixed with H_2) on well-dispersed transition metal particles (Fe, Co, Ni, Cu) supported on graphite or SiO_2. Co-SiO_2 was found to be the best catalyst/support combination for the preparation of MWNTs, but most of the other combinations led to carbon filaments, sometimes covered with amorphous-like carbon. In order to improve the dispersion of metals, the same group developed a precipitation-ion-exchange method on silica then proposed the use of a zeolite-supported Co catalyst [8.68, 69], resulting in very finely dispersed metal particles (from 1 to 50 nm diameter). In the latter work, they reported the preparation of MWNTs with a diameter around 4 nm and only two or three walls. More recently, ethanol solutions containing a $NiCl_2$ salt were used to impregnate a carbon fiber paper. After drying

and reduction under H_2 at 1023 K, randomly orientated MWNTs with a diameter of ≈ 20 nm were grown on the paper from a C_2H_4/H_2 mixture, thus forming 3-D conductive layers for electrochemical capacitor applications [8.70].

Results with CCVD Involving Solid Solution–Based Catalysts. A solid solution of two metal oxides is formed when ions of one metal mix with ions of the other metal. For example, Fe_2O_3 can be prepared in solid solution in Al_2O_3 to give a $Al_{2-2x}Fe_{2x}O_3$ solid solution. The use of a solid solution allows homogeneous dispersion of one oxide in the other to be obtained. These solid solutions can be prepared in different ways; coprecipitation of mixed oxalates and combustion synthesis are the most common methods used to prepare nanotubes. Nanotube synthesis by catalytic decomposition of CH_4 over a $Al_{2-2x}Fe_{2x}O_3$ solid solution was first developed by *Peigney* et al. [8.62] and then studied extensively by the same group using different oxides such as spinel-based solid solutions ($Mg_{1-x}M_xAl_2O_4$ with M = Fe, Co, Ni, or a binary alloy [8.65, 71]) or magnesia-based solid solutions [8.65, 72] ($Mg_{1-x}M_xO$, with M = Fe, Co, or Ni). The homogeneous dispersion of the catalytic oxide makes it possible to produce very small catalytic metal particles at the high temperature required for CH_4 decomposition (chosen for its greater thermal stability compared to other hydrocarbons). This method involves heating the solid solution from room temperature to a temperature between 850–1050 °C in a mixture of H_2 and CH_4, typically containing 18 mol% of CH_4. The nanotubes obtained are typically gathered into small-diameter bundles (less than 15 nm) with lengths up to 100 µm. The nanotubes are mainly SWNTs and DWNTs, with diameters between 1–3 nm. Their properties depend upon the nature of the metal (or alloy) and the inert oxide (matrix); the latter because the Lewis acidity seems to play an important role [8.73]. For example, in the case of solid solutions containing around 10 wt% of Fe, the amount of CNTs obtained decreases in the following order depending on the matrix oxide: $MgO < Al_2O_3 < MgAl_2O_4$ [8.65].

Obtaining pure nanotubes by CCVD requires removing the catalyst (as for all other techniques). In the case of catalysts supported (impregnated) in a solid solution, the supporting – and catalytically inactive – oxide is the main impurity, both in weight and volume. Removing oxides such as Al_2O_3 or SiO_2 requires aggressive treatments (hot caustic solutions of KOH and NaOH for Al_2O_3, or an acidic solution of HF for SiO_2). However, such treatments have no effect on other impurities such as other carbon forms (amorphous-like carbon, graphitized carbon particles and shells, and so on). Oxidizing treatments (air oxidation, use of strong oxidants such as HNO_3, $KMnO_4$, H_2O_2) are thus required but inevitably result in low final yield of CNTs, which are often quite damaged. *Flahaut* et al. were the first to use MgO-based solid solutions (for example $Mg_xCo_yO_z$) to prepare SWNTs and DWNTs that could be easily separated without incurring any damage via fast and safe washing with an aqueous HCl solution [8.72].

In most cases, only very small quantities of catalyst (typically less than 500 mg) are used, and most claims for high-yield productions of nanotubes are based on laboratory experimental data, without taking into account all of the technical problems related to scaling up to a laboratory-scale CCVD reactor. At the present time, although the production of MWNTs is possible on an industrial scale, the production of affordable SWNTs is still a challenge, and controlling the arrangement of and the number of walls in the nanotubes is also problematic. For example, adding small amounts of molybdenum to the catalyst [8.74] can lead to drastic modifications of the nanotube type (from regular nanotubes to carbon nanofibers – Sect. 8.1). *Flahaut* et al. have recently shown that the method used to prepare a particular catalyst can play a very important role [8.75]. Double-walled CNTs (DWNTs) represent a special case: they are at the frontier between single- (SWNTs) and multiwalled nanotubes (MWNTs). Because they are the MWNTs with the lowest possible number of walls, their structures and properties are very similar to those of SWNTs. Any subsequent functionalization, which is often required to improve the compatibility of nanotubes with their external environment (composites) or to give them new properties (solubility, sensors), will partially damage the external wall, resulting in drastic modifications in terms of both electrical and mechanical properties. This is a serious drawback for SWNTs. In the case of DWNTs, the outer wall can be modified (functionalized) while retaining the structure of the inner tube. CCVD allows synthesis of DWNTs on a gram-scale [8.66] with high purity and high selectivity (around 80% DWNTs) (Fig. 8.13b).

Homogeneous Processes

The homogenous route, also called the *floating catalyst method*, differs from the other CCVD-based methods because it only uses gaseous species as catalyst and carbon precursors. The basic principle of this technique, whose mechanisms are similar to the other CCVD processes, is to produce in situ the catalyst particles from a gaseous precursor and then to decompose a gaseous carbon source in order to obtain CNTs.

Experimental Devices. The typical reactor used in this technique is a quartz tube placed in a furnace into

which the gaseous feedstock, containing the metal precursor, the carbon source, sometimes hydrogen, and a vector gas (N_2, Ar, or He), is passed. The first zone of the reactor is kept at a lower temperature to produce the catalyst particles, and the second zone is heated to 700–1200 °C to form the CNTs. The metal precursor is generally a metal-organic compound, such as a zero-valent carbonyl compound like [$Fe(CO)_5$] [8.76], or a metallocene [8.77–79] such as ferrocene, nickelocene, or cobaltocene. In recent works, ferrocene is diluted in ethanol [8.80] or in benzene [8.81] and the liquid solution is vaporized just before or at the entrance of the reactor to form a gaseous mixture containing both the metal and the carbon precursors. The use of metal salts, such as cobalt nitrate, has also been reported [8.82].

Results. This technique is quite flexible and SWNTs [8.80, 81], DWNTs [8.83], and MWNTs [8.84] have been obtained, in proportions depending on the carbon feedstock gas. The technique has also been exploited for some time in the production of vapor-grown carbon nanofibers [8.85].

The main drawback of this process concerns the difficulty to control the size of the metal nanoparticles, hence the CNT formation is often accompanied by the production of undesired carbon forms (amorphous carbon or polyaromatic carbon phases found as various phases or as coatings). In particular, encapsulated forms have been found as the result of the formation of metal particles that are too large to promote nanotube growth. The same operating parameters as for heterogeneous CCVD processes must be controlled in order to selectively obtain the desired CNTs morphology and structure, including the choice of the carbon source, the reaction temperature [8.86], the composition of the incoming gaseous feedstock [8.77], the presence and proportion of hydrogen, which can influence the orientation of the graphene with respect to the nanotube axis, thus switching from c-MWNT to h-MWNT [8.60].

The overall process can be improved by adding other compounds such as ammonia or sulfur-containing species to the reactive gas phase. The former allows aligned nanotubes and mixed C–N nanotubes [8.87] to be obtained, while the latter results in a significant increase in productivity [8.85, 88]. An interesting result is the increase in yield and purity brought about by a small input of oxygen, as achieved by using alcohol vapors instead of hydrocarbons as feedstock [8.89]. It is assumed that the oxygen (i) preferably burns the poorly organized carbon out into CO_2, thereby enhancing the purity; (ii) prevents the catalyst particles from being encapsulated in carbon shells too early (making them inactive), thereby enhancing the nanotube yield. Moreover, it was found to promote the formation of SWNTs over MWNTs, because suppressing the carbon shell formation limits the formation of MWNTs too.

It should be emphasized that only small amounts of CNTs have been produced so far, and scaling-up to industrial levels seems quite difficult. One reason is that the in situ produced catalyst particles stick to the reactor wall so that CNT growth from these particles may lead to clogging of the reactor [8.80]. Another bottleneck is to be able to increase the quantity of gaseous precursors fed in the reactor in order to increase production, without obtaining catalyst particles that are too big. This problem has not yet been solved.

A significant breakthrough concerning this technique could be the HiPco process developed at Rice University, which produces SWNTs of high purity [8.90, 91]. The current production yields approach 450 mg/h and nanotubes typically have no more than 7 mol% of iron impurities. In the HiPco process, nanotubes grow in high-pressure, high-temperature flowing CO on catalytic clusters of iron. Catalyst is formed in situ by thermal decomposition of $Fe(CO)_5$, which is delivered within a cold CO flow and then rapidly mixed with hot CO in the reaction zone [8.91]. Upon heating, the $Fe(CO)_5$ decomposes into atoms that condense into large clusters. SWNTs nucleate and grow on these particles via CO disproportionation (the Boudouard reaction, see Sect. 8.2.2, *Heterogeneous Processes*).

The company Continental Carbon Nanotechnologies Inc. (Houston, USA) currently sells raw SWNT materials prepared by this process, at a market price of 575 $/g, or 800 $/g if purified (2016 data). Other companies that specialize in MWNTs include Applied Sciences Inc. (Cedarville, USA), which currently has a production facility of ≈ 40 t/yr of ≈ 100 nm large MWNTs (Pyrograf-III), and Hyperion Catalysis (Cambridge, USA), which produces MWNT-based materials. Though prepared in a similar way by CCVD-related processes, MWNTs remain far less expensive than SWNTs, yet over a large price range depending on the manufacturer, the CNT type (herringbone, concentric etc.) and quality, and the amount to be ordered. Hence, prices can be as low as ≈ 0.5 $/g (2016 average ASI fare for Pyrograf-III grade), although this can be cut by a factor of 10 for very large quantities.

Templating

The templating technique is the only other method aside from the electric-arc technique that is able to synthesize CNTs without a catalyst. Another useful aspect of this approach is that it allows aligned nanotubes to be obtained naturally, provided the substrate is not removed.

The principle of this technique is to deposit the solid carbon coating obtained from the CVD method onto the walls of a porous substrate whose pores are arranged in

parallel channels. The substrate can be alumina or zeolite for instance, which present natural channel pores, while the whole system is heated to a temperature that cracks the hydrocarbon molecules selected as the carbon source [8.92].

Provided the chemical vapor deposition mechanism is well controlled, the synthesis results in the channel pore walls being coated with a variable number of graphenes. Either MWNTs (exclusively concentric type) or SWNTs can be obtained. The smallest SWNTs (diameters ≈ 0.4 nm) ever obtained were actually synthesized using this technique [8.9]. The nanotube lengths are directly determined by the channel lengths, in other words by the thickness of the substrate plate. One main advantage of the technique is the purity of the tubes (no catalyst remnants and few other carbon phases), and their well-calibrated, single-value diameters. Conversely, the nanotube structure is open at one end or both, which can be an advantage or a drawback depending on the application, and make them more sensitive to chemical attacks. The porous matrix can be kept or dissolved in order to recover the tubes, as needed. The method is not suitable for mass production.

8.2.3 Miscellaneous Techniques

In addition to the major techniques described in Sect. 8.2.1 and Sect. 8.2.2, other ways to produce nanotubes can be found in the literature. They can offer alternatives such as a low-cost or a catalyst-free production. However, none has been convincing enough or economically preferable so far with respect to the major processes described previously. Some examples are provided in the following, the three first sharing similarities (high temperature and solid carbon source) with the dc arc process.

For the laser ablation process a graphite pellet containing catalysts is vaporized using a focused laser beam [8.93]. The carbon species, swept along by a flow of neutral gas, are then deposited as soot in different regions: on a conical water-cooled copper collector, on the containment quartz tube walls, and on the backside of the pellet. Improvements of this method include the use of a second pulsed laser to get better pellet irradiation [8.94] or the use of two pellets (one made of graphite and one made of catalysts) simultaneously irradiated [8.95]. Laser-based methods can lead to high-quality SWNTs but are generally not considered to be competitive in the long term for the low-cost production of SWNTs.

Three-phase AC arc plasma has been used for the processing of carbon nanomaterials [8.96]. An electric arc is established between three graphite electrodes, the three-phase AC power supply being operated at 600 Hz and with arc currents of 250–400 A. Carbon precursors, gaseous, liquid or solid, are injected at the desired (variable) position into the plasma zone. It has been demonstrated that this plasma technology can be used to produce a wide range of carbon nanostructures ranging from carbon blacks to CNTs over fullerenes with a high product selectivity.

In solar furnace devices, solar rays are focused onto a graphite pellet containing catalysts in a neutral gas chamber. They have been used by several groups to first produce fullerenes but the original devices were then modified to achieve CNT production [8.97, 98]. It was demonstrated that solar energy-based synthesis is a versatile method for obtaining SWNTs that could be scaled up from 0.1–0.2 to 10 g/h and then to 100 g/h productivity using existing solar furnaces. A numerical reactor simulation was performed in order to improve the quality of the product, which was subsequently observed to reach 40% SWNT in the soot [8.99].

MWNTs (including coiled MWNTs, a peculiar morphology resembling a spring) have been successfully prepared by a catalyst-free (although Li was present) electrolytic method, by running a 3–5 A current between a graphite crucible (as anode) and a graphite rod (as cathode) [8.100]. The graphite crucible was filled with lithium chloride, while the whole system was heated in air or argon at $\approx 600\,^{\circ}$C. As with many other techniques, by-products such as encapsulated metal particles, carbon shells, amorphous carbon, and so on, are formed.

A pure chemistry route has also been proposed, using the polyesterification of citric acid onto ethylene glycol at 50 °C, followed by polymerization at 135 °C and then carbonization at 300 °C under argon, followed by oxidation at 400 °C in air [8.101]. Despite the latter oxidation step, the solid product contains short MWNTs, although they obviously have poor nanotextures. By-products such as carbon shells and amorphous carbon are also formed.

Short MWNTs have been obtained through a catalyst-free (although Si is present) pyrolytic method which involves heating silicon carbonitride nanograins in a BN crucible to 1200–1900 °C in nitrogen within a graphite furnace [8.102]. No details are given about the possible occurrence of by-products, but they are likely considering the complexity of the chemical system (Si-C-B-N) and the high temperatures involved.

The catalyzed reaction between a solid carbon source and atomic hydrogen has been investigated [8.103]. Graphite nanoparticles (≈ 20 nm) are sent with a stream of H_2 onto a Ta filament heated at 2200 °C. The species produced, whatever they are, then

hit a Si polished plate warmed to 900 °C that supports transition metal particles. The whole chamber is kept in a dynamic vacuum of 40 Torr. SWNTs are supposed to form according to the authors. One major drawback of the method, besides its complexity compared to the others, is that it is difficult to recover the nanotubes from the Si substrates to which they seem to be firmly bonded.

Diffusion flame synthesis was also attempted [8.104]. A regular gaseous hydrocarbon source (ethylene, and others) along with ferrocene vapor is passed into a laminar diffusion flame derived from air and CH_4 of temperature 500–1200 °C. SWNTs are formed, together with encapsulated metal particles, soot, and so on. In addition to a low yield, the SWNT structure is quite poor.

Finally, a recent and original technology should be mentioned, called the Graphetron 1.0 process, which is based on the decomposition of CO_2 into SWNTs by a high-temperature electrochemical reaction, likely assisted by an iron catalyst. The Russian company OCSiAl is claiming to be the world's largest facility with a 10 t/yr SWNT production complex, with prices as low as a few $/g for an SWNT-grade product (TUBALL) containing 75% SWNT, 15% catalyst, and 10% of other carbon forms, including carbon shells encapsulating the catalyst (2016 data).

8.2.4 Synthesis with Controlled Orientation

Several applications (such as field emission-based displays, see Sect. 8.7) require that CNTs grow as highly aligned individuals or bunches, in highly ordered arrays, or that they are located at specific positions. In this case, the purpose of the process is not mass production but controlled growth and purity, with subsequent control of nanotube morphology, texture, and structure. Generally speaking, the more promising methods for the synthesis of aligned nanotubes are based on CCVD processes (see Sect. 8.2.2). Although this approach initially produced mainly MWNTs, DWNT [8.105] and SWNT [8.106] arrays can be selectively obtained. Generally speaking, SWNTs and DWNTs nucleate at higher temperatures than MWNTs [8.107].

During CCVD growth, CNTs can self-assemble into nanotube bunches aligned perpendicularly to the substrate if the catalyst film on the substrate has a critical thickness [8.108, 109]. The driving forces for this alignment are the van der Waals interactions between the nanotubes, which allow them to grow perpendicularly to the substrates. When the catalyst nanoparticles are deposited onto a mesoporous substrate (such as mesoporous silica [8.110, 111] or anodic alumina [8.112]), the mesoscopic pores control the growth direction of the nanotubes (of course, removing the substrate may result in loosing the alignment).

Here is a list of the most common deposition methods:

1. Metal salt impregnation followed by oxidation/reduction steps for thin film deposition on alumina substrates [8.113].
2. Embedding catalyst particles in mesoporous silica by sol–gel processes [8.110].
3. Thermal evaporation of Fe, Co, Ni, or Co-Ni metal alloys on SiO_2 or quartz substrates under high vacuum [8.114, 115].
4. Photolithographic patterning of metal-containing photoresist polymer using conventional black and white films as a mask [8.116] or photolithography and inductive plasma deep etching [8.117].
5. Electrochemical deposition into pores in anodic aluminum oxide templates [8.112].
6. Spin-coating colloidal suspensions of catalyst particles with tailored diameters on a support [8.118–121].
7. Stamping a catalyst precursor over a patterned Si wafer has been used to grow networks of nanotubes parallel to the substrate (Fig. 8.14b), or more generally to localize the growth of individual CNTs [8.122]. Combining e-beam lithography and template methods has also been reported for large-scale production of ordered MWNTs [8.123] or atomic force microscope/microscopy (AFM) tips [8.124].

Depositing the catalyst nanoparticles onto a prepatterned substrate allows control of the frequency of local occurrence and the arrangement of the nanotube bunches formed. The materials produced mainly consist of arrayed, densely packed, freestanding, aligned MWNTs (Fig. 8.14a), which are suitable for field emission-based applications for instance [8.125]. SWNTs have also been produced, and it was reported that the introduction of water vapor during the CVD process allows impurity-free SWNTs to be synthesized [8.126], due to a mechanism related to that previously proposed for the effect of using alcohol instead of hydrocarbon feedstock [8.89].

The pyrolysis of hydrocarbons in the presence of organometallic precursor molecules such as metallocene or iron pentacarbonyl, operating in a dual furnace system (Fig. 8.15a,b) leads to densely packed vertically aligned MWNTs (Fig. 8.15c). The precursor is first sublimed at low temperature in the first furnace or injected as a solution along with the hydrocarbon feedstock, and then the whole system is pyrolyzed at higher temperature in the second furnace [8.83, 129, 130]. Important synthesis parameters include the heat-

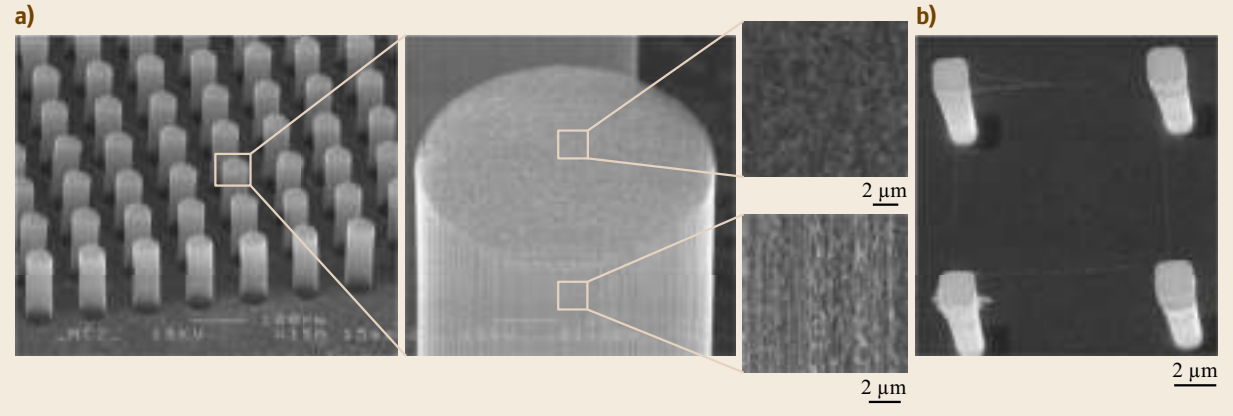


Fig. 8.14a,b Examples of directional growth of CNTs obtained from the pyrolysis of a gaseous carbon source over catalyst nanoparticles previously deposited onto a patterned substrate: **(a)** a controlled pattern of nanotube pillars grown perpendicular to the substrate: each pillar is made of thousands of DWNTs aligned vertically (after [8.127]); **(b)** suspended SWNTs grown parallel to the substrate interconnecting silicon pillars (after [8.128])

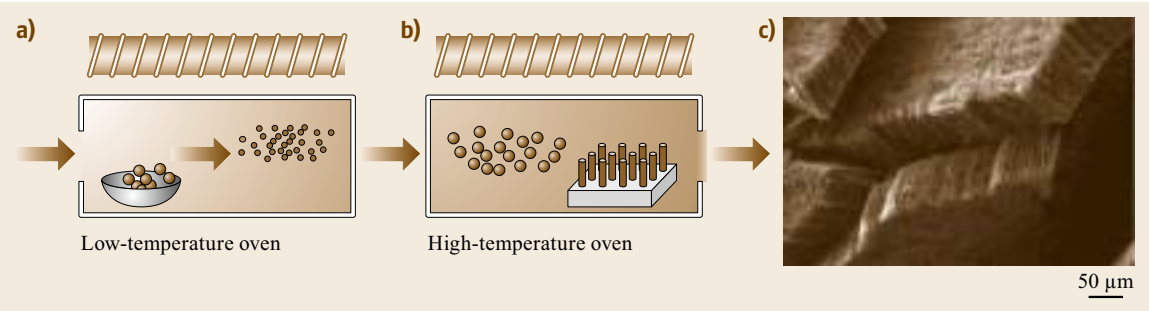


Fig. 8.15a–c Sketch of a double-furnace CCVD device used in the organometallic/hydrocarbon copyrolysis process. **(a)** Sublimation of the precursor. **(b)** Decomposition of the precursor and MWNT growth onto the substrate. **(c)** Example of the densely packed and aligned MWNT material obtained (after [8.129])

ing and feeding rate of the precursors, vector gas (Ar or N_2) and gaseous hydrocarbon flow rates, and the pyrolysis temperature (650–1050 °C). Using [$Fe(CO)_5$] as the catalyst source results in thermal decomposition at elevated temperatures, producing atomic iron that deposits on the substrates in the hot zone of the reactor. CNT growth occurs immediately after the introduction of [$Fe(CO)_5$] so the growth temperature depends mainly on the carbon feedstock used (from 750 °C for acetylene to 1100 °C for methane).

The nanotube yield and quality are directly linked to the amount and size of the catalyst particles, and since the planar substrates used do not exhibit high surface areas,the catalyst dispersion is a critical step in the process. NH_3-etching of the catalyst thin film provides the appropriate metal particle size distribution and may also favor the MWNT alignment and prevent the formation of amorphous carbon [8.131]. The use of phthalocyanines of Co, Fe, and Ni has also been reported, and in this case the organometallic precursors pyrolysis is also the carbon source for the vertically aligned MWNT growth [8.132].

Densely packed coatings of vertically aligned MWNTs may also be produced over metal-containing deposits, such as iron oxides on aluminum [8.133]. Well-aligned MWNT arrays can be produced on a large scale on ceramic spheres using the floating catalyst technique [8.134]. Finally, the Langmuir–Blodgett method can be used to produce densely packed monolayers of aligned noncovalently functionalized SWNTs from an organic solvent [8.135]. This method seems valid for bulk materials with various diameters and offers the advantage that the SWNT monolayers are readily patterned for device integration by microfabrication.

8.3 Growth Mechanisms of Carbon Nanotubes

The growth mechanisms of CNTs have been highly debated for the main reason that the conditions that allow carbon nanofilaments to grow are very diverse, which means that there are many related growth mechanisms. Another reason is that the phenomena that occur during growth are pretty rapid and difficult to observe in situ. It is generally agreed, however, that growth occurs such that the number of dangling bonds is minimized, for energetic reasons. Much related information can be found in review papers such as [8.136–138].

8.3.1 Catalyst-Free Growth

In addition to the templating technique, which is merely a chemical vapor infiltration mechanism for pyrolytic carbon, the growth of c-MWNT as a deposit on the cathode in the electric-arc method is a rare example of catalyst-free carbon nanofilament growth. The driving force is related to charge transfer from one electrode to the other via the particles contained in the plasma. It is not clear how the MWNT nucleus is formed, but once it has, it may include the direct incorporation of C_2 species into the primary graphene structure, as was previously proposed for fullerenes [8.139].

8.3.2 Catalytically Activated Growth

Although growth mechanisms involving catalysts include a more or less extensive contribution from a VLS [8.140] mechanism, it is quite difficult to find comprehensive and plausible explanations that are able to account for both the various conditions used and the various morphologies observed. What follows is an attempt to provide overall plausible explanations of most of the phenomena, while remaining consistent with the experimental data. From various results, it appears that the most important parameters are probably the thermodynamic ones (only temperature will be considered here), the catalyst particle size, and the presence of a substrate.

Low-Temperature Conditions

In CCVD, nanotubes are frequently found to grow far below 1000 °C, yet 800–1200 °C is the most common temperature range for growing well-structured nanotubes with the usual catalysts.

If the conditions are such that the catalyst is a crystallized solid, the nanofilament is probably formed via a mechanism similar to a VLS mechanism, in which three steps are defined:

1. Adsorption then decomposition of C-containing gaseous moieties at the catalyst surface
2. Dissolution then diffusion of the C species through the catalyst, thus forming a solid solution
3. Back-precipitation of solid carbon as nanotube walls. The texture is then determined by the orientation of the crystal faces relative to the filament axis [8.15, 16].

This mechanism can therefore provide either c-MWNT, h-MWNT, or platelet nanofibers. The latter, however, are mainly formed when catalyst particles are crystallized and large enough to exhibit fully developed crystal faces with extensions in the range of at least 40–50 nm onto which graphenes form and then pile up. Indeed, platelet nanofibers with low diameters (< 40 nm) have never been observed. The reasons for this are related to graphene energetics, such as the need to reach the optimal ratio between the amount of edge carbon atoms (with dangling bonds) and inner carbon atoms (where all of the σ and π orbitals are shared).

If conditions are such that the catalyst is a liquid droplet, due to the use of temperatures beyond the catalyst melting temperature, a mechanism similar to that described above can still occur, which is really VLS (vapor = gaseous C species, liquid = molten catalyst, S = graphenes), but there are no crystal faces to preferentially orient the rejected graphenes. Energy minimization requirements will therefore make them concentric and parallel to the filament axis. With large catalyst particles (or in the absence of any substrate), the mechanisms above will generally follow a *tip-growth* scheme: the catalyst moves forward while the rejected carbon forms the nanotube behind, whether there is a substrate or not. One end might be left open. Conversely, when the catalyst particles deposited onto the substrate are small enough to be held in place by interaction forces with the substrate, the growth mechanism will follow a *base-growth* scheme, where the carbon nanofilament grows away from the substrate, leaving the catalyst nanoparticle attached to it (Fig. 8.16).

The bamboo texture that affects both the herringbone and the concentric textures may reveal a distinguishing aspect of the dissolution-rejection mechanism: the periodic, discontinuous dynamics of the phenomenon. Once the catalyst has reached the saturation threshold in terms of its carbon content, it expels it quite suddenly. Then it becomes able to incorporate a given amount of carbon again without having any catalytic activity for a little while. Then over-saturation is reached again, and so on [8.141].

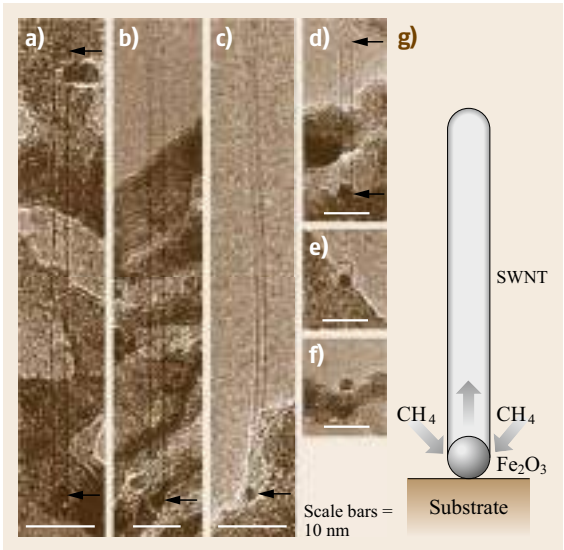


Fig. 8.16a–g High-resolution TEM images of several SWNTs grown from iron-based nanoparticles using the CCVD method, showing that particle sizes determine SWNT diameters in this case (after [8.142]). Yet the catalyst crystal imaged (the *dark spot* at the bottom of each tube) is different for each figure, considering images backward from (f) to (c) illustrates what could be a sequence of growth of an SWNT from a single nanocrystal, as sketched in (g). (a) and (b) show additional examples of fully grown SWNTs, similar to (c) ◄

Therefore, it is clear that 1 catalyst particle = 1 nanofilament in any of the mechanisms above. This explains why achieving a narrow distribution of nanotube diameters in CCVD is quite challenging, particularly when nanosizes are required for the growth of SWNTs. Only particles < 2 nm are useful for this (Fig. 8.16), since larger SWNTs are not favored energetically [8.8]. Another distinguishing aspect of the CCVD method is that growth can occur all along the isothermal zone of the reactor furnace as it is continuously fed with a carbon-rich feedstock, which is generally in excess, with a constant composition at a given species time of flight. Roughly speaking, the longer the isothermal zone, the longer the nanotubes. This is why the lengths of the nanotubes can be much longer than those obtained using solid carbon source-based methods.

The comments above are summarized in Table 8.1, compared with basic information relating to the high-temperature growth conditions (Sect. 8.3.2, *High-Temperature Conditions* below).

High-Temperature Conditions

High-temperature conditions (several thousands of °C) are typically used in solid carbon source-based methods such as the electric-arc method. Both the carbon source and the catalysts are atomized. Of course, catalyst-grown SWNTs do not form in the plasma, where temperatures are the highest whereas c-MWNTs do. The plasma medium is a mixture of atoms and radicals,

Table 8.1 Guideline indicating the relationships between possible carbon nanofilament morphologies and some basic synthesis conditions. Columns (1) and (2) mainly relate to CCVD-based methods; column (3) mainly relates to plasma-based methods

	Increasing temperature … → … and physical state of catalyst			Substrate		Thermal gradient	
Catalyst particle size	Solid (crystallized) (1)	Liquid from melting (2)	Liquid from condensing atoms (3)	Yes	No	Low	High
$\lesssim 3$ nm	SWNT	SWNT	–	Base growth	Tip growth	Long length	Short length
$\gtrsim 3$ nm	MWNT (c,h,b) platelet nanofiber	c-MWNT	SWNT	Tip growth	Tip growth	Long length	Short length
	Nanotube diameter						
	Heterogeneous related to catalyst particle size		Homogeneous (independent of particle size)		[a]		
	Nanotube/particle						
	One nanotube/particle		Several SWNTs/particle		[a]		

[a] Except for SWNTs growing from case (3) catalyst

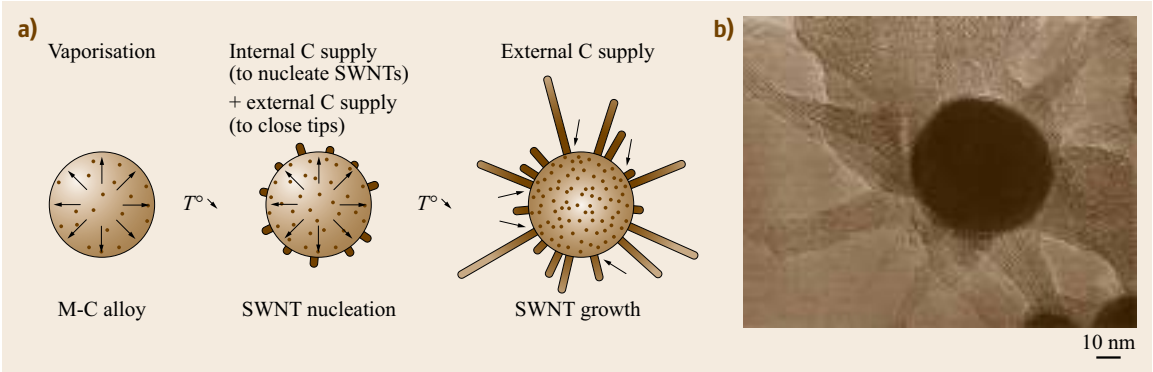


Fig. 8.17 **(a)** Mechanism proposed for SWNT growth in high-temperature conditions (see text). **(b)** TEM image of SWNT growing radially from the surface of a large Ni catalyst particle in an electric-arc experiment (after [8.18])

some of which are likely to combine and condense into liquid droplets. At some distance from the atomization zone, the medium is therefore made of carbon-metal alloy droplets and of secondary carbon species that range from C_2 to higher order molecules such as corannulene, which is made of a central pentagon surrounded by five hexagons (as the fastest way to limit dangling bonds at low energetic cost), thereby providing a fixation site for other carbon atoms (or C_2) which also will tend to close into a ring. Since adjacent pentagons are not energetically favored, these cycles will be hexagons. Such a molecule is thought to be a probable precursor for fullerenes. Fullerenes are actually always produced, even in conditions that produce SWNTs. The same saturation in C described in Sect. 8.3.2 occurs for the carbon-metal alloy droplets as well, resulting in the precipitation of excess C outside the particle due to the effect of the decreasing thermal gradient in the reactor, which decreases the solubility threshold of C in the metal [8.143]. Once the *inner* carbon atoms reach the surface of the catalyst particle, they meet the *outer* carbon species, including corannulene, that will contribute to capping the merging nanotubes. Once formed and capped, nanotubes can grow both from the inner carbon atoms (Fig. 8.17a), according to the VLS mechanism proposed by *Saito* et al. [8.143], and from the outer carbon atoms, according to the adatom mechanism proposed by *Bernholc* et al. [8.144]. In the latter, carbon atoms from the surrounding medium in the reactor are attracted then stabilized by the carbon/catalyst interface at the nanotube/catalyst surface contact, promoting their subsequent incorporation at the tube base. The growth mechanism therefore mainly follows the base growth scheme. The occurrence of a nanometer-thick surface layer of yttrium carbide (onto the main Ni-containing catalyst core), the lattice distance of which is commensurable with that of the C−C distance in graphene, could possibly play a beneficial role in stabilizing the nanotube/catalyst interface, which could explain why the SWNT yield is enhanced by bimetallic alloys (as opposed to single metal catalysts) [8.41]. But this needs the catalyst to be crystallized during CNT growth, which is not ascertained. For instance, as SWNTs only grow as bundles, it was proposed that those could result from the occurrence of a cell structuration phenomenon of the liquid catalyst according to the so-called solutal Bénard–Marangoni instability mechanism [8.145]. Hence, alternatively, Y could mostly play a major role in providing the plasma with the favorable thermal properties.

A major difference from the low-temperature mechanisms described for CCVD methods is that many nanotubes are formed from a single, relatively large ($\approx$ 10−50 nm) catalyst particle (Fig. 8.17b). This is why the diameters of SWNTs grown at high temperature are much more homogeneous than those associated with CCVD methods. The reason that the most frequent diameter is $\approx$ 1.35 nm is again a matter of energetic compromise. Larger SWNTs are less stable [8.8]. Conversely, the strain on the C−C bond increases as the radius of curvature decreases. Another difference from the low-temperature mechanism for CCVD is that temperature gradients in high-temperature methods are huge (Fig. 8.10), and the gas-phase composition surrounding the catalyst droplets is also subjected to rapid changes. This explains why nanotubes from arc are generally shorter than nanotubes from CCVD, and why mass production by CCVD is favored.

8.4 Properties of Carbon Nanotubes

In MWNTs that consist of stacked graphenes, the bond strength varies significantly depending on whether the in-plane direction (characterized by very strong covalent and therefore very short – 0.142 nm – bonds) or the direction perpendicular to it (characterized by very weak van der Waals and therefore very loose – ≈ 0.34 nm – bonds) is considered. Such heterogeneity is not found in single SWNTs. However, the heterogeneity returns, along with the related consequences, when SWNTs associate into bundles. Therefore, the properties – and applicability – of SWNTs may also change dramatically depending on whether single SWNTs or SWNT ropes are involved.

In the following, we will emphasize the properties of SWNTs, since their unique structures often lead to different properties to regular polyaromatic solids. However, we will also sometimes discuss the properties of MWNTs for comparison.

8.4.1 Overall Properties of SWNTs

With diameters in the nanometer range, and lengths in the range of micrometer to hundreds of micrometers, SWNTs exhibit huge aspect ratios. They are stable up to 600–700 °C in air and up to ≈ 1500–1800 °C in an inert atmosphere, beyond which they transform into regular graphene-based solids [8.146]. They have half the mass density of aluminum. The properties of SWNTs, like any molecule, are heavily influenced by the way their atoms are arranged, hence by their helicity [8.147].

8.4.2 Adsorption Properties

Accessible Specific Surface Area of CNTs

Various studies dealing with the adsorption of nitrogen onto MWNTs [8.148–150] and SWNTs [8.150, 151] have highlighted the porous nature of both materials. CNTs present large surface areas because of their structure and physical form. Their actual surface areas may be influenced by a variety of properties such as tube/fiber diameter, bundling and agglomeration, purity, and surface functionalization [8.150]. The theoretical surface area of CNTs has a broad range, from 50–1315 $m^2\,g^{-1}$ depending on the number of walls, the diameter, and the number of nanotubes in a bundle of CNTs [8.152]. Experimentally, the surface area of an SWNT is often larger than that of an MWNT. The total surface area of as-grown SWNTs is typically between 400–900 $m^2\,g^{-1}$ (micropore volume 0.15–0.3 $cm^3\,g^{-1}$), whereas values of 200–400 $m^2\,g^{-1}$ for as-produced MWNTs are often reported (Table 8.2).

However, the mesopores of the SWNT, obtained by subtracting the micropore volume from the total, constitute the major fraction of the pores [8.154]. The pores in MWNTs can be divided mainly into hollow inner cavities with small diameters (with narrow size distributions, mainly 3–10 nm) and aggregated pores (with wide size distributions, 20–40 nm), formed by interactions between isolated MWNTs. It is worth noting that opening or closing the central canal significantly influences the adsorption properties of CNTs. Opening, closing, or cutting CNTs, as well as heating them can considerably affect their surface area and pore structure. For instance, activating MWNTs with KOH or NaOH was shown to promote microporosity and reaching surface areas as high as 1050 $m^2\,g^{-1}$ [8.155, 156].

Adsorption Sites and Binding Energy of the Adsorbates

It is expected that adsorption properties of small-diameter CNTs such as SWCNTs are strongly affected by tube diameter and helicity, as the electronic structure is. Generally speaking, three types of sites exist on a graphene lattice: top (T), bridge (B), and center (C) (Fig. 8.18a). Additionally, the possibility to adsorb species on the convex (exohedral adsorption) or concave (endohedral adsorption) surface should be taken into account [8.157]. For SWNT bundles, adsorption may occur onto the various types of surface available (Fig. 8.18b). For MWCNTs, adsorption can occur in the aggregated pores, inside the tube or on the external walls; in this latter case, the presence of structural defects (vacancies, Dienes/Stone–Wales defects, dopants) plays a significant role in the adsorption as they can be seen as islands with elevated reactivity. Additionally, modeling studies have shown that the CNT convex (outer) surface is chemically reactive because the convex arrangement of pyramidalized sp^2 carbon atoms makes the formation of chemical bonds with reagent species possible. For a related reason, the concave (inner) surface should be more inert and can withstand the presence of highly reactive species encapsulated within the nanotubes. This difference in reactivity between convex and concave surfaces increases as the tube diameter decreases [8.158, 159].

Generally speaking, the adsorbates can be either charge donors or acceptors to the nanotubes. Trends in the binding energies of gases with different van der Waals radii suggest that the groove sites of SWNTs are the preferred low-coverage adsorption sites due to their higher binding energies (Table 8.2). In the first stages of adsorption (occurring at the most attractive sites), it

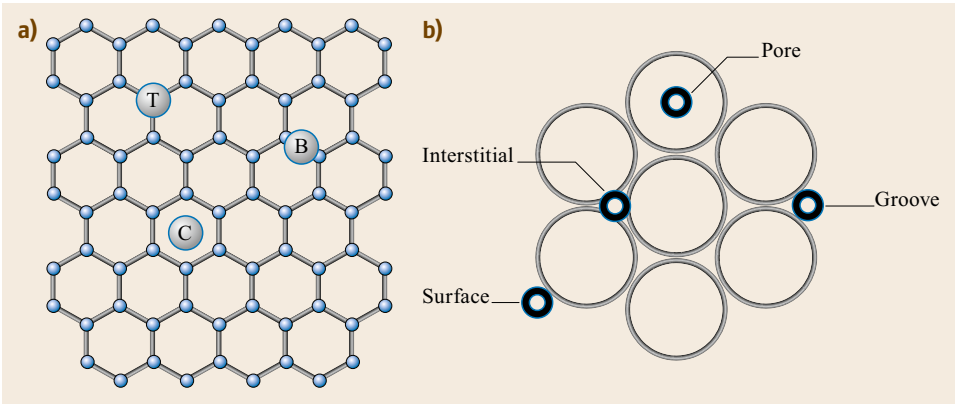


Fig. 8.18 **(a)** Adsorption sites on graphene; **(b)** sketch of an SWNT bundle cross-section, illustrating the four different adsorption sites

Table 8.2 Adsorption properties and sites of SWNTs and MWNTs. The data in the last two columns are from [8.153]

Type of nanotube	Porosity (cm^3/g)	Surface area (m^2/g)	Binding energy of the adsorbate	Adsorption sites	Attractive potential per site (eV)	Surface area per site (m^2/g)
SWNTs (bundle)	Microporous (0.15–0.3)	400–900	Low, mainly physisorption (25–75% higher than graphene)	Surface Groove Pore Interstitial	0.049 0.089 0.062 0.119	483 22 783 45
MWNT	Mesoporous	200–400	Physisorption	Surface pore Aggregated pores	–	–

seems that adsorption or condensation in the interstitial channels of the SWNT bundles depends on the size of the molecules (and/or on the SWNT diameters) and on their interaction energies [8.160–162]. For hydrogen and other small molecules like CO, computational methods have shown that, for open SWNTs, the pore, interstitial and groove sites are energetically more favorable than surface sites [8.163, 164]. Quantum chemical calculations suggest that adsorption in interstitial channels of bundles formed of large-diameter SWCNTs is possible and can account for high desorption barriers, a result of strong dispersion interactions between neighboring SWCNTs [8.165]. Opening CNTs favors gas (including O_2, N_2) adsorption within the inner walls [8.166, 167]. Finally, several studies have shown that, at low coverage, the binding energy of the adsorbate on an SWNT is between 25–75% higher than the binding energy on a single graphene (Table 8.2). This discrepancy can be attributed to an increase of effective coordination at the binding sites in SWNT bundles [8.168, 169]. Only a few studies deal with adsorption sites in MWNTs, but it has been shown that butane adsorbs more onto MWNTs with smaller outside diameters, which is consistent with another statement that the strain on curved graphene surfaces affects sorption. Most of the butane adsorbs to the external surface of the MWNTs while only a small fraction of the gas condenses in the pores [8.170].

A limited number of theoretical as well as experimental studies on the binding energies of gases onto CNTs exist. While most of these studies report low binding energies on SWNTs, consistent with physisorption, some experimental results, in particular for hydrogen, are still controversial. The energetics of hydrogen chemisorption on a (6,6) nanotube has been modeled using gradient-corrected density functional theory [8.171]. The chemical reaction of hydrogen with carbon nanotubes is expected to become increasingly exothermic and also to proceed more rapidly at high coverages. Oxygen gas could be physisorbed via dispersive van der Waals forces or chemisorbed via formation of a chemical bond, and adsorption could take place either on perfect nanotube walls or at defect sites [8.172]. The adsorption of single atoms on a semiconducting and metallic SWNT was studied from first principles for a large number of foreign atoms [8.173]. The stable adsorption sites, binding energy, and the resulting electronic properties were analyzed. The character of the bonding and associated physical properties exhibit dramatic variations depending on the type of adsorbed atom. While the atoms of good conducting metals, such as Cu and Au, form very weak bonding, atoms such as Ti, Sc, Nb, and Ta are adsorbed with relatively high binding energy. Adsorption geometries and binding affinities of metal nanoparticles onto CNTs were also investigated through DFT (density functional the-

ory) calculations [8.174]. Pd, Pt, and Ti nanoparticles strongly chemisorb to the CNT surface. Unlike the cases of atomic adsorptions, the aluminum particle has the weakest binding affinity with the CNT. Aluminum or gold nanoparticles accumulated on the CNT develop the triangular bonding network of the metal surfaces in which the metal–carbon bond is not favored.

8.4.3 Electronic and Optical Properties

The electronic states in SWNTs are strongly influenced by their 1-D cylindrical structures. One-dimensional sub-bands are formed that have strong singularities in the density of states (Van Hove singularities) [8.175]. By rolling the graphene sheet to form a tube (n, m), new periodic boundary conditions are imposed on the electronic wavefunctions, which give rise to one-dimensional sub-bands: $C_\mathrm{h}\,\mathrm{K} = 2q$ where q is an integer and C_h is the helicity vector defined in Sect. 8.1. Much of the electronic band structure of an SWNT can be derived from the electronic band structure of graphene by applying the periodic boundary conditions of the tube considered. The graphene conduction and valence bands touch each other at six corners (K points) of the Brillouin zone [8.176]. If one of these sub-bands passes through the K point, the nanotube is metallic; otherwise it is semiconducting. This is a unique property that is not found in any other 1-D system, which means that for certain orientations of the honeycomb lattice with respect to the tube axis (chirality), some nanotubes are semiconducting and others are metallic. The band gap for semiconducting tubes is found to be inversely proportional to the tube diameter. As pointed out in Sect. 8.1, knowing (n, m) allows – in principle – prediction of whether the tube is metallic or not. However, the gap energy of semiconducting SWNTs decreases as their diameter increases, whatever their (n, m) values, up to a value of ≈ 14 nm at which point the gap energy approaches the energy range brought by the room temperature [8.10]. Hence MWNTs with diameters beyond this value are no longer semiconducting and exhibit properties similar to other forms of regular, polyaromatic solids. It has been shown that electronic conduction mostly occurs through the external tube for MWNTs; even so, interactions with internal tubes often cannot be neglected and they depend upon the helicity of the neighboring tubes [8.177].

The electronic and optical properties of SWNTs are considerably influenced by the environment [8.178], starting with the bundle configuration as compared to isolated. In such a case, due to the interactions between neighboring tubes, a bundle of identical, metallic SWNTs can behave as a semiconducting object [8.179]. When multiple-wall tubes such as DWNTs are subjected to external pressure, the small interaction between the tube walls results in the internal tubes experiencing reduced pressure [8.180]. The electronic transition energies are in the infrared and visible spectral range. The 1-D Van Hove singularities have a large influence on the optical properties of CNTs. Visible light is selectively and strongly absorbed, which can lead to the spontaneous burning of agglomerated SWNTs in air at room temperature [8.181]. Strong Coulomb interaction in quasi-one dimension leads to the formation of excitons with very large binding energies in CNTs (200–400 meV), and degenerated states at the K, K′ points lead to multiple exciton states with dipole-allowed (bright) and dipole-forbidden transitions (dark) [8.182, 183]. Photoluminescence can be observed in individual SWNT aqueous suspensions stabilized by the addition of surfactants. Detailed photoexcitation maps provide information about the helicity-dependent transition energies and the electronic band structures of CNTs [8.184]. The agglomeration of SWNTs into bundles influences their electronic states. Photoluminescence signals are quenched for bundles.

CNTs are model systems for the study of 1-D transport in materials. Apart from the singularities in the density of states, electron–electron interactions are expected to show drastic changes at the Fermi edge; the electrons in CNTs are not described by a Fermi liquid, but instead by a Luttinger liquid model [8.185] which describes electronic transport in 1-D systems. It is expected that the variation of electronic conductance versus temperature follows a power law, with zero conductance at low temperatures. Depending on how L_ϕ (the coherence length) on the one hand and L_m (the electronic mean free path) on the other hand compare to L (the length of the nanotube), different conduction modes are observed: ballistic if $L \ll L_\phi$, $L \ll L_\mathrm{m}$, diffusive if $L_\phi \ll L_\mathrm{m} < L$, and localization if $L_\mathrm{m} \ll L_\phi \ll L$. Fluctuations in the conductance can be seen when $L \approx L_\phi$. For ballistic conduction (a small number of defects) [8.186–188], the predicted electronic conductance is independent of the tube length. The conductance value is twice the quantum of conductance (the fundamental conductance unit) $G_0 = 4e/h$ due to the existence of two propagating modes. Because of the reduced electron scattering observed for metallic SWNTs and their stability at high temperatures, CNTs can support high current densities (max. $10^9\,\mathrm{A/cm^2}$): about three orders of magnitude higher than Cu. Structural defects can, however, lead to quantum interference of the electronic wave function, which localizes the charge carriers in 1-D systems and increases resistivity [8.185, 189, 190]. Localization and quantum interference can be strongly influenced by applying a magnetic field [8.191]. At low temperatures,

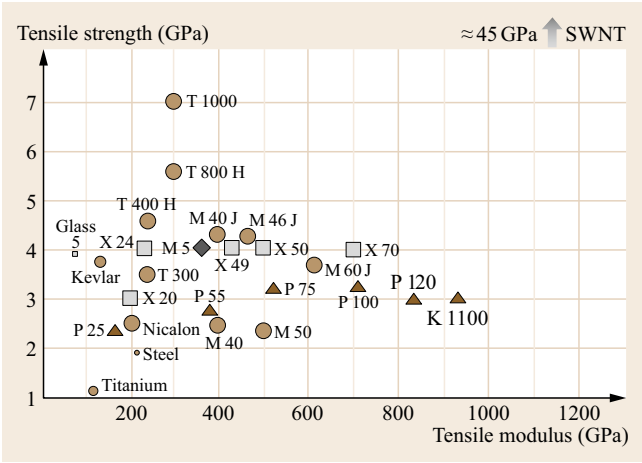


Fig. 8.19 Plot of the tensile strength versus the tensile modulus for various fibrous materials compared to SWNTs. *Large circles* are PAN-based carbon fibers, *triangles* are pitch-based carbon fibers

the discrete energy spectrum leads to a Coulomb blockade resulting in oscillations in the conductance as the gate voltage is increased [8.190]. In order to observe the different conductance regimes, it is important to consider the influence of the electrodes where Schottky barriers are formed. Palladium electrodes have been shown to form excellent junctions with nanotubes [8.192]. The influence of superconducting electrodes or ferromagnetic electrodes on electronic transport in CNTs due to spin polarization has also been explored [8.193, 194].

As a probable consequence of both the small number of defects (at least the kind of defects that oppose phonon transport) and the cylindrical topography, SWNTs exhibit a large phonon mean free path, which results in a high thermal conductivity. The thermal conductivity of SWNTs is comparable to that of a single, isolated graphene layer or high purity diamond [8.195], or possibly higher ($\approx 6000\,\mathrm{W\,m^{-1}\,K^{-1}}$).

8.4.4 Mechanical Properties

CNTs are unique due to the particularly strong bonding between the carbons (sp^2 hybridization) of the curved graphene sheet, which is stronger than in diamond (sp^3 hybridization; 0.142 versus 0.154 nm for the C−C bond length in graphene and diamond respectively). This makes CNTs – SWNTs or c-MWNTs – particularly stable against deformations. The tensile strength of SWNTs can be 20−100 times that of steel [8.196] (while being 6 times lighter) and has actually been measured as ≈ 45 GPa [8.197]. Very high tensile strength values are also expected for ideal (defect-free) c-MWNTs, since combining perfect tubes concentrically is not supposed to be detrimental to the overall tube strength, provided the tube ends are well capped (otherwise, concentric tubes could glide relative to each other, inducing high strain). Tensile strength values as high as ≈ 150 GPa have actually been measured for perfect MWNTs from an electric arc [8.198]. The flexural modulus of perfect MWNTs should logically be higher than that for SWNTs [8.196], with a flexibility that decreases as the number of walls increases. Conversely, measurements performed on defective MWNTs obtained from CCVD exhibit a range of 3−30 GPa [8.199]. Values of tensile modulus are also the highest values known, in the range of 1−1.3 TPa for both SWNTs and MWNTs [8.200–202]. Figure 8.19 illustrates how defect-free CNTs could spectacularly revolutionize the field of high mechanical performance fibrous materials. However, making micrometer-sized fibers out of carbon nanotubes while maintaining the ultimate CNT properties remains an unaddressed challenge due to the problem of creating covalent bonds between the CNTs that have to be as strong as that of the CNT C–C bonds without creating defects in CNT structure [8.203].

8.4.5 Reactivity

CNTs have a large surface to interact with their environment (Sect. 8.4.2, *Accessible Specific Surface Area of CNTs*). Although SWNTs belong to the graphene-based material family, unlike graphite, they have no (chemically active) dangling bonds, and unlike fullerenes, the ratio of *weak* sites (C−C bonds involved in heterocycles) over strong sites (C−C bonds between regular hexagons) only deviates slightly from 0 whereas it is 1 for C_{60} fullerenes. Although graphene faces are chemically relatively inert, the radius of curvature imposed on the graphene in nanotubes causes the three normally planar sp^2 C−C bonds to undergo distortions, resulting in bond angles that are closer to the sp^3 C−C bond angle in diamond as the radius of curvature shortens. Even though it is not enough to make the carbon atoms chemically reactive, one consequence is that nesting sites are created at the concave surface and physisorption sites are created on the convex surface, both with an adsorption efficiency that increases as the nanotube diameter decreases.

The chemical reactivities of SWNTs (and c-MWNTs) are believed to derive mainly from the caps, since they contain six pentagons each, as opposed to the tube body, which supposedly only contains hexagons. Indeed, applying oxidizing treatments to CNTs (air oxidation, wet-chemistry oxidation) selectively opens the nanotube tips [8.204]. However, that SWNTs can be filled with foreign molecules such as fullerenes (Sect. 8.5.2) with ≈ 90% filling rate suggests the occurrence of side defects (typically Dienes defects, oth-

erwise wrongly named Stone–Wales defects [8.205]) able to be chemically attacked [8.206, 207], thereby creating additional entry ports along the tube. An average of one Dienes defect every 5 nm along the tube length, involving about 2% of the carbon atoms in a regular (10,10) SWNT has been estimated [8.207]. This means that the overall chemical reactivity of CNTs should depend strongly on how they are synthesized. For example, SWNTs prepared by the arc-discharge method, because they grow very fast in nonsteady conditions, are likely to contain more structural defects than CCVD-synthesized SWNTs. It is however the contrary for arc-prepared c-MWNTs, which exhibit a perfect, defect-free nanotexture [8.20] as compared to CCVD-prepared c-MWNTs. Finally, the reactivity of h-MWNT-type nanotubes is intrinsically higher, due to the occurrence of accessible graphene edges at the nanotube surface.

8.5 Carbon Nanotube-Based Nano-Objects (Carbon Meta-Nanotubes)

Carbon nanotubes are beautiful objects exhibiting a variety of amazing properties from a theoretical point of view or when investigated at lab-scale. However, as soon as actual devices and materials have to be prepared, and moreover made compatible with upscaled processes, they more often than not need to be modified. Several modification routes have been distinguished [8.208], which are listed in the following sub-sections. As it is a very active, multiple-field research area, only the basis is provided here. Further information can be found in related chapters and review papers from the current literature, beginning with the book which was dedicated to it [8.209].

8.5.1 Heteronanotubes

It is possible to replace some or all of the carbon atoms in a nanotube with atoms of other elements – mostly boron and nitrogen [8.210], but also phosphorus [8.211] and possibly sulfur – without affecting the overall honeycomb lattice-based graphene structure. Nanotubes modified in this way are termed *heteronanotubes* [8.210]. They can exhibit new behaviors (for example, BN nanotubes are electrical insulators) or modified properties (e.g., enhanced resistance to oxidation, modulation of electronegativity). This is indeed a way to enable better control over such properties. For instance, one current challenge in carbon-SWNT synthesis is to obtain selective formation of the desired SWNT structure, as it drives the metallic or semiconductor character. In this regard, it was demonstrated that replacing some C atoms with N or B atoms leads to SWNTs with systematic metallic electrical behavior [8.212, 213]. Most examples of heteronanotubes found in the literature are MWNTs. The heteroatom usually involved is nitrogen, due to the ease with which gaseous or solid nitrogen- and/or boron-containing species (such as N_2, NH_3, BN, HfB_2) can be passed into existing equipment for synthesizing MWNTs [8.212, 214] until complete substitution of carbon occurs [8.215, 216]. An amazing result of such attempts to synthesize hetero-MWNTs is the subsequent formation of c-MWNTs made up of coaxial alternate carbon graphene tubes and boron nitride graphene tubes [8.217]. Conversely, examples of hetero-SWNTs are few, although the synthesis of B- or N-containing SWNTs has recently been reported [8.213, 218, 219]. Only one successful synthesis of genuine BN-SWNTs has been reported so far [8.220].

8.5.2 Filled Carbon Nanotubes

SWNTs or MWNTs can have their inner cavities filled (partially or entirely) with foreign atoms, molecules, compounds, or crystals [8.207, 221]. The terminology X@SWNT (or X@MWNT, if appropriate, where X is the atom, molecule and so on involved) has been used for such structures since 1999 [8.222]. The very small inner cavity of nanotubes is indeed an amazing tool for preparing and studying the properties of confined nanostructures of any type, such as salts, metals, oxides, gases, or even discrete molecules like C_{60}, for example. Because of the almost one-dimensional structure of CNTs (particularly for SWNTs), encapsulated materials might have different physical and/or chemical properties and stability with respect to the unencapsulated or bulk material. Indeed, if the volume available inside a CNT is small enough, the foreign material is mostly made of *surface atoms* with reduced coordination. Filling material/nanotube interactions can also result in the filled nanotube itself behaving differently from a *pure* nanotube. This the basis of the current efforts being made to develop SWNT-based cables exhibiting electrical conductivity reaching that of metals such as copper, thanks to interactions of the SWNT bundle with iodine and sulfuric acid molecules as *dopents* [8.223].

Because of the limitations of some desirable CNT filling materials, such as very high melting temperature or poor solubility, it is also possible to fill CNTs

with appropriate precursors (i. e., able to sublime, melt, or solubilize) which will later be transformed into the required material by a physical interaction mechanism such as electron beam irradiation for example [8.224], or by in situ chemical transformations such as reduction by H_2 or H_2S [8.225].

Generally speaking, it is difficult to estimate the filling rate, and this is usually achieved in published papers through TEM observations, although they are barely based on reliable statistics regarding the number of tubes observed. Moreover, as far as SWNTs are concerned, the fact that the nanotubes are gathered into bundles makes it difficult to observe the exact number of filled tubes, as well as to estimate the filled length for each tube. However, an estimation of filling rates can also be obtained from x-ray studies and Raman spectroscopy.

It has to be pointed out that the driving forces involved in filling events are not fully yet understood [8.207, 221].

In Situ Filling Method

Initially, most hybrid CNTs synthesized were based on MWNTs prepared using the electric-arc method, and were obtained directly during processing. The filling materials were easily introduced into the system by drilling a central hole in the anode and filling it with the heteroelement. The first hybrid products obtained using this approach were all reported in the same year [8.226–230] for heteroelements such as Pb, Bi, YC_2, and TiC. Later on, *Loiseau* and *Pascard* [8.231] showed that MWNTs could also be filled to several μm in length by elements such as Se, Sb, S, and Ge, but only for nanoparticles of elements such as Bi, B, Al, and Te. Sulfur was suggested to play an important role during the in situ formation of filled MWNTs using arc discharge [8.232]. This technique is no longer preferred because it is difficult to control the filling ratio and yield and to achieve mass production, and it is not suitable for filling SWNTs, nor for filling MWNTs with transition metals as this induces the formation of (empty) SWNTs instead. Conversely, significant in situ filling can occur during CVD-based processes, but only with the catalyst that helped with MWNT growth (most often Fe, but possibly Co and Ni as well).

Wet Chemistry Filling Method

For the wet chemistry method the nanotube tips must be opened by chemical oxidation prior to the filling step. This is generally achieved by refluxing the nanotubes in dilute nitric acid [8.233–235], although other oxidizing liquid media may work as well, such as [HCl $+CrO_3$] [8.236] or chlorocarbenes formed from the photolytic dissociation of $CHCl_3$ [8.207], a rare example of a nonacidic liquid route to opening SWNT tips. If a dissolved form (such as a salt or oxide) of the desired metal is introduced during the opening step, some of it will get inside the nanotubes. An annealing treatment (after washing and drying the treated nanotubes) may then lead to the oxide or to the metal, depending on the annealing atmosphere [8.204]. The wet chemistry method initially looked promising because a wide variety of materials can be introduced into nanotubes in this way and because it operates at temperatures that are not much different from room temperature, allowing the filling of CNTs with materials with low temperature stability. However, close attention must be paid to the oxidation method that is used. The damage caused to nanotubes by severe treatments (such as using nitric acid) makes them unsuitable for use with SWNTs. Moreover, the filling yield is not very good, obviously because the solvent molecules also enter the tube cavity, hence filled lengths rarely exceed 100 nm.

Molten-State Filling Method

The physical filling method involving a liquid (molten) phase is more restrictive, firstly because some materials can decompose when they melt, and secondly because the melting point must be compatible with the nanotubes, so the thermal treatment temperature should remain below the temperature of transformation or the nanotubes will be damaged. Because the filling occurs due to capillarity hence obeying Jurin's law, the surface tension of the filling material in the molten state is probably important, and a threshold of 100−200 N/cm^2 was proposed [8.237]. However, a related study investigated MWNTs, whose inner diameters (5−10 nm) are generally larger than for SWNTs (1−2 nm). In addition, a major problem is that related data are barely available in the literature, therefore a lot of uncertainties remain regarding the various factors affecting filling rates. In a typical filling experiment, the MWNTs are well mixed with the desired amount of filler by gentle grinding, and the mixture is then vacuum-sealed in a silica ampoule. The ampoule is then slowly heated to a temperature above the melting point of the filler and then slowly cooled. This method does not always require that the nanotubes are opened prior to heat treatment, as some molten filling materials (e.g., halides) are able to react with CNT tips and open them (definitely thanks to the presence of heterocycles – pentagons for example – which are more reactive).

Most of the works involving the application of this method to SWNTs come from Oxford University [8.238–242], although other groups have followed the same procedure [8.233, 235, 243]. As the precursors used were mainly metal salts, the crystallization of molten halides within small-diameter SWNTs has been

studied in detail, and the resulting one-dimensional crystals have been shown to interact strongly with the surrounding graphene wall. For example, *Sloan* et al. [8.240] described two-layer 4 : 4 coordinated KI crystals that formed within SWNTs that were ≈ 1.4 nm in diameter. These two-layer crystals were *all surface* and had no *internal* atoms. Significant lattice distortions occurred compared to the bulk structure of KI, where the normal coordination is 6 : 6 (meaning that each ion is surrounded by six identical close neighbors). Indeed, the distance between two ions across the SWNT capillary is 1.4 times as much as the same distance along the tube axis. This suggests an accommodation of the KI crystal into the confined space provided by the inner nanotube cavity in the constrained crystal direction (across the tube axis). This implies that the interactions between the ions and the surrounding carbon atoms are strong. The volume available within the nanotubes thus somehow controls the crystal structures of inserted materials. Taking the example of PbI_2, the structures and orientations of encapsulated PbI_2 crystals were found to depend on the diameter of the confining nanotubes [8.238]. For SWNTs, most of the encapsulated one-dimensional PbI_2 crystals obtained exhibited a strong preferred orientation, with their (110) planes aligning at an angle of around 60 °C to the SWNT axes, as shown in Fig. 8.20a,b. Because of the extremely small diameters of the nanotube capillaries, individual crystallites are often only a few polyhedral layers thick, as outlined in Fig. 8.20d–h. The unusual lattice terminations enforced by capillary confinement, cause the edging polyhedra to be of reduced coordination, as indicated in Fig. 8.20g,h. However, as the diameter of the encapsulating capillary increases, different preferred orientations were frequently observed [8.238].

High-yield filling of CNTs by the capillary method is generally difficult but fillings of more than 60% have been reported for different halides, with filling lengths of up to several hundred nm [8.244]. Results from the imaging and characterization of individual molecules and atomically thin, effectively one-dimensional crystals of rock salt and other halides encapsulated within SWNTs were reviewed by *Sloan* et al. [8.245].

Sublimation Filling Method

This method is more restrictive than the previous one, since it is only applicable to compounds able to sublimate within the temperature range of thermal stability of the nanotubes. Examples are therefore scarce. The first and still most successful example published so far is the formation of C_{60}@SWNT (nicknamed *peapods*), reported for the first time in 1998 [8.246], where regular ≈ 1.4 nm-large SWNTs were filled with C_{60} fullerene molecule chains. On the basis of this work, many fullerene-related molecules have been successfully encapsulated as well, e.g., C_{70} (Fig. 8.21a), or endofullerenes (La_2@C_{80} [8.247], Gd @C_{82} [8.248], and $Er_xSc_{3-x}N$@C_{80} [8.249] among other examples). Other attempts have followed, for instance involving $ZrCl_4$ [8.241], selenium [8.250], or iodine [8.243, 251] (Fig. 8.21b). Of course, the process requires that the SWNTs are opened by some method, as discussed previously, typically by either acid attack [8.252] or heat treatment in air [8.253]. The opened SWNTs are then in-

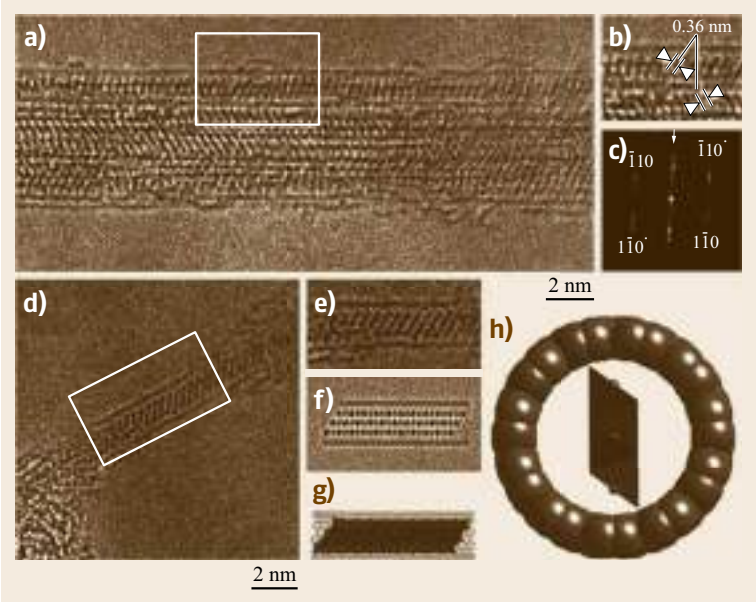


Fig. 8.20a–h High-resolution transmission electron microscope (HRTEM) images and corresponding structural model for PbI_2-filled SWNTs. **(a)** Image of a bundle of SWNTs, all of them being filled with PbI_2. **(b)** Enlargement of the portion framed in **(a)**. **(c)** Fourier transform obtained from **(b)** showing the 110 distances at 0.36 nm of a single PbI_2 crystal. **(d)** Image of a single PbI_2-filled SWNT. **(e)** Enlargement of the portion framed in **(b)**. **(f)** Simulated HRTEM image, corresponding to **(e)**. **(g)** Structural model corresponding to **(f)**. **(h)** Structural model of an SWNT filled with a PbI_2 crystal as seen in cross section (after [8.238])

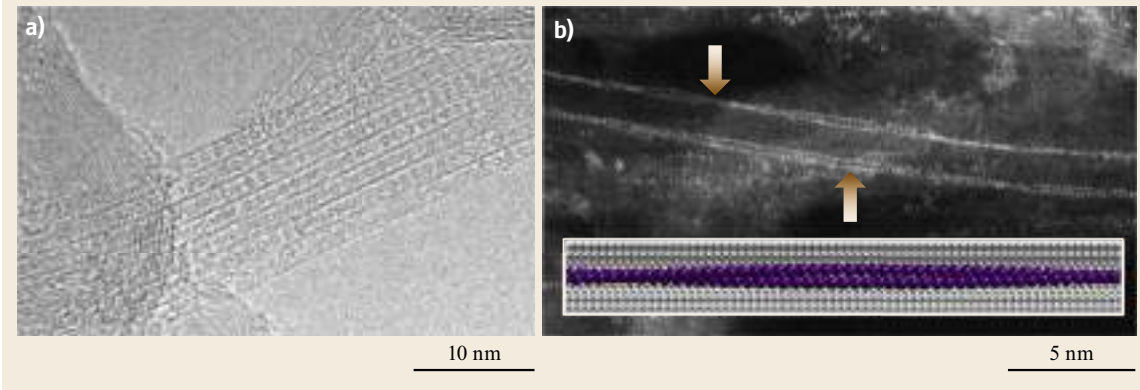


Fig. 8.21 **(a)** HRTEM image of an example of so-called peapods, here SWNTs filled with C_{70} molecules (image by L. Noé, CNRS). **(b)** Annular dark-field mode TEM image of an example of DWNTs filled with iodine chains. The specific imaging mode used prevents the carbon wall of the host nanotubes from being seen. Each white dot is a single iodine atom. The iodine chains are twinned and helically arranged in the DWNT cavities (image by C. Nie, CNRS). Inserted model by J. Sloan (University of Warwick)

serted into a glass or quartz tube together with fullerene powder, which is sealed and placed into a furnace heated above the sublimation temperature for fullerite ($> 350\,^{\circ}C$). As gaseous molecules are involved, the filling mechanism somewhat relates to capillary condensation instead of capillary wetting, meaning that there are no filling limitations related to Jurin's law or the presence of solvent. Consequently, filling efficiency may actually reach $\approx 100\%$ with this filling route [8.253].

8.5.3 Functionalized Nanotubes

Nanotube functionalization reactions can be divided into two main groups: (i) the chemical oxidation of the tips and structural defects leading to carboxylic, carbonyl, and/or hydroxyl functions. These functions are then used for additional reactions, to attach oligomeric or polymeric functional entities; (ii) the direct addition to the graphenic surface (without any intermediate step). Examples of the latter reactions include oxidation or fluorination (an important first step for further functionalization with other organic groups). The properties and applications of functionalized nanotubes have been reviewed in [8.254, 255]. It must be noted that these chemical reactions also affect any graphene-based impurities, making quantification difficult, either because raw yet pure CNT materials barely exist, or because purification procedures induce their own functionalization.

Oxidation of Carbon Nanotubes

CNTs are often oxidized and therefore opened before chemical functionalization in order to increase their chemical reactivity (by creating dangling bonds). The chemical oxidation of nanotubes is mainly performed using either wet chemistry or gaseous oxidants such as oxygen (typically air), ozone, or CO_2. Depending on the process they were synthesized with, the oxidation resistance of nanotubes can vary. When oxidation is achieved using a gas phase, thermogravimetric analysis (TGA) is of great help for determining the temperature at which the treatment should apply (attention should be paid to the use of low heating rates so that overestimated oxidation temperatures are not provided). Differences in the presence of catalyst remnants (metals or, more rarely, oxides) and other by-products, variations in the type of nanotube-oxidizing agent and TGA conditions (heating rate, flow rate) used make published results difficult to compare. It is generally agreed, however, that amorphous carbon burns first, followed by SWNTs and then multiwall materials (shells, MWNTs), even if TGA is often unable to separate the different oxidation steps clearly. Aqueous solutions of oxidizing reagents are often used for nanotube oxidation. The main reagent is nitric acid, either concentrated or diluted (around 3 mol/l in most cases), but oxidants such as potassium dichromate ($K_2Cr_2O_7$), hydrogen peroxide (H_2O_2), or potassium permanganate ($KMnO_4$) are often used as well. HCl, like HF, does not damage nanotubes because it is not oxidizing. However, the liquid wastes generated by the wet oxidation methods and the necessity of filtration/drying steps are not well suited for industrial use. Gas-phase oxidation of CNTs using either oxygen [8.227, 256] or nitric acid vapors [8.257] at moderate or high temperature removes disordered carbon but can lead to weight loss due to the etching of nanotube outer walls. As a consequence, high purity materials are obtained, but the yield is generally low. Conversely, gas-phase oxidation from ozone (O_3) per-

formed at ambient temperature is an environmentally friendly and low cost route to form oxygen-containing groups on the whole CNT surface without too much structural damage. It is suited for mass treatment since it can be performed in a fluidized bed [8.258, 259].

Functionalization of Oxidized Carbon Nanotubes

The defects on CNTs created by oxidants are stabilized by bonding with carboxylic acid (–COOH) or hydroxyl (–OH) groups. These functional groups have rich chemistry and the CNTs can be used as precursors for further chemical reactions, such as silanation, polymer grafting, esterification, thiolation, alkylation, arylation, and even some biomolecules. Oxidized nanotubes are usually reacted with thionyl chloride ($SOCl_2$) to generate the acyl chloride. The reaction of SWNTs with octadecylamine was reported by *Chen* et al. [8.260] after reacting oxidized SWNTs with $SOCl_2$. Many other reactions between functionalized nanotubes (after reaction with $SOCl_2$) and amines have been reported in the literature [8.261]. The common outcome of these functionalizations is the increased solubility of CNTs either in the organic solvent or water. Noncovalent reactions between the carboxylic groups of oxidized nanotubes and octadecylammonium ions are also possible [8.262], providing solubility in tetrahydrofuran (THF) and CH_2Cl_2. The attachment of proteins onto CNTs using diimide-activated amidation (through direct coupling of the carboxylic acid to proteins using a coupling agent) is possible [8.263]. DNA functionalization of CNTs was also reported [8.264].

Functionalization with lipophilic (long alkyl chains) and hydrophilic (oligomeric poly(ethyleneglycol) groups) dendra has been achieved via amination and esterification reactions [8.265]. In that case, the functional groups could be removed simply by modifying the pH of the solution [8.266]. The possible interconnection of nanotubes was reported by *Chiu* et al. [8.267] using the acyl chloride method and a bifunctionalized amine to link the nanotubes through the formation of amide bonds.

Finally, it has been discovered that imidazolium ion-functionalized CNTs are highly dispersible in ionic liquids of analogous chemical structure and that mixtures of functionalized CNT and ionic liquids can form gels upon sonication or waxes that could find applications as soft composite materials for electrochemistry [8.268].

Sidewall Functionalization of Carbon Nanotubes

There are two main approaches for the side-wall functionalization of CNTs: i) the covalent attachment of chemical groups through reactions onto the π-conjugated skeleton; and ii) the noncovalent adsorption or wrapping of various functional molecules [8.261, 269]. Covalent functionalization is based on the formation of a covalent linkage between functional entities and the carbon skeleton of nanotubes. It could also be divided into direct covalent sidewall functionalization and indirect covalent functionalization with carboxylic groups on the surface of CNTs (Fig. 8.22). Direct covalent sidewall functionalization is associated with a change in hybridization from sp^2 to sp^3 and a simultaneous loss of conjugation (e.g., fluorination of nanotubes [8.270, 271]). Indirect covalent functionalization takes advantage of chemical transformations of carboxylic groups at the open ends and holes in the sidewalls (see above). The drawback of covalent functionalization is that the perfect structure of CNTs has to be destroyed, resulting in significant changes in their physical properties. Noncovalent functionalization is mainly based on supramolecular complexation using various adsorption forces, such as Van der Waals force, hydrogen bonds, electrostatic force, and π-stacking interactions [8.269]. Compared to chemical functionalization, the advantages of noncovalent functionalization are that it can operate under relatively mild reaction conditions and that the perfect graphene structure of CNTs can be maintained.

Numerous chemical reactions including hydrogenation, halogenation, esterification, amidation, cycloaddition, radical addition, azide chemistry, nucleophilic addition, electrophilic addition, inorganic complex addition, ozonolysis, surface-initiated polymerization, and physical methods such as plasma treatment, π–π stacking and solid-phase mechanical milling have been tried to functionalize CNTs, offering alternative approaches to the solubilization of CNTs with desired functionality and surface characteristics [8.261, 269, 272–274].

8.5.4 Coated/Decorated Nanotubes

In contrast to functionalization, the decoration of CNTs consists in depositing solids, mainly nanoparticles (NPs), physi- or chemisorbed onto the nanotube's outer surface. Because of their interaction with the nanotube structure, the deposited metals, semiconductors, oxides, or organic materials often present unique properties [8.275]. Consequently, the coated CNTs find numerous applications in catalysis, electronics, energy conversion and storage, environment, and medicine [8.275, 276].

Even if the outer surface of tubes and tube ends present potential adsorption sites, numerous studies have shown that surface activation is essential to reach a high coating density. Oxidation (Sect. 8.5.3, *Oxida-*

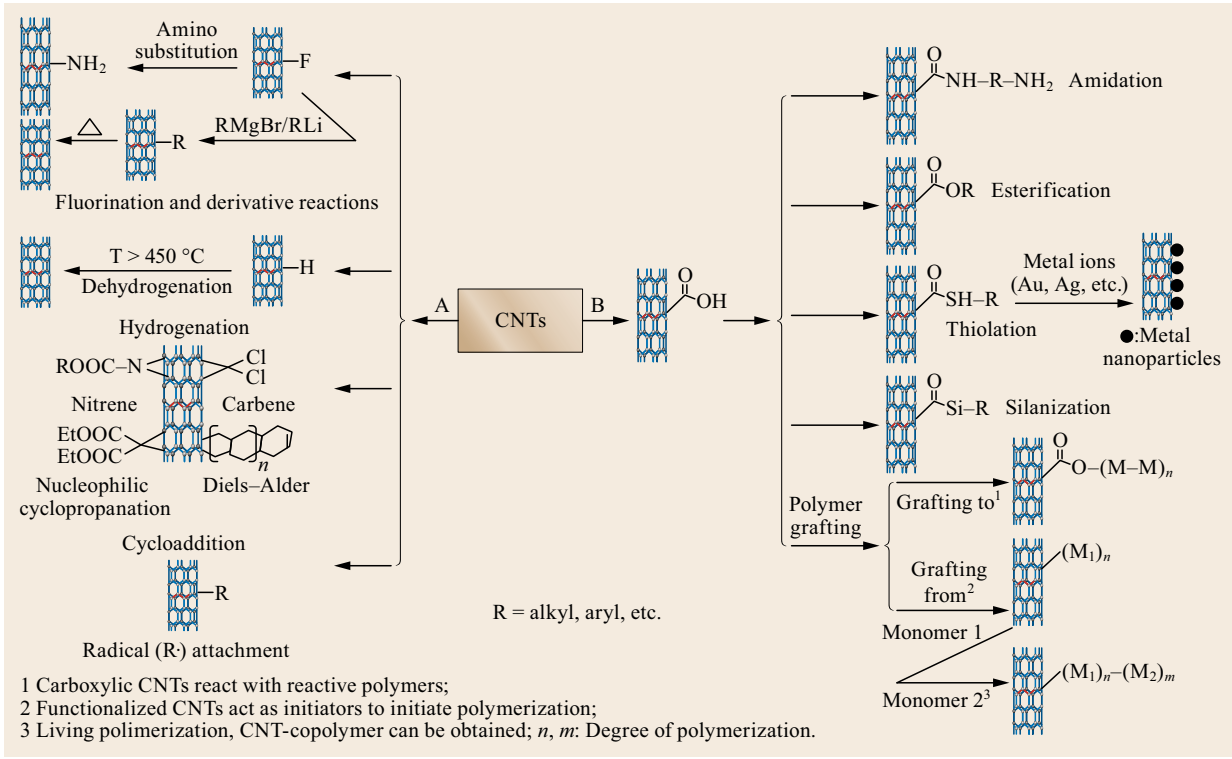


Fig. 8.22 Strategies for covalent functionalization of CNTs (*A*: direct sidewall functionalization; *B*: defect functionalization)

tion of Carbon Nanotubes) is the most popular route to activate the nanotube surface by creating surface carbonyl and carboxyl groups. There are many methods for the deposition of nanoparticles onto nanotubes, each offering different degrees of control of NP size and distribution along the tubes [8.275, 277]. Their choice often depends on the application field and on the scale of treatment. They are generally classified into two groups: the wet methods including the electrochemical, sol-gel and impregnation methods, and the gas processes such as physical vapor deposition (PVD), atomic layer deposition (ALD), and chemical vapor deposition (CVD) [8.275, 278–280].

The wet impregnation method is the most widely used technique even though it is multistep, because of its simplicity and cheapness [8.275]. It involves the impregnation of nanotubes by a solution of precursor salt or organometallic complex, followed by drying, calcination, and reduction. However, the size of the coated particles may be difficult to control precisely [8.278]. As an example, Pd nanoparticles were deposited on MWNTs by wet impregnation and by a functionalization using poly-alcohol based ligands, in order to increase the hydrogen absorption properties of nanotubes [8.280]. The polyol route gave much better dispersion of Pd on the MWNTs with smaller Pd nanoparticle size distribution.

The gas processes yield products often purer than the wet methods. They are single step and generally involve high coating rates [8.275]. CVD on MWNTs has often been used to deposit metallic (Pt, Pd, Ru, Rh) or semiconducting (InP, CdSe) nanoparticles on nanotubes [8.275]. The deposition of silicon layers on MWNT networks has been explored to prepare anodes of Li-ion batteries by combining the high specific energy capacity of silicon and the good mechanical strength and structural flexibility of nanotubes [8.281]. Composites made of SnO_2 NPs deposited on MWNTs have demonstrated excellent sensing responses for gas detection at room temperature [8.282]. As an alternative to oxidizing pretreatments, the predeposition of a titanium carbide (TiC) layer by CVD acting as a surface functionalization has successfully increased the surface coverage of MWNTs by Ni, Pd and Pt thin layers deposited by CVD [8.283]. Such depositions can be performed at large scale using the fluidized bed CVD process (Sect. 8.2.2, *Heterogeneous Processes*) [8.259, 284].

8.6 Carbon-Nanotube-Containing Materials (Composites)

Because of their unique and exceptional morphological characteristics and properties, CNTs make particularly promising component materials of composites with metals, ceramics, or polymer matrices. Key issues to address include the homogeneous dispersion of the CNTs (without damaging the CNTs), the control of the CNT/matrix bonding, the densification of bulk composites, and the possibility of aligning CNTs. In addition, the CNT type (SWNT, DWNT, MWNT, etc.) and production origin (arc, laser, CCVD, etc.) are also important variables that control structural perfection, surface reactivity, and aspect ratio of the reinforcement. The application of CNTs in this field is expected to lead to major advances in composites.

8.6.1 Metal Matrix Composites

CNT-metal matrix composites are less studied despite promising applications as structural materials in aeronautics, aerospace, and automotive fields. The main matrices are Al, Cu, Mg, Ti, and Ni. The materials are generally prepared by standard powder metallurgy techniques [8.285]. The spark plasma sintering (SPS) technique is sometimes used to densify the composites whilst avoiding matrix-grain growth [8.286]. Other techniques such as liquid metallurgy, thermal spray, and electrochemical deposition are also currently employed [8.287]. To achieve better dispersion of CNTs, new processing methods such as molecular level mixing and stir friction processing have been developed in recent years [8.287]. The room temperature electrical resistivity of hot-pressed MWNT-Al composites increases slightly upon increasing the MWNT volume fraction [8.288]. The tensile strengths and elongations of unpurified arc-discharge MWNT-Al composites are only slightly affected by annealing at 873 K in contrast to those of pure Al [8.289]. The addition of 1 vol.% MWNT to Al greatly enhances its tensile strength without sacrifices to its tensile elongation [8.290]. The elastic modulus, yield strength, and fracture toughness of 1.5, 3, and 4.5 vol.% MWNT-Al composites increase with carbon content [8.291]. The coefficient of thermal expansion (CTE) of 1 wt% MWNT-Al composite fabricated by cold isostatic pressing and the hot squeeze technique is 11% lower than that of pure Al or 2024 Al matrix, showing some promise as low-CTE materials. Associated with a high thermal conductivity, such materials would be interesting for applications such as packaging and space structures [8.292]. The stiffness, yield, and tensile strengths of spark-plasma-sintered MWNT-Cu composites increase with increasing carbon content up to 15 vol.% [8.293, 294]. 0.5 vol.% DWNT-Cu composite macroscopic wires prepared using a combination of SPS and room-temperature wire drawing present an increase of ultimate tensile strength by about 10% compared to the corresponding pure copper wires [8.295]. DWNT- and MWCNT-Cu composites also show a higher hardness and a lower friction coefficient and wear loss [8.296]. Hot-extruded CNT-Mg nanocomposites showed a simultaneous increase in yield strength, ultimate tensile strength and ductility, until a threshold of 1.3 wt% was reached [8.297]. Compared to pure Mg, a 45% increase of tensile strength was obtained for 2.4 wt% MWNT-Mg composites prepared by in situ synthesis of the CNTs within the Mg powder followed by hot extrusion [8.298]. The Young's modulus of nonpurified arc-discharge MWNT-Ti composite is about 1.7 times that of pure Ti [8.299]. The formation of TiC, probably resulting from a reaction between amorphous carbon and the matrix, was observed, but the MWNTs themselves were not damaged. An increase in the Vickers hardness by a factor of 5.5 over that of pure Ti was associated with the suppression of coarsening of the Ti grains, TiC formation, and the addition of MWNTs. The addition of small MWNT loadings of 0.18–0.35 wt% also improves the yield stress and tensile strength of Ti markedly [8.300]. Ni-plated MWNTs give better results than unplated MWNTs in strength tests. Indeed, CNT coating is a promising way to improve the strength of bonding with the matrix [8.301]. Composite films and coatings deposited by electroless or electrodeposition techniques on various substrates have also been studied. The addition of up to 15 vol.% purified SWNTs to nanocrystalline Al films reduces the coefficient of thermal expansion by as much as 65% and the resulting material could be a promising electronic packaging material [8.302]. CNT-Ni coatings deposited on a carbon steel plate by electroless deposition show significantly increased resistance to corrosion [8.303] and higher Vickers microhardness, higher wear resistance, and lower friction coefficient than SiC-reinforced composite deposits [8.304].

8.6.2 Ceramic Matrix Composites

Many different ceramic matrices have been studied over the years, although oxides (in particular Al_2O_3) are still the most studied [8.305–307]. There are three main methods for the preparation of CNT-ceramic nanocomposite powders. One is mechanical milling which usually involves long times that could damage the CNTs. Wet-milling is preferred but often requires the addition of organic additives to stabilize both the CNTs and the ceramic powder. This is also true for a second

method, i. e., the in situ synthesis of the matrix on preformed CNTs. It can lead to a good adhesion between the CNTs and the ceramic, but can be rather complex to implement. A third method is the in situ synthesis of the CNTs within the ceramic powder using procedures closely related to those described in Sect. 8.2.2, *Heterogeneous Processes*. The densification of the nanocomposite powders is made difficult by the detrimental influence of the CNTs. The most common methods are hot pressing (HP) and spark plasma sintering (SPS). Most of the works report that increasing the carbon content inhibits the densification of the composite, all the more when CNTs are homogeneously dispersed in the matrix. The SPS technique has been reported as an efficient method to achieve the total densification of CNT-oxide composites without damaging the CNTs [8.308, 309]. Full densification can be reached with SPS at comparatively lower temperatures with substantial shorter holding times. However, the successful densification by SPS at a lower temperature than for HP suggests that matrix grains are nonagglomerated and with sizes in the range of a few tens of nanometers. The influence of CNT dispersion on mechanical properties, in particular on toughness, has been controversial. Indeed, strong increases in toughness derived from the measurement of Vickers indentation cracks have been reported [8.307], but they were shown to be probably widely overestimated because such materials are very resistant to contact damage [8.310, 311]. *Xia* et al. [8.312] reported microstructural investigations on MWNTs well aligned in the pores of an Al_2O_3 membrane. Different possible reinforcement mechanisms induced by the MWNTs have been evidenced, such as crack deflection, crack bridging, MWNT pulling-out, and MWNT collapsing in shear bands. Indeed, although so far neither single-edge notched beam nor single-edge V-notched beam results have evidenced that CNTs can significantly reinforce Al_2O_3 ceramics [8.313], a strong and unambiguous increase in toughness has been obtained for DWNT-MgO composites in which the CNTs are very homogeneously dispersed [8.314]. Enhanced tribological properties (low friction and/or wear) of composites have been reported [8.315–318]. The microhardness is found to either increase or decrease, and this depends greatly on the powder preparation route. As noted in [8.311], processing-induced changes in the matrix may have greater effects on the mechanical properties than the actual presence of CNTs. Regarding the thermal properties, CNT-ceramic composites often show a lower thermal conductivity than the corresponding ceramics, probably caused by too high thermal contact resistances at CNT-CNT and CNT-ceramic grain junctions [8.319, 320]. By contrast, CNTs increase the electrical conductivity of insulating ceramic nanocomposites [8.321–324], with a low percolation threshold (less than 1 vol.%) due to their very high aspect ratio [8.322]. The electrical conductivity can be tailored within several orders of magnitude directly by the CNT quantity and is well fitted by the scaling law of the percolation theory with the exponent close to the theoretical value characteristic of a three-dimensional network [8.322]. An anisotropic conductivity is obtained when the CNTs are aligned within the composite by high-temperature extrusion [8.325].

8.6.3 Polymer-Matrix Composites

CNT-polymer composites, first reported by *Ajayan* et al. [8.326], are still being intensively studied, especially epoxy- and polymethylmethacrylate (PMMA)-matrix composites. A review on the chemistry, processing, mechanical, and electrical properties can be found in [8.327]. In terms of mechanical characteristics, the three key issues that affect the performance of a fiber-polymer composite are the strength and toughness of the fibrous reinforcement, its orientation, and good interfacial bonding, which is crucial to load transfer [8.328]. The ability of the polymer to form large-diameter helices around individual CNTs favors the formation of a strong bond with the matrix [8.328]. Isolated SWNTs may be more desirable than MWNTs or bundles for dispersion in a matrix because of the weak frictional interactions between layers of MWNTs and between SWNTs in bundles [8.328]. The main mechanisms of load transfer are micromechanical interlocking, chemical bonding, and van der Waals bonding between the CNTs and the matrix. A high interfacial shear stress between the fiber and the matrix will transfer the applied load to the fiber over a short distance [8.329]. SWNTs longer than 10−100 μm would be needed for significant load-bearing ability in the case of nonbonded SWNT-matrix interactions, whereas the critical length for SWNTs cross-linked to the matrix is only 1 μm [8.330]. Defects are likely to limit the working length of SWNTs, however [8.331]. The load transfer to MWNTs dispersed in an epoxy resin was much higher in compression than in tension [8.329]. It was proposed that all of the walls of the MWNTs are stressed in compression, whereas only the outer walls are stressed in tension because all of the inner tubes are sliding within the outer tube. Mechanical tests performed on 5 wt% SWNT-epoxy composites [8.332] showed that SWNT bundles were pulled out of the matrix during the deformation of the material. The influence of the interfacial CNT/matrix interaction was demonstrated by *Gong* et al. [8.333]. It was also reported that coating regular carbon fiber with MWNTs prior to their dispersion into an epoxy matrix

improves the interfacial load transfer, possibly via local stiffening of the matrix near the interface [8.334]. DWNT-epoxy composites prepared by a standard calendaring technique were shown to possess higher strength, Young's modulus, and strain to failure at a carbon content of only 0.1 wt% [8.335]. Significantly improved fracture toughness was also observed. The influence of the different types of CNTs (SWNTs, DWNTs, and MWNTs) on the mechanical properties of epoxy-matrix composites is discussed in [8.336]. The stiffness and damping properties of SWNT- and MWNT-epoxy composites were investigated for use in structural vibration applications [8.337]. It was shown that enhancement in damping ratio is more dominant than enhancement in stiffness, MWNTs making a better reinforcement than SWNTs. Indeed, up to 700% increase in damping ratio is observed for MWNT-epoxy beam as compared to the plain epoxy beam. The outstanding potential of CNTs as reinforcements in polymer composites is evident from the supertough composite fibers fabricated by the groups of *Baughman* et al. [8.338] and *Poulin* et al. [8.339]. By optimizing the *coagulation spinning* method [8.340, 341], they spun several hundred meters of CNT-polyvinyl alcohol composite fibers, which have tensile properties comparable to that of a spider silk. As for ceramic matrix composites, the electrical characteristics of CNT-polymer composites are well described by the percolation theory [8.327, 342]. Polymer hosting CNTs such as PMMA, epoxy, polyimide, and polystyrene exhibit extremely low percolation threshold values (much below 1 wt%) [8.327]. An ultralow electrical percolation threshold of 0.0025 wt% was reported for aligned MWNT-epoxy composites [8.343]. The thermal conductivity of composites can also be greatly enhanced by the alignment of CNTs [8.344]. Polymer composites with other matrices include CCVD-prepared MWNT-polyvinyl alcohol [8.345], arc-prepared MWNT-polyhydroxyaminoether [8.346], arc-prepared MWNT-polyurethane acrylate [8.347, 348], SWNT-polyurethane acrylate [8.349], SWNT polycarbonate [8.350], MWNT-polyaniline [8.351], MWNT-polystyrene [8.352], CCVD double-walled nanotubes-polystyrene-polymethylacrylate [8.353], MWNT-polypropylene [8.354, 355], SWNT-polyethylene [8.356–358], SWNT-poly(vinyl acetate) [8.357, 358], CCVD-prepared MWNT-polyacrylonitrile [8.359], SWNT-polyacrylonitrile [8.360], MWNT-oxotitanium phthalocyanine [8.361], arc-prepared MWNTpoly(3-octylthiophene) [8.362], SWNT-poly-(3-octylthiophene) [8.363], and CCVD MWNT-poly(3-hexylthiophene) [8.364]. These works deal mainly with films 100–200 μm thick, and aim to study the glass transition of the polymer, its mechanical and electrical characteristics, as well as the photoconductivity. A great deal of work has also been devoted to the applications of CNT-polymer composites as materials for molecular optoelectronics, using primarily poly[(*m*-phenylenevinylene)-*co*-(2,5-dioctoxy-*p*-phenylenevinylene)] (PmPV) as the matrix. This conjugated polymer tends to coil, forming a helical structure. The electrical conductivity of the composite films (4–36 wt% MWNTs) is increased by eight orders of magnitude compared to that of PmPV [8.365]. Using MWNT-PmPV composites as electron transport layers in light-emitting diodes results in a significant increase in brightness [8.366]. The SWNTs act as a hole-trapping material that blocks the holes in the composites; this is probably induced through long-range interactions within the matrix [8.367]. Similar investigations were carried out on arc-discharge SWNT-polyethylene dioxythiophene (PEDOT) composite layers [8.368] and MWNT-polyphenylenevinylene composites [8.369].

To conclude, two critical issues must be considered when using nanotubes as components for advanced composites. One is to choose between SWNTs, DWNTs, and MWNTs. The former seem more beneficial to mechanical strengthening, provided that they are isolated or arranged into cohesive yarns so that the load can be conveniently transferred from one SWNT to another. Unfortunately, despite many advances [8.340, 341, 370–373], this remains a technical challenge. The other issue is to tailor the CNT/matrix interface with respect to the matrix. In this case, DWNTs and MWNTs may be more efficient than SWNTs.

8.6.4 Composites as Multifunctional Materials

One of the major benefits expected from incorporating CNTs into other solid or liquid materials is that they endow the material with some electrical conductivity while leaving other properties or behaviors unaffected. As already mentioned in the previous section, the percolation threshold is reached at very low CNT loadings. Tailoring the electrical conductivity of a bulk material is then achieved by adjusting the CNT volume fraction in the formerly insulating material while making sure that this fraction is not too large. According to *Maruyama* [8.374], there are three areas of interest regarding the electrical conductivity:

1. Electrostatic discharge (for example, preventing fire or explosion hazards in combustible environments or perturbations in electronics, which requires an electrical resistivity of less than $10^{12}\ \Omega$cm)
2. Electrostatic painting (which requires the material to be painted to have enough electrical conductiv-

Table 8.3 Applications of nanotube-based multifunctional materials (from [8.374]), copyright B. Maruyama (WPAFB, Dayton, Ohio)

Fiber fraction	Applications	Mechanical			Electrical			Thermal		Thermo-mechanical	
		Strength/ stiffness	Specific strength	Through-thickness strength	Static dissipation	Surface conduction[a]	EMI shielding	Service[b] temperature	Conduction/ dissipation[c]	Dimensional stability[d]	CTE reduction[e]
Low volume fraction (fillers)											
Elastomers	Tires	✓			✓				✓		
Thermoplastics	Chip package				✓				✓		
	Electronics/housing	✓					✓	✓	✓		
Thermosets	Epoxy products	✓	✓	✓		✓				✓	✓
	Composites			✓						✓	
High volume fraction											
Structural composites	Space/aircraft components		✓	✓							
High conduction composites	Radiations	✓							✓	✓	
	Heat exchangers	✓						✓	✓		✓
	EMI shield	✓					✓				

[a] For electrostatic painting, to mitigate lightning strikes on aircraft etc.
[b] To increase service temperature rating of product
[c] To reduce operating temperatures of electronic packages
[d] Reduces warping
[e] Reduces microcracking damage in composites

ity – an electrical resistivity below $10^6\,\Omega$cm – to prevent the charged paint droplets from being repelled)
3. Electromagnetic interference shielding (which is achieved for an electrical resistivity of less than $10\,\Omega$cm).

Materials are often required to be multifunctional; for example, to have both high electrical conductivity and high toughness, or high thermal conductivity and high thermal stability. An association of several materials, each of them bringing one of the desired features, generally meets this need. The exceptional features and properties of CNTs make them likely to be a perfect multifunctional material in many cases. For instance, materials used in satellites are often required to be electrically conductive, mechanically self-supporting, able to transport away excess heat, and often to be robust against electromagnetic interference, while being of minimal weight and volume. All of these properties should be possible with a single nanotube-containing composite material instead of complex multimaterials combining layers of polymers, aluminum, copper, and so on. Table 8.3 provides an overview of various fields in which nanotube-based multifunctional materials should find application.

8.7 Current Applications of Carbon Nanotubes (on the Market)

Perfect CNTs exhibit high aspect ratio, high tensile strength, low mass density, high heat conductivity, large surface area, poor chemical reactivity, and versatile electronic behavior, including high electron conductivity. However, while these are the main characteristics of individual nanotubes, many of them can form secondary structures such as ropes, fibers, papers, and thin films with aligned tubes, all with their own specific properties. These properties make them potential candidates for a nearly infinite variety of applications, regardless of their cost [8.375]. Cost is an issue, though, which strongly depends in first place on both the CNT type and quality and on the production process type (e.g., all semiconducting, high-quality SWNTs can cost ≈ 50 times more than gold to date (2016)!). Cost is also strongly affected by the modifications to be brought to production lines for being nanocompatible, as well as that of the specific studies for assessing potential public health issues all along the material/device life (i. e., from production to waste and recycling). All this can prevent a CNT-based material or device, although capable of superior performance, from reaching the market. Because of this, we only list below CNT-based applications which actually are on the market, and readers who are interested in the vast potentiality of CNTs may refer to other literature sources [8.375, 376].

8.7.1 Carbon Nanotubes and Master Batches

Carbon nanotubes have become commercial products by themselves, for supplying research laboratories and start-ups. Table 8.4 provides a list of CNT-selling companies around the world. This list was probably nonexhaustive, and is certainly already partly obsolete, and partly incomplete. Examples of current price ranges are provided in the various sub-sections of Sect. 8.2.

Companies also (or instead) sell *CNT master batches*, i. e., premixed CNT-polymers for instance, or transformed products containing nanotubes (e.g., sheets, tapes, yarns, etc.) to be subsequently used by the buyer as starting materials for fabricating parts and devices for their own uses or for final products to sell.

8.7.2 Near-Field Microscopy Probes

The high mechanical strength and aspect ratio of carbon nanotubes makes them almost ideal candidates for use as force sensors in scanning probe microscopy (SPM). They provide higher durability, low sensitivity to electrostatic charges, and the ability to image surfaces with a high lateral resolution, with respect to conventional (ceramic-based) force sensors. The idea was first proposed and tested by *Dai* et al. [8.377] using c-MWNTs. Although it was extended to SWNTs by *Hafner* et al. [8.378], commercial nanotube-based probes use MWNTs for processing convenience. CNT-based SPM tips also offer the potential to be functionalized, leading to the prospect of selective imaging based on chemical discrimination in *chemical force microscopy* (CFM), allowing the chemical mapping of molecules [8.379, 380]. CNT-based SPM tips are not drawback-free, however. They may generate specific image artifacts (e.g., *ringing* effect), and exhibit highly variable geometry and alignment from a tip to another even within the same probe batch and brand. Current nanotube-based SPM tips or high-resolution AFM tips can be quite expensive, in the range of ≈ 450 \$/tip on the average, depending on the geometrical characteristics and the supplier (CDI, NanoSensors). The market for nanotube AFM tips is growing, and was estimated to reach $\approx 150\,$M€/yr in 2016.

Table 8.4 Listing of companies selling nanotubes (all types included) around the world (2015 data)

Company Name	Website	Company Name	Website	Company Name	Website
American Dye Source Inc. (USA)	http://adsdyes.com	Honjo Chemical Corporation (Japan)	http://honjo-chem.co.jp	NanoNB Corp (Canada)	http://nanonb.com
Applied Science Inc. (USA)	http://apsci.com	Hyperion Catalysis International, Inc (USA)	http://fibrils.com	Nanoselect Inc (USA)	http://nanoselect-sensors.com
Arkema (France)	http://arkema.com	Innovations Unified Technologies (India)	http://iutechnologies.com	Nanoshel LLC (USA)	http://nanoshel.com
Arknano Technology Development Co. (China)	http://arknano.com	Luna Nanoworks (USA)	http://lunananoworks.com	Nanostructured & Amorphous Materials, Inc (USA)	http://nanoamor.com
Arry International Group (Hong Kong)	http://arry-nano.com	MER Corporation (USA)	http://mercorp.com	NanoTechLabs, Inc (USA)	http://nanotechlabs.com
Bayer Material Science (Germany)	http://baytubes.com	Mitsui Chemicals, Inc (Japan)	http://jp.mitsuichem.com	Nanothink SA (Greece)	http://nanothinx.com
BioNano Internat. Singapore Pte Ltd (Singapore)	http://nanotech-now.com	Mknano (Canada)	http://mknano.com	Natural Nano, Inc (USA)	http://naturalnano.com
Bucky USA (USA)	http://buckyusa.com	Molecular Nanosystems Inc (USA)	http://monano.com	N-Tec (Norway)	http://n-tec.no
C-Polymers GmbH (Austria)	http://c-polymers.com	MTR Ltd (USA)	http://mtr-ltd.com	Raymor Industries, Inc (Canada)	http://raymor.com
Carbon NT and F (Austria)	http://carbon-nanofiber.com	Nano-C, Inc (USA)	http://nano-c.com	Rossetter Holdings Ltd (Cyprus)	http://e-nanoscience.com
Carbon Solutions Inc (USA)	http://carbonsolution.com	NanoCarbLab (Russia)	http://nanocarblab.com	SES Research (USA)	http://sesres.com
Catalytic Materials LLC (USA)	http://catalyticmaterials.com	Nanocs (USA)	http://nanocs.com	Shenzhen Nanotech Port Co, Ltd (NTP) (China)	http://nanotubes.com.cn
Cheap Tubes (USA)	http://Cheaptubesinc.com	Nanocomp Technologies, Inc (USA)	http://nanocomptech.com	SouthWest NanoTechnol., Inc (USA)	http://swentnano.com
Chengdu Organic Chemicals Co (China)	http://timesnano.com	Nanocomposites Inc (USA)	http://nanocompositesinc.com	Sun Nanotech Co (China)	http://sunnano.com
Continental Carbon Company (USA)	http://continentalcarbon.com	NanoCompound GmbH (Germany)	http://nanocompound.de	Surrey NanoSystems (UK)	http://surreynanosystems.com
DuPont (USA)	http://dupont.com	Nanocyl SA (Belgium)	http://nanocyl.com	Thomas Swan and Co (UK)	http://thomas-swan.co.uk
Frontier Carbon Corp (Japan)	http://f-carbon.com	NanoIntegris (USA)	http://nanointegris.com	Unidym Inc (USA)	http://unidym.com
Fujitsu Laboratories (Japan)	http://jp.fujitsu.com/group/labs/en	NanoLab, Inc (USA)	http://nano-lab.com	Xintek, Inc (USA)	http://xintek.com
Future Carbon GmbH (Germany)	http://future-carbon.de	Nanoledge Inc (Canada)	http://nanoledge.com	YTC America, Inc (USA)	http://ytca.com/carbon_nanotubes
Helix Material Solutions (USA)	http://Helixmaterial.com	NanoMas Technologies (USA)	http://nanomastech.com		

8.7.3 Electron Emitter

In a pioneering work by *de Heer* et al. [8.381], CNTs were shown to be efficient field emitters. Briefly, a potential difference is set up between the emitting tip(s) and an extraction grid (or an anode) so that electrons are pulled from the tip(s) onto an electron-sensitive screen layer. With respect to regular (Mo, doped-Si, W) electron-emitting tips, the structural perfection of CNTs is supposed to provide higher electron emission stability, higher mechanical resistance, longer lifetimes, and $1/3-1/10$ lower energy consumption (for example, nanotubes are able to produce a current density of $1\,\mathrm{mA/cm^2}$ for a threshold voltage of $3\,\mathrm{V/\mu m}$, while regular Mo or Si tips require $100\,\mathrm{V/\mu m}$). Generally speaking, the maximum current density that can be obtained with CNTs ranges from 10^6 to $10^8\,\mathrm{A/cm^2}$ depending on the nanotubes involved (SWNT or MWNT, opened or capped, aligned or not, and so on) [8.382–384]. Those properties have made CNTs potential candidates for a variety of applications, yet comprising two categories: (i) those in which electron emitters are displayed by thousands according to a periodic pattern, as for flat panel displays for television sets and computers, and (ii) those merely requiring a single electron-producing cathode, as for x-ray sources, and electron sources for electron microscopes. In spite of the potentiality of CNTs in this field and the convincing existence of demonstrators and prototypes (by Samsung, for instance), no applications within the former categories have reached the market so far, indicating that some cost and/or technological issues are yet to be fixed. Conversely, regarding the latter category, companies such as Oxford Instruments and Medirad have commercialized miniature x-ray generators for medical applications which use nanotube-based cold cathodes developed by Applied Nanotech Inc. Also, a new cold-field emission gun for electron microscopes based on a single, small-diameter MWNT (yet thickened into a cone in order to prevent the CNT from vibrating while emitting, which is a fatal drawback for the quality of the emitted beam) is currently under development at Hitachi High Technologies (Japan) based on work carried out at the University of Toulouse (France) [8.385].

8.7.4 Flexible and Touch-Screen Displays

Nowadays, the majority of touch-screen displays in smartphones and tablets are made of Indium Tin Oxide (ITO) transparent conductive electrodes which are brittle and thus unusable for flexible displays. Chemically doped polymers exhibit both significant electrical conductivities and flexibility but demonstrate undesirable optical transmittance over the visible region [8.386]. Owing to their very high aspect ratio, CNTs increase the electrical conductivity of polymer composites with a very low percolation threshold and allow the fabrication of transparent conductive CNT-polymer films. Commercial developments by Eikos resulted in SWNT-polymer films with optoelectronic performances (90–97% visible-light transmittance and $200-500\,\Omega/\mathrm{square}$ sheet resistance) very close to ITO [8.387]. Moreover, CNTs can be deposited by cost-effective solution-based processes, a supplementary advantage over ITO coatings produced by sputtering at low pressure [8.388].

8.7.5 Nonvolatile Random Access Memory

Nonvolatile random access memories (NRAM) have recently been developed by Nantero, based on a concept by *Rueckes* et al. [8.389] which exploits a suspended SWNT crossbar array for both I/O and switchable, bistable device elements with well-defined OFF and ON states. This crossbar consists of a set of parallel SWNTs on a substrate and a set of perpendicular SWNTs that are suspended on a periodic array of supports. Each cross point in this structure corresponds to a device element with an SWNT suspended above a perpendicular nanowire. For practical manufacturing purposes, the actual switches do not rely on single SWNTs but rather on a random mesh of hundreds of SWNTs with many different intersection points (thin film of tangled CNTs patterned on silicon substrate). The devices reach switching speeds of 100 to 200 GHz, operate at orders of magnitude more cycles than flash memories, and present lower power consumption compared to dynamic random access memories (RAMs).

8.7.6 Light Absorbants

Optical properties of carpet-like deposits of CNTs (Fig. 8.15c) are being used by Surrey NanoSystems to create and sell a coating called Vantablack absorbing 99.6% of light, which is useful in optical devices for instance. Hence it is not a painting, but it is claimed to be the darkest black *color* available. This has attracted the interest of various artists to such an extent that a British (and rich) one, Anish Kapoor, bought in 2016 the exclusive rights to use Vantablack for artistic purposes. However, Frederic De Wilde, a Belgian artist, claims anteriority with a coating named NANOblck-Sqr#1, also based on CNTs, jointly developed by Rice University and NASA for the fabrication of materials with light-adsorbing surfaces.

8.7.7 Automotive and Aeronautic Industry

As reported in Sect. 8.6.3, CNTs increase the electrical conductivity of polymer composites at very low percolation threshold (below 1 wt%) due to their high aspect ratio. In the automotive industry, this property is used to enable electrostatic-assisted painting of polymer parts such as mirror housings and bumpers [8.376]. Since 2009, BASF has produced a CNT-polyoxymethylene composite fuel filter housing for the Audi A4 and A5. The high conductivity of the composite allows suppression of electrostatic charging and sparking as fuel flows through the filter. Moreover, this material is more creep-resistant than conventional polyoxymethylene and less costly than other conductive polymer composites. CNT-polymer composites are also widely used as anticorrosion coatings for metals. CNTs enhance mechanical and tribological properties of coatings while providing an electric pathway for cathodic protection [8.390]. Solvent-based zinc, aluminium and magnesium CNT-epoxy/polyamide paints for primer anticorrosion coatings were designed by Tesla Nanocoatings. CNT sheets are used in the aerospace industry to protect polymer composites against electrostatic discharge (ESD), to protect electronic components against electromagnetic interference (EMI), and to create honeycomb structures with higher stiffness than aluminum with virtually no thermal expansion, reduced weight, and improved vibrational damping properties.

8.7.8 High-Tech Goods and Clothes

Enhancement in damping ratios of CNT-polymer composites (up to 700% for MWNT-epoxy composites [8.337]) is used to improve sporting goods like baseball bats, tennis and badminton racquets, hockey sticks, racing bicycles, golf balls/clubs, skis, fly-fishing rods, and archery arrows. Easton Sports, manufacturer of sporting goods, moved new markets from wood and steel sports equipment to lightweight, high-strength CNT-polymer products. Considering the still-current difficulty of dispersing CNTs in polymers well, which is likely to generate flaws in the material from a mechanical point of view, one might wonder whether the improvement is real, or whether a minimized amount of CNTs is merely added to take benefit of the mediatic impact of using new and high-tech materials.

Far from sport issues yet a serious matter nowadays, CNT sheets are applied in light armors and projectile-protection wearables reducing the weight of the finished products. For instance, bullet-proof vests as manufactured by China Xingxing Guangzhou Imp. & Exp. Corp. (China) or AR500 Armor (USA) can weigh 25% less than their Kevlar-based equivalents, at lower cost.

8.7.9 Anodes for Li-Ion Batteries

A Li-ion battery is composed of three main components, the anode intercalated with Li^+ ions, the cathode, and the electrolyte. During charging and discharging cycles Li^+ ions diffuse between the anode and cathode through the electrolyte which is typically a lithium salt dissolved in an organic solvent. Graphite has long been the preferred anode material for commercial batteries due to its relatively high capacity, energy storage density, good cyclability, and a low redox potential for lithium. The capacity of graphite in Li-ion batteries is limited by the number of Li ions that can intercalate between individual graphene layers. Assuming a maximum of 1 Li atom for every 6 carbon atoms (LiC_6) one obtains a theoretical maximal capacity of $372\,mA\,hg^{-1}$ [8.391]. Owing to their high aspect ratio, high specific surface and high conductivity, CNTs have been shown to be excellent conductive additives for Li-ion batteries [8.392]. Li-ion storage capacity in CNTs results from the availability of the central core of the tube, the interlayer space (for MWNTs), and the interstitial space between CNTs when they are assembled in bundles [8.393]. The measured reversible Li-ion capacities for CNT-based anodes can exceed $1000\,mA\,hg^{-1}$, which is a 3-fold improvement over conventional graphite anodes [8.394]. The addition of CNTs on the graphite anode can also efficiently improve rate capability and cyclability of the resulting CNT-graphite composite electrodes [8.395]. For example, the cyclic efficiency of a graphite anode was maintained at almost 100% for up to 50 cycles by adding 10 wt% MWNTs [8.396]. CNTs are thus widely used as anode additives in commercial Li-ion batteries and more than 50% of mobile phones and notebook batteries contain CNTs [8.395]. Several CNT-producing companies such as ANS, BlueNano, Showa Denko K.K and Ube Industries provide CNT-based powders for battery electrodes.

8.7.10 Chemical Sensors

The development of CNT-based sensors has attracted intensive interest in the last several years because of their excellent sensing properties such as prompt response and high selectivity for a wide variety of analytes with detection limits reaching ppb levels [8.397, 398]. Different kinds of gas/chemical sensors, based on different working principles, have been fabricated thanks to CNTs: miniaturized ionizing gas sensors (also called *microgun* sensors), CNT thin films with variable resistance as a function of the adsorbed gas properties (chemiresistors), and CNT field-effect transistor (FET)-based sensors [8.399]. To overcome some of

the limitations posed by pristine CNT sensors (lack of specificity to different gaseous analytes, and low sensitivity towards analytes that have no affinity to CNTs), functionalization of CNTs with various methods was considered, and with a wide spectrum of materials such as conducting polymers and metal nanoparticles [8.397].

The commercialization of nanoenabled sensors is in the early stages. Research efforts are still needed in particular to control the reproducibility and specificity (chirality, diameter, and length) of commercial CNT batch characteristics [8.400, 401]. CNT-based sensors are currently being developed by large and small companies, mainly in the USA. Nanomix Incorporated (California, USA) has commercialized a CNT-based nanosensor for hydrogen detection (operating range of 2000–20 000 ppm, which is 5–50% of the lower explosive limit for hydrogen). NASA Ames Research Center is developing a low-power-consumption gas sensor based on coated (e.g., with chlorosulfonated polyethylene) or decorated (e.g., with Pd, Pt, Au, Cu, and Rh nanoparticles) SWNTs, so far tested for NO_2, NH_3, CH_4, Cl_2, HCl, toluene, benzene, acetone, formaldehyde, and nitrotoluene. Northeastern's Center for High-rate Nanomanufacturing has developed a simple and highly sensitive multibiosensor containing semiconductor SWNTs that are enzyme-immobilized for detecting D-glucose, L-lactate, and urea in sweat. The core technology platform developed by Nano Engineered Applications, Inc. (Riverside, California) is a gas sensor using SWNTs and solid-state electrical sensing, capable of real-time detection of airborne gases at the ppb level. Sensigent (Baldwin Park, California) chemical sensing technology includes CNT-based nanosensors with high sensitivity and selectivity for small molecule gases. Design West Technologies, Inc., (Southern California) proprietary CNT technology can be incorporated into existing chemical detector products as standalone chemical sensors or as a supporting orthogonal technology to detect various chemical vapors. C2Sense (Cambridge, MA) aims to provide sensors for meat spoilage detection in smart packaging. The sensing material is based on CNTs, which are functionalized with proprietary selector materials. Brewer Science (Rolla, Missouri) developed a fully integrated temperature and humidity sensor based on highly purified aqueous-based SWNT ink. It is capable of providing real-time simultaneous measurement of the temperature and rapid fluctuations of humidity in a surrounding environment.

8.7.11 Catalyst Support

Carbon-based materials make good supports in heterogeneous catalytic processes due to their ability to be tailored to the specific needs, such as high surface area, high amount of mesopores, good stability at high temperatures (under nonoxidizing atmospheres), and the possibility of controlling the chemical nature of their surfaces [8.402]. Nanosized carbon fibrous morphologies have appeared over the last decade, which show great potential for use as supports [8.403] superior to regular ones such as activated carbon, carbon blacks, or graphite [8.404]. The field of catalysis to which the specific features of CNTs could bring the most significant advancements is perhaps the oxygen reduction reaction in fuel cell electrocatalysis. Carbon materials with both high surface area and good crystallinity such as CNTs not only provide a high dispersion of the catalyst (Pt) nanoparticles, but also limit carbon oxidation at high potentials, and facilitate electron transfer, resulting in better device performance [8.405]. Moreover, CNTs have a positive effect on Pt structure, resulting in a higher catalytic activity and a higher stability than with carbon black supports [8.406]. In that sense, CNTs offer great promise for overcoming the problems of existing fuel cells, such as Pt loading. However, a problem for the commercialization of CNT-based electrodes is their higher cost compared to that of carbon blacks.

Currently, Bing Energy Inc. (Tallahassee, Florida) has commercialized a proton-exchange membrane fuel cell, comprised of highly conductive CNT buckypaper coated with Pt nanoparticles, which was developed at Florida State University to improve membrane manufacturing processes, fuel cell efficiency, and lifetime.

8.8 Toxicity and Environmental Impact of Carbon Nanotubes

It is first very important to consider that the large variety of CNTs (SWNT, DWNT, MWNT, hetero-CNTs, hybrid CNTs, etc.) and of synthesis routes (arc discharge, laser ablation, CCVD, ...) as well as the lack of standardized testing procedures make the investigation of the toxicity of CNTs very difficult, and the comparison of the already published results almost impossible [8.407]. CNTs are mostly found as bundles rather than as individual objects, or more likely as large micrometric agglomerates. All CNT materials contain various levels of residual catalyst(s), depending on the synthesis route and purification steps they have been subjected to. Usual purification treatments involve the combination of acids and oxidizing agents, which

leads to partial functionalization of the outer wall, making the treated samples more hydrophilic. SWNTs and DWNTs usually form long and flexible bundles (typically hundreds of micrometers long) whereas MWNTs are generally shorter (tens of micrometers) and more rigid. MWNTs also have generally more surface defects, which enhances their chemical reactivity. The specific surface area can range from a few tens of m^2/g in the case of densely packed MWNTs to nearly $1000\,m^2/g$ in the case of SWNTs and DWNTs (the theoretical limit being $\approx 1300\,m^2/g$ in the case of individual closed SWNTs). This is a very important parameter as the surface area has recently been shown to be the most relevant metric for the comparison of different carbon nanoparticles, even with very different morphologies [8.408].

The main exposure routes for dry CNTs are inhalation and dermal contact (also possible in the case of suspensions). Ingestion is generally not considered (would be accidental), although it is in fact more or less related to inhalation. In the case of suspensions, the main issue concerns their stability. This question has been widely studied worldwide and the general approach is the addition of a surfactant in order to stabilize the CNTs in the liquid. The main problem is that all commonly used surfactants are toxic to a certain extent and thus cannot be used in the presence of living cells or animals for in vivo or in vitro investigations, or at such low concentrations that really they no longer play the role they are supposed to play. Although a few natural surfactants have been investigated, the stability of the suspensions in the presence of living organisms is often very different (fast destabilization leading to flocculation). Injection in the bloodstream is envisaged, but would not be accidental (biological applications such as imaging, targeted cell delivery, hyperthermia, etc.). After the CNTs have entered the body, they could travel following different routes depending on the entry point (movements from one organ to another are called translocation) but also mainly on their physicochemical characteristics. Objects recognized as nonself by the immune system usually end up in the liver or the kidneys if they can be transported there, and could possibly be excreted (eliminated) from the body. In the general case, CNTs will just accumulate (biopersistance). They are usually intercepted by macrophages (cells present in all tissues and whose role is to phagocyte (engulf and then digest – when this is possible) cellular debris and pathogens as well as to stimulate lymphocytes and other immune cells to respond to the pathogens). Taking into account the small size of macrophages as compared to that of agglomerates, bundles or even individual CNTs, macrophages usually do not manage to get rid of the CNTs by phagocytose. However, they try to do so and thus release reactive oxygen species, enzymes, cytokines (interferons), etc. and agglomerate around them to isolate them from the body. Proteins present in blood and most of the biological fluids (complement system – innate immunity) will play a similar role by *labeling* the CNTs (opsonization) and possibly generating some inflammatory reactions. The complement system strongly interacts with the lymphocytes. These natural phenomena have deleterious consequences on the surrounding tissues: inflammation in a first instance, formation of granuloma (commonly observed in the lungs after exposure to CNTs). Each target organ has its own phagocyte cells (Kupffer cells in the liver, Langerhans cells in the skin, etc.).

Toxicity can be assessed both by in vitro and in vivo experiments. In the case of in vitro assays, cell cultures (usually immortalized cancer cells, but also primary cultures or even stem cells) are exposed to suspensions of CNTs. In the case of in vivo assays, the animals (mice, rats, worms, amphibians, fishes, etc.) are exposed either to aerosols (inhalation) or mainly again to suspensions of CNTs which will be administrated according to different protocols depending on the study (intratracheal instillation, injection, contact with the skin, etc.). Extrapolating the toxicity results from animals (or even worse, from cells) to humans is very delicate but the data are however very useful for the sake of comparison in a given system and with given experimental conditions. As soon as CNTs are in contact with a biological fluid, their surface chemistry is likely to be modified very quickly by adsorption of proteins (complement system, surfactants [8.409], etc.); this adsorption can be very specific [8.408, 409], and is likely to be dynamic and controlled by the affinity of the molecules for the surface of the CNTs (pristine or functionalized). It is thus obvious that the surface chemistry of the CNTs will play a very important role.

The potential use of CNTs in commercial products (Sect. 8.7) begs the question of their fate at the end of their lifecycle. If the impact of CNTs on human health has been under investigation for a few years already, it is noteworthy that environmental impact has only scarcely been taken into account. Only a few publications (less than 80) are available to date and the concentration at which ecotoxic effects are evidenced is in most cases much higher than what could be reasonably found in the environment (unless very local and specific conditions apply). Because of the potentially very high specific surface area of CNTs, they could act as vectors for pollutants adsorbed on their surface (polycyclic aromatic hydrocarbons or heavy metal ions, for example), even if themselves do not show any obvious sign of toxicity.

There is currently no consensus about the toxicity of CNTs [8.407], although about 600 papers have al-

ready been published on this topic within the last 5 years (2016 data). Despite the worldwide efforts devoted to this field of research, the huge variety of CNT types, shapes, composition, etc. still makes it very difficult to answer this simple question: *are CNTs toxic*? The *principle of precaution* should not stop all research in this area but only draw attention to a more responsible attitude for people working on their synthesis or manipulating them, and for industries willing to include them in consumer products. Gloves should be worn at all times as well as an adapted (FFP3 (filtering facepiece protection class 3) type) disposable dust mask. Wearing a lab coat is recommended to limit contamination of clothes, and CNT wastes should be burnt.

8.9 Concluding Remarks

Carbon nanotubes have been the focus of a lot of research work (and therefore a lot of funding) for more than two decades now. Although some experts initially predicted the same disappointment that followed the *fullerene era* (originally believed to be so promising, but which has resulted in no significant applications to date, in spite of a considerable investment of time and money), it is clear that applications including carbon nanotubes are now coming onto the market. It took time indeed, but that is the price to pay when introducing a new material into technology. It is an almost general rule, which accounts for the time needed for production lines to reach the requested mass-scale, reproducibility, reliability, and economic viability according to industrial standards. Taking the example of carbon fibers, first developed in the 1960s, how long did it take to see them in mass consumption, high-tech products (besides sporting goods, for which cost is less of an issue)? It is happening only now, as commercial planes (A350 from Airbus, Dreamliner from Boeing) incorporate significant amounts of carbon fibers in their structures, whereas previously only the military field had been able to afford the benefits of such a technological jump. Despite this, carbon fibers have not yet reached the general automobile field, except for luxury or race cars. In addition, carbon nanotubes have a supplementary drawback: they deal with a nano-world that most existing industrial plants are not familiar with, and not designed for. That is why most applications are currently being taken to the market by start-ups, and it will take time for those to reach economic reliability and expand to become major companies. But some will.

What are the perspectives for carbon nanotubes? They are quite positive, and it is not being overoptimistic to say so: nanotubes exhibit an extraordinary diversity of morphologies, textures, structures and nanotextures, far beyond that provided by fullerenes, and now graphene. As most of their properties have now been experimentally studied or predicted, research should focus on solving the problems related to their incorporation into the tremendous variety of devices and materials that would benefit from the CNT technology. In this regard, considering the nanotube-based nano-objects and composites described in Sect. 8.5 and Sect. 8.6 is of primary importance. Of course, this comes with some uncertainty regarding the potential toxicity related to CNTs at various points of their life cycle, but this issue concerns all nanomaterials in general, and is the most critical challenge upon which the future of nanotechnologies depends.

References

8.1 M. Monthioux, V.L. Kuznetsov: Who should be given the credit for the discovery of carbon nanotubes?, Carbon **44**, 1621–1623 (2006)

8.2 L.V. Radushkevich, V.M. Lukyanovich: O strukture ugleroda, obrazujucegosja pri termiceskom razlozenii okisi ugleroda na zeleznom kontakte, Zurn. Fis. Chim. **26**, 88–95 (1952)

8.3 R.T.K. Baker, P.S. Harris: The formation of filamentous carbon. In: *Chemistry and Physics of Carbon*, Vol. 14, ed. by P.L. Walker Jr., P.A. Thrower (Marcel Dekker, New York 1978) pp. 83–165

8.4 S. Iijima: Helical microtubules of graphite carbon, Nature **354**, 56–58 (1991)

8.5 S. Iijima, T. Ichihashi: Single-shell carbon nanotubes of 1 nm diameter, Nature **363**, 603–605 (1993)

8.6 D.S. Bethune, C.H. Kiang, M.S. de Vries, G. Gorman, R. Savoy, J. Vazquez, R. Bayers: Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls, Nature **363**, 605–607 (1993)

8.7 N. Hamada, S.I. Sawada, A. Oshiyama: New one-dimensional conductors, graphite microtubules, Phys. Rev. Lett. **68**, 1579–1581 (1992)

8.8 J. Tersoff, R.S. Ruoff: Structural properties of a carbon-nanotube crystal, Phys. Rev. Lett. **73**, 676–679 (1994)

8.9 N. Wang, Z.K. Tang, G.D. Li, J.S. Chen: Single-walled 4 Å carbon nanotube arrays, Nature **408**, 50–51

(2000)
8.10 M.S. Dresselhaus, G. Dresselhaus, P.C. Eklund: *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego 1995)
8.11 R.C. Haddon: Chemistry of the fullerenes: The manifestation of strain in a class of continuous aromatic molecules, Science **261**, 1545–1550 (1993)
8.12 M. Monthioux, B.W. Smith, B. Burteaux, A. Claye, J. Fisher, D.E. Luzzi: Sensitivity of single-wall nanotubes to chemical processing: An electron microscopy investigation, Carbon **39**, 1261–1272 (2001)
8.13 H. Allouche, M. Monthioux: Chemical vapor deposition of pyrolytic carbon onto carbon nanotubes. Part II – Structure and texture, Carbon **43**, 1265–1278 (2005)
8.14 M. Audier, A. Oberlin, M. Oberlin, M. Coulon, L. Bonnetain: Morphology and crystalline order in catalytic carbons, Carbon **19**, 217–224 (1981)
8.15 N.M. Rodriguez, A. Chambers, R.T. Baker: Catalytic engineering of carbon nanostructures, Langmuir **11**, 3862–3866 (1995)
8.16 M. Monthioux, L. Noé, L. Dussault, J.-C. Dupin, N. Latorre, T. Ubieto, E. Romeo, C. Royo, A. Monzón, C. Guimon: Texturising and structurising mechanisms of carbon nanofilament during growth, J. Mater. Chem. **17**, 4611–4618 (2007)
8.17 J. Vera-Agullo, H. Varela-Rizo, J.A. Conesa, C. Almansa, C. Merino, I. Martin-Gullon: Evidence for growth mechanism and helix-spiral cone structure of stacked-cup carbon nanofibers, Carbon **45**, 2751–2758 (2007)
8.18 Y. Saito: Nanoparticles and filled nanocapsules, Carbon **33**, 979–988 (1995)
8.19 P.J.F. Harris: *Carbon Nanotubes and Related Structures* (Cambridge Univ. Press, Cambridge 1999)
8.20 M. Monthioux: Introduction to carbon nanotubes. In: *Carbon Meta-Nanotubes: Synthesis, Properties, Applications*, ed. by M. Monthioux (Wiley-Blackwell, Chichester 2012) pp. 8–39
8.21 M. Monthioux, P. Serp, E. Flahaut, M. Razafinimanana, C. Laurent, A. Peigney, W. Bacsa, J.-M. Broto: Introduction to carbon nanotubes. In: *Nanotechnology Handbook*, 3rd edn., ed. by B. Bhushan (Springer, Berlin, Heidelberg 2010) pp. 47–118
8.22 W. Krätschmer, L.D. Lamb, K. Fostiropoulos, D.R. Huffman: Solid C_{60}: A new form of carbon, Nature **347**, 354–358 (1990)
8.23 L. Fulchieri, Y. Schwob, F. Fabry, G. Flamant, L.F.P. Chibante, D. Laplaze: Fullerene production in a 3-phase AC plasma process, Carbon **38**, 797–803 (2000)
8.24 K. Saidane, M. Razafinimanana, H. Lange, A. Huczko, M. Baltas, A. Gleizes, J.L. Meunier: Fullerene synthesis in the graphite electrode arc process: local plasma characteristics and correlation with yield, J. Phys. D Appl. Phys. **37**, 232–239 (2004)
8.25 T.W. Ebbesen, P.M. Ajayan: Large-scale synthesis of carbon nanotubes, Nature **358**, 220–221 (1992)
8.26 D. Ugarte: Morphology and structure of graphitic soot particles generated in arc-discharge C_{60} production, Chem. Phys. Lett. **198**, 596–602 (1992)
8.27 T.W. Ebbesen: Carbon nanotubes, Ann. Rev. Mater. Sci. **24**, 235–264 (1994)
8.28 T. Beltz, J. Find, D. Herein, N. Pfänder, T. Rühle, H. Werner, M. Wohlers, R. Schlögl: On the production of different carbon forms by electric arc graphite evaporation, Ber. Bunsen. Phys. Chem. **101**, 712–725 (1997)
8.29 C. Bernier, W.K. Maser, P. Bernier, A. Loiseau, M.L. de la Chapelle, S. Lefrant, P. Deniard, R. Lee, J.E. Fischer: Large-scale production of single-walled carbon nanotubes by the electric-arc technique, Nature **388**, 756–758 (1997)
8.30 H. Allouche, M. Monthioux, M. Pacheco, M. Razafinimanana, H. Lange, A. Huczko, T.P. Teulet, A. Gleizes, T. Sogabe: Physical characteristics of the graphite-electrode electric-arc as parameters for the formation of single-wall carbon nanotubes. In: *Proc. Eurocarbon*, Vol. 2 (Deutsche Keramische Gesellschaft, Cologne 2000) pp. 1053–1054
8.31 M. Razafinimanana, M. Pacheco, M. Monthioux, H. Allouche, H. Lange, A. Huczko, A. Gleizes: Spectroscopic study of an electric arc with Gd and Fe doped anodes for the carbon nanotube formation. In: *Proc. 25th Int. Conf. Phenom. Ioniz. Gases*, ed. by E. Goto (Nagoya Univ., Nagoya 2001) pp. 297–298
8.32 M. Razafinimanana, M. Pacheco, M. Monthioux, H. Allouche, H. Lange, A. Huczko, P. Teulet, A. Gleizes, C. Goze, P. Bernier, T. Sogabe: Influence of doped graphite electrode in electric arc for the formation of single wall carbon nanotubes. In: *Proc. 6th Eur. Conf. Therm. Plasma Process. – Prog. Plasma Process. Mater.*, ed. by P. Fauchais (Begell House, New York 2001) pp. 649–654
8.33 M. Pacheco, H. Allouche, M. Monthioux, A. Razafinimanana, A. Gleizes: Correlation between the plasma characteristics and the morphology and structure of the carbon phases synthesised by electric arc discharge. In: *Proc. 25th Bienn. Conf. Carbon, Lexington 2001*, ed. by F. Derbyshire (2001)
8.34 M. Pacheco, M. Monthioux, M. Razafinimanana, L. Donadieu, H. Allouche, N. Caprais, A. Gleizes: New factors controlling the formation of single-wall carbon nanotubes by arc plasma. In: *Proc. Carbon 2002 Int. Conf.*, ed. by H.-M. Cheng (Shanxi Chunqiu Audio-Visual, Beijing 2002)
8.35 M. Monthioux, M. Pacheco, H. Allouche, M. Razafinimanana, N. Caprais, L. Donnadieu, A. Gleizes: New data about the formation of SWNTs by the electric arc method. In: *Electron. Prop. Mol. Nanostructures AIP Conf. Proc.*, ed. by H. Kuzmany, J. Fink, M. Mehring, S. Roth (Springer, Berlin, Heidelberg 2002) pp. 182–185
8.36 H. Lange, A. Huczko, M. Sioda, M. Pacheco, M. Razafinimanana, A. Gleizes: Influence of gadolinium on carbon arc plasma and formation of fullerenes and nanotubes, Plasma Chem. Plasma Process **22**, 523–536 (2002)
8.37 A.G. Rinzler, J. Liu, H. Dai, P. Nikolaev, C.B. Huffman, F.J. Rodriguez-Macias, P.J. Boul, A.H. Lu, D. Heymann, D.T. Colbert, R.S. Lee, J.E. Fischer, A.M. Rao,

P.C. Eklund, R.E. Smalley: Large scale purification of single wall carbon nanotubes: Process, product and characterization, Appl. Phys. A **67**, 29–37 (1998)

8.38 M. Ishigami, J. Cumings, A. Zettl, S. Chen: A simple method for the continuous production of carbon nanotubes, Chem. Phys. Lett. **319**, 457–459 (2000)

8.39 Y.L. Hsin, K.C. Hwang, F.R. Chen, J.J. Kai: Production and in-situ metal filling of carbon nanotube in water, Adv. Mater. **13**, 830–833 (2001)

8.40 H.W. Zhu, X.S. Li, B. Jiang, C.L. Xu, C.L. Zhu, Y.F. Zhu, D.H. Wu, X.H. Chen: Formation of carbon nanotubes in water by the electric arc technique, Chem. Phys. Lett. **366**, 664–669 (2002)

8.41 T. Sogabe, T. Masuda, K. Kuroda, Y. Hirohaya, T. Hino, T. Ymashina: Preparation of B_4C-mixed graphite by pressureless sintering and its air oxidation behavior, Carbon **33**, 1783–1788 (1995)

8.42 W.K. Maser, P. Bernier, J.M. Lambert, O. Stephan, P.M. Ajayan, C. Colliex, V. Brotons, J.M. Planeix, B. Coq, P. Molinie, S. Lefrant: Elaboration and characterization of various carbon nanostructures, Synth. Met. **81**, 243–250 (1996)

8.43 A. Mansour, M. Razafinimanana, M. Monthioux, M. Pacheco, A. Gleizes: A significant improvement of both yield and purity during SWCNT synthesis via the electric arc process, Carbon **45**, 1651–1661 (2007)

8.44 A. Mansour: *Caractérisation expérimentale d'un plasma d'arc électrique en vue du contrôle de la synthèse des nanotubes de carbone monoparois*, Ph.D. Thesis (University Paul Sabatier, Toulouse 2007)

8.45 J. Gavillet, A. Loiseau, J. Thibault, A. Maigné, O. Stéphan, P. Bernier: TEM study of the influence of the catalyst composition on the formation and growth of SWNT. In: *Proc. Electron. Prop. Nov. Mater. – XVI Int. Wintersch. – AIP Conf.*, ed. by H. Kuzmany, J. Fink, M. Mehring, S. Roth (Springer, Berlin, Heidelberg 2002) pp. 202–206

8.46 V. Ramarozatovo, A. Mansour, M. Razafinimanana, M. Monthioux, F. Valensi, L. Noé, M. Masquère: Influence of chamber volume in single wall carbon nanotube synthesis by an electric arc, J. Phys. D: Appl. Phys. **45**, 345204 (2012)

8.47 G.G. Tibbetts, M. Endo, C.P. Beetz: Carbon fibers grown from the vapor phase: A novel material, SAMPE J. **22**, 30 (1989)

8.48 R.T.K. Baker: Catalytic growth of carbon filaments, Carbon **27**, 315–323 (1989)

8.49 E. Lamouroux, P. Serp, P. Kalck: Catalytic chemical vapor deposition routes towards single-walled and double-walled carbon nanotubes, Catal. Rev. Sci. Eng. **49**, 341–405 (2007)

8.50 R. Philippe, A. Morançais, M. Corrias, B. Caussat, Y. Kihn, P. Kalck, D. Plee, P. Gaillard, D. Bernard, P. Serp: Catalytic production of carbon nanotubes by fluidized-bed CVD, Chem. Vap. Depos. **13**, 447–457 (2007)

8.51 R.T.K. Baker, P.S. Harris, R.B. Thomas, R.J. Waite: Formation of filamentous carbon from iron, cobalt, and chromium catalyzed decomposition of acetylene, J. Catal. **30**, 86–95 (1973)

8.52 T. Koyama, M. Endo, Y. Oyuma: Carbon fibers obtained by thermal decomposition of vaporized hydrocarbon, Jpn. J. Appl. Phys. **11**, 445–449 (1972)

8.53 M. Endo, A. Oberlin, T. Koyama: High resolution electron microscopy of graphitizable carbon fiber prepared by benzene decomposition, Jpn. J. Appl. Phys. **16**, 1519–1523 (1977)

8.54 N.M. Rodriguez: A review of catalytically grown carbon nanofibers, J. Mater. Res. **8**, 3233–3250 (1993)

8.55 W.R. Davis, R.J. Slawson, G.R. Rigby: An unusual form of carbon, Nature **171**, 756 (1953)

8.56 H.P. Boehm: Carbon from carbon monoxide disproportionation on nickel and iron catalysts; morphological studies and possible growth mechanisms, Carbon **11**, 583–590 (1973)

8.57 M. Audier, A. Oberlin, M. Coulon: Crystallographic orientations of catalytic particles in filamentous carbon; case of simple conical particles, J. Cryst. Growth **55**, 546–549 (1981)

8.58 M. Audier, M. Coulon: Kinetic and microscopic aspects of catalytic carbon growth, Carbon **23**, 317–323 (1985)

8.59 A. Thaib, G.A. Martin, P. Pinheiro, M.C. Schouler, P. Gadelle: Formation of carbon nanotubes from the carbon monoxide disproportionation reaction over Co/Al_2O_3 and Co/SiO_2 catalysts, Catal. Lett. **63**, 135–141 (1999)

8.60 P. Pinheiro, M.C. Schouler, P. Gadelle, M. Mermoux, E. Dooryhée: Effect of hydrogen on the orientation of carbon layers in deposits from the carbon monoxide disproportionation reaction over Co/Al_2O_3 catalysts, Carbon **38**, 1469–1479 (2000)

8.61 C. Laurent, E. Flahaut, A. Peigney, A. Rousset: Metal nanoparticles for the catalytic synthesis of carbon nanotubes, New J. Chem. **22**, 1229–1237 (1998)

8.62 A. Peigney, C. Laurent, F. Dobigeon, A. Rousset: Carbon nanotubes grown in situ by a novel catalytic method, J. Mater. Res. **12**, 613–615 (1997)

8.63 D. Kunii, O. Levenspiel: *Fluidization Engineering* (Butterworth-Heinemann, Boston 1991)

8.64 S. Shukrullah, N.M. Mohamed, M.S. Shaharun, M.Y. Naz: Mass production of carbon nanotubes using fluidized bed reactor: A short review, Trends Appl. Sci. Res. **9**, 121–131 (2014)

8.65 E. Flahaut: *Synthèse par voir catalytique et caractérisation de composites nanotubes de carbone-metal-oxyde poudres et matériaux denses*, Ph.D. Thesis (Univers. Paul Sabatier, Toulouse 1999)

8.66 E. Flahaut, R. Bacsa, A. Peigney, C. Laurent: Gram-scale CCVD synthesis of double-walled carbon nanotubes, Chem. Commun. (2003) doi:10.1039/B301514A

8.67 V. Ivanov, J.B. Nagy, P. Lambin, A. Lucas, X.B. Zhang, X.F. Zhang, D. Bernaerts, G. Van Tendeloo, S. Amelinckx, J. Van Landuyt: The study of nanotubules produced by catalytic method, Chem. Phys. Lett. **223**, 329–335 (1994)

8.68 V. Ivanov, A. Fonseca, J.B. Nagy, A. Lucas, P. Lambin, D. Bernaerts, X.B. Zhang: Catalytic production and purification of nanotubules having fullerene-scale diameters, Carbon **33**, 1727–1738 (1995)

8.69 K. Hernadi, A. Fonseca, J.B. Nagy, D. Bernaerts, A. Fudala, A. Lucas: Catalytic synthesis of carbon nanotubes using zeolite support, Zeolites **17**, 416–423 (1996)

8.70 Y.H. Li, H.Q. Gao, J.H. Yang, W.L. Gao, J. Xiang, Q.Y. Li: Multi-wall carbon nanotubes supported on carbon fiber paper synthesized by simple chemical vapor deposition, Mat. Sci.Eng. B **187**, 113–119 (2014)

8.71 A. Govindaraj, E. Flahaut, C. Laurent, A. Peigney, A. Rousset, C.N.R. Rao: An investigation of carbon nanotubes obtained from the decomposition of methane over reduced $Mg_{1-x}M_xAl_2O_4$ spinel catalysts, J. Mater. Res. **14**, 2567–2576 (1999)

8.72 E. Flahaut, A. Peigney, C. Laurent, A. Rousset: Synthesis of single-walled carbon nanotube-Co-MgO composite powders and extraction of the nanotubes, J. Mater. Chem. **10**, 249–252 (2000)

8.73 J. Kong, A.M. Cassel, H. Dai: Chemical vapor deposition of methane for single-walled carbon nanotubes, Chem. Phys. Lett. **292**, 567–574 (1998)

8.74 E. Flahaut, A. Peigney, W.S. Bacsa, R.R. Bacsa, C. Laurent: CCVD synthesis of carbon nanotubes from (Mg,Co,Mo)O catalysts: Influence of the proportions of cobalt and molybdenum, J. Mater. Chem. **14**, 646–653 (2004)

8.75 E. Flahaut, C. Laurent, A. Peigney: Catalytic CVD synthesis of double and triple-walled carbon nanotubes by the control of the catalyst preparation, Carbon **43**, 375–383 (2005)

8.76 R. Marangoni, P. Serp, R. Feurrer, Y. Kihn, P. Kalck, C. Vahlas: Carbon nanotubes produced by substrate free metalorganic chemical vapor deposition of iron catalyst and ethylene, Carbon **39**, 443–449 (2001)

8.77 R. Sen, A. Govindaraj, C.N.R. Rao: Carbon nanotubes by the metallocene route, Chem. Phys. Lett. **267**, 276–280 (1997)

8.78 Y.Y. Fan, H.M. Cheng, Y.L. Wei, G. Su, S.H. Shen: The influence of preparation parameters on the mass production of vapor grown carbon nanofibers, Carbon **38**, 789–795 (2000)

8.79 L. Ci, J. Wei, B. Wei, J. Liang, C. Xu, D. Wu: Carbon nanofibers and single-walled carbon nanotubes prepared by the floating catalyst method, Carbon **39**, 329–335 (2001)

8.80 F. Toni, H. Xing, J. Walter, V. Strauss, T.J. Nacken, C. Damma, K.E. Wirth, D. Guldi, W. Peukert: Production of well dispersible single walled carbon nanotubes via a floating catalyst-method, Chem. Eng. Sci. **138**, 385–395 (2015)

8.81 N. Sankararamakrishnan, D. Chauhan, J. Dwivedi: Synthesis of functionalized carbon nanotubes by floating catalytic chemical vapor deposition method and their sorption behavior towards arsenic, Chem. Eng. J. **284**, 599–608 (2016)

8.82 M. Glerup, H. Kanzow, R. Almairac, M. Castignolles, P. Bernier: Synthesis of multi-walled carbon nanotubes and nano-fibres using aerosol method with metal-ions as the catalyst precursors, Chem. Phys. Lett. **377**, 293–298 (2003)

8.83 Z. Zhou, L. Ci, L. Song, X. Yan, D. Liu, H. Yuan, Y. Gao, J. Wang, L. Liu, W. Zhou, G. Wang, S. Xie: Producing cleaner double-walled carbon nanotubes in a floating catalyst system, Carbon **41**, 2607–2611 (2003)

8.84 F. Rohmund, L.K.L. Falk, F.E.B. Campbell: A simple method for the production of large arrays of aligned carbon nanotubes, Chem. Phys. Lett. **328**, 369–373 (2000)

8.85 G.G. Tibbetts, C.A. Bernardo, D.W. Gorkiewicz, R.L. Alig: Role of sulfur in the production of carbon fibers in the vapor phase, Carbon **32**, 569–576 (1994)

8.86 H. Chen, M. Chen, Y. Zhang, Q. Li: Rational control on floating catalysts for the growth of carbon nanotube assemblies: From vertically aligned carbon nanotube arrays to carbon nanotube films, Appl. Surf. Sci. **353**, 651–661 (2015)

8.87 W.Q. Han, P. Kholer-Riedlich, T. Seeger, F. Ernst, M. Ruhle, N. Grobert, W.K. Hsu, B.H. Chang, Y.Q. Zhu, H.W. Kroto, M. Terrones, H. Terrones: Aligned CN_x nanotubes by pyrolysis of ferrocene under NH_3 atmosphere, Appl. Phys. Lett. **77**, 1807–1809 (2000)

8.88 L. Ci, Z. Rao, Z. Zhou, D. Tang, X. Yan, Y. Liang, D. Liu, H. Yuan, W. Zhou, G. Wang, W. Liu, S. Xie: Double wall carbon nanotubes promoted by sulfur in a floating iron catalyst CVD system, Chem. Phys. Lett. **359**, 63–67 (2002)

8.89 S. Maruyama, R. Kojima, Y. Miyauchi, S. Chiashi, M. Kohno: Low-temperature synthesis of high-purity single-walled carbon nanotubes from alcohol, Chem. Phys. Lett. **360**, 229–234 (2002)

8.90 R.E. Smalley, J.H. Hafner, D.T. Colbert, K. Smith: Catalytic growth of single-wall carbon nanotubes from metal particles, US Patent (Application) 19980601010903 (1998)

8.91 P. Nikolaev: Gas-phase production of single-walled carbon nanotubes from carbon monoxide: A review of the HiPco process, J. Nanosci. Nanotechnol. **4**, 307–316 (2004)

8.92 T. Kyotani, L.F. Tsai, A. Tomita: Preparation of ultrafine carbon tubes in nanochannels of an anodic aluminum oxide film, Chem. Mater. **8**, 2109–2113 (1996)

8.93 T. Guo, P. Nikolaev, A. Thess, D.T. Colbert, R.E. Smalley: Catalytic growth of single-walled nanotubes by laser vaporisation, Chem. Phys. Lett. **243**, 49–54 (1995)

8.94 A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y.H. Lee, S.G. Kim, D.T. Colbert, G. Scuseria, D. Tomanek, J.E. Fischer, R.E. Smalley: Crystalline ropes of metallic carbon nanotubes, Science **273**, 487–493 (1996)

8.95 M. Yudasaka, T. Komatsu, T. Ichihashi, S. Iijima: Single wall carbon nanotube formation by laser ablation using double targets of carbon and metal, Chem. Phys. Lett. **278**, 102–106 (1997)

8.96 T.M. Gruenberger, J. Gonzalez-Aguilar, F. Fabry, L. Fulchieri, E. Grivei, N. Probst, G. Flamant, H. Okuno, J.C. Charlier: Production of carbon nanotubes and other nanostructures via continuous 3-phase AC plasma processing, Fuller. Nanotub. Carbon Nanostruct. **12**, 571–581 (2004)

8.97 M.J. Heben, T.A. Bekkedhal, D.L. Schultz, K.M. Jones, A.C. Dillon, C.J. Curtis, C. Bingham, J.R. Pitts, A. Lewandowski, C.L. Fields: Production of single wall carbon nanotubes using concentrated sunlight. In: *Proc. Symp. Recent Adv. Chem. Phys. Fuller. Rel. Mater.*, ed. by K.M. Kadish, R.S. Ruoff (Electrochemical Society, Pennington 1996) pp. 803–811

8.98 D. Laplaze, P. Bernier, C. Journet, G. Vié, G. Flamant, E. Philippot, M. Lebrun: Evaporation of graphite using a solar furnace. In: *Proc. 8th Int. Symp. Solar Conc. Technol., Köln, 1996*, ed. by M. Becker, M. Balmer (Müller, Heidelberg 1997) pp. 1653–1656

8.99 G. Flamant, M. Bijeire, D. Luxembourg: Modelling of a solar reactor for single wall nanotubes synthesis, ASME J. Solar Energ. Eng. **128**, 1–124 (2006)

8.100 W.K. Hsu, J.P. Hare, M. Terrones, H.W. Kroto, D.R.M. Walton, P.J.F. Harris: Condensed-phase nanotubes, Nature **377**, 687 (1995)

8.101 W.S. Cho, E. Hamada, Y. Kondo, K. Takayanagi: Synthesis of carbon nanotubes from bulk polymer, Appl. Phys. Lett. **69**, 278–279 (1996)

8.102 Y.L. Li, Y.D. Yu, Y. Liang: A novel method for synthesis of carbon nanotubes: Low temperature solid pyrolysis, J. Mater. Res. **12**, 1678–1680 (1997)

8.103 M.L. Terranova, S. Piccirillo, V. Sessa, P. Sbornicchia, M. Rossi, S. Botti, D. Manno: Growth of single-walled carbon nanotubes by a novel technique using nanosized graphite as carbon source, Chem. Phys. Lett. **327**, 284–290 (2000)

8.104 R.L. Vander Wal, T. Ticich, V.E. Curtis: Diffusion flame synthesis of single-walled carbon nanotubes, Chem. Phys. Lett. **323**, 217–223 (2000)

8.105 I. Gunjishima, T. Inoue, S. Yamamuro, K. Sumiyama, A. Okamoto: Synthesis of vertically aligned, double-walled carbon nanotubes from highly active Fe-V-O nanoparticles, Carbon **45**, 1193–1199 (2007)

8.106 G. Zhong, T. Iwasaki, J. Robertson, H. Kawarada: Growth kinetics of 0.5 cm vertically aligned single-walled carbon nanotubes, J. Phys. Chem. B **111**, 1907–1910 (2007)

8.107 H. Cui, G. Eres, J.Y. Howe, A. Puretzki, M. Varela, D.B. Geohegan, D.H. Lowndes: Growth behavior of carbon nanotubes on multilayered metal catalyst film in chemical vapor deposition, Chem. Phys. Lett. **374**, 222–228 (2003)

8.108 Y.Y. Wei, G. Eres, V.I. Merkulov, D.H. Lowdens: Effect of film thickness on carbon nanotube growth by selective area chemical vapor deposition, Appl. Phys. Lett. **78**, 1394–1396 (2001)

8.109 I.T. Han, B.K. Kim, H.J. Kim, M. Yang, Y.W. Jin, S. Jung, N. Lee, S.K. Kim, J.M. Kim: Effect of Al and catalyst thickness on the growth of carbon nanotubes and application to gated field emitter arrays, Chem. Phys. Lett. **400**, 139–144 (2004)

8.110 W.Z. Li, S.S. Xie, L.X. Qian, B.H. Chang, B.S. Zou, W.Y. Zhou, R.A. Zha, G. Wang: Large scale synthesis of aligned carbon nanotubes, Science **274**, 1701–1703 (1996)

8.111 F. Zheng, L. Liang, Y. Gao, J.H. Sukamto, L. Aardahl: Carbon nanotubes synthesis using mesoporous silica templates, Nano Lett. **2**, 729–732 (2002)

8.112 S.H. Jeong, O.-K. Lee, K.H. Lee, S.H. Oh, C.G. Park: Preparation of aligned carbon nanotubes with prescribed dimension: Template synthesis and sonication cutting approach, Chem. Mater. **14**, 1859–1862 (2002)

8.113 N.S. Kim, Y.T. Lee, J. Park, H. Ryu, H.J. Lee, S.Y. Choi, J. Choo: Dependence of vertically aligned growth of carbon nanotubes on catalyst, J. Phys. Chem. B **106**, 9286–9290 (2002)

8.114 C.J. Lee, D.W. Kim, T.J. Lee, Y.C. Choi, Y.S. Park, Y.H. Lee, W.B. Choi, N.S. Lee, G.-S. Park, J.M. Kim: Synthesis of aligned carbon nanotubes using thermal chemical vapor deposition, Chem. Phys. Lett. **312**, 461–468 (1999)

8.115 W.D. Zhang, Y. Wen, S.M. Liu, W.C. Tjiu, G.Q. Xu, L.M. Gan: Synthesis of vertically aligned carbon nanotubes on metal deposited quartz plates, Carbon **40**, 1981–1989 (2002)

8.116 S. Huang, L. Dai, A.W.H. Mau: Controlled fabrication of large scale aligned carbon nanofiber/nanotube patterns by photolithography, Adv. Mater. **14**, 1140–1143 (2002)

8.117 T. Sun, G. Wang, H. Liu, L. Feng, D. Zhu: Control over the wettability of an aligned carbon nanotube film, J. Am. Chem. Soc. **125**, 14996–14997 (2003)

8.118 C.L. Cheung, A. Kurtz, H. Park, C.M. Lieber: Diameter-controlled synthesis of carbon nanotubes, J. Phys. Chem. B **106**, 2429–2433 (2002)

8.119 Y. Huh, J.Y. Lee, J. Cheon, Y.K. Hong, J.Y. Koo, T.J. Lee, C.J. Lee: Controlled growth of carbon nanotubes over cobalt nanoparticles by thermal chemical vapor deposition, J. Mater. Chem. **13**, 2297–2300 (2003)

8.120 M. Paillet, V. Jourdain, P. Poncharal, J.-L. Sauvajol, A. Zahab, J.C. Meyer, S. Roth, N. Cordente, C. Amiens, B. Chaudret: Versatile synthesis of individual single-walled carbon nanotubes from nickel nanoparticles for the study of their physical properties, J. Phys. Chem. B **108**, 17112–17118 (2004)

8.121 Y. Kobayashi, H. Nakashima, D. Takagi, Y. Homma: CVD growth of single-walled carbon nanotubes using size-controlled nanoparticle catalyst, Thin Solid Films **464/465**, 286–289 (2004)

8.122 S. Casimirius, E. Flahaut, C. Laurent, C. Vieu, F. Carcenac, C. Laberty-Robert: Optimized microcontact printing process for the patterned growth of individual SWNTs, Microelectron. Eng. **73/74**, 564–569 (2004)

8.123 Y. Lei, K.S. Yeong, J.T.L. Thong, W.K. Chim: Large-scale ordered carbon nanotubes arrays initiated from highly ordered catalyst arrays on silicon substrates, Chem. Mater. **16**, 2757–2761 (2004)

8.124 Q. Ye, A.M. Cassel, H. Liu, K.J. Chao, J. Han, M. Meyyappan: Large-scale fabrication of carbon nanotube probe tips for atomic force microscopy critical dimension imaging applications, Nano Lett. **4**, 1301–1308 (2004)

8.125 S. Fan, M. Chapline, N. Franklin, T. Tombler, A.M. Cassel, H. Dai: Self-oriented regular arrays of carbon nanotubes and their field emission properties, Science **283**, 512–514 (1999)

8.126 K. Hata, D.N. Futaba, K. Mizuno, T. Namai, M. Yumara, S. Iijima: Ware-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes, Science **306**, 1362–1364 (2004)

8.127 Y. Fu, S. Chen, J. Bielecki, A. Matic, T. Wang, L.-L. Ye, J. Liu: Selective growth of double walled carbon nanotubes on gold films, Mater. Lett. **72**, 78–80 (2012)

8.128 A.M. Cassel, N.R. Franklin, T.W. Tombler, E.M. Chan, J. Han, H. Dai: Directed growth of free-standing single-walled carbon nanotubes, J. Am. Chem. Soc. **121**, 7975–7976 (1999)

8.129 R. Andrews, D. Jacques, A.M. Rao, F. Derbyshire, D. Qian, X. Fan, E.C. Dickey, J. Chen: Continous production of aligned carbon nanotubes: A step closer to commercial realization, Chem. Phys. Lett. **303**, 467–474 (1999)

8.130 C.N.R. Rao, R. Sen, B.C. Satishkumar, A. Govindaraj: Large aligned carbon nanotubes bundles from ferrocene pyrolysis, Chem. Commun. (1998) doi:10.1039/A802258E

8.131 K.S. Choi, Y.S. Cho, S.Y. Hong, J.B. Park, D.J. Kim: Effects of ammonia on the alignment of carbon nanotubes in metal-assisted chemical vapor deposition, J. Eur. Ceram. Soc. **21**, 2095–2098 (2001)

8.132 N.S. Kim, Y.T. Lee, J. Park, J.B. Han, Y.S. Choi, S.Y. Choi, J. Choo, G.H. Lee: Vertically aligned carbon nanotubes grown by pyrolysis of iron, cobalt, and nickel phthalocyanines, J. Phys. Chem. B **107**, 9249–9255 (2003)

8.133 C. Emmeger, J.M. Bonard, P. Mauron, P. Sudan, A. Lepora, B. Grobety, A. Züttel, L. Schlapbach: Synthesis of carbon nanotubes over Fe catalyst on aluminum and suggested growth mechanism, Carbon **41**, 539–547 (2003)

8.134 Q. Zhang, J. Huang, F. Wei, G. Xu, Y. Wang, W. Qian, D. Wang: Large scale production of carbon nanotubes arrays on the sphere surface from liquefied petroleum gas at low cost, Chin. Sci. Bull. **52**, 2896–2902 (2007)

8.135 X. Li, L. Zhang, X. Wang, I. Shimoyama, X. Sun, W.-S. Seo, H. Dai: Assembly of densely aligned single-walled carbon nanotubes from bulk materials Langmuir–Blodgett, J. Am. Chem. Soc. **129**, 4890–4891 (2007)

8.136 M. Kumar, Y. Ando: Chemical vapor deposition of carbon nanotubes: A review on growth mechanism and mass production, J. Nanosci. Nanotechnol. **10**, 3739–3758 (2010)

8.137 J.-P. Tessonnier, D.S. Su: Recent progress on the growth mechanism of carbon nanotubes: A review, Chem. Sus. Chem. **4**, 824–847 (2011)

8.138 V. Jourdain, C. Bichara: Current understanding of the growth of carbon nanotubes in catalytic chemical vapour deposition, Carbon **58**, 2–39 (2013)

8.139 M. Endo, H.W. Kroto: Formation of carbon nanofibers, J. Phys. Chem. **96**, 6941–6944 (1992)

8.140 R.S. Wagner: VLS mechanisms of crystal growth. In: *Whisker Technology*, ed. by P.A. Levit (Wiley, New York 1970) pp. 47–72

8.141 V. Jourdain, H. Kanzow, M. Castignolles, A. Loiseau, P. Bernier: Sequential catalytic growth of carbon nanotubes, Chem. Phys. Lett. **364**, 27–33 (2002)

8.142 H. Dai: Carbon nanotubes: Synthesis, integration, and properties, Acc. Chem. Res. **35**, 1035–1044 (2002)

8.143 Y. Saito, M. Okuda, N. Fujimoto, T. Yoshikawa, M. Tomita, T. Hayashi: Single-wall carbon nanotubes growing radially from Ni fine particles formed by arc evaporation, Jpn. J. Appl. Phys. **33**, L526–L529 (1994)

8.144 J. Bernholc, C. Brabec, M. Buongiorno Nardelli, A. Malti, C. Roland, B.J. Yakobson: Theory of growth and mechanical properties of nanotubes, Appl. Phys. A **67**, 39–46 (1998)

8.145 F. Larouche, O. Smiljanic, X. Sun, B.L. Stanfield: Solutal Bénard–Marangoni instability as a growth mechanism for single-walled carbon nanotubes, Carbon **43**, 986–993 (2005)

8.146 K. Méténier, S. Bonnamy, F. Béguin, C. Journet, P. Bernier, L.M. de la Chapelle, O. Chauvet, S. Lefrant: Coalescence of single walled nanotubes and formation of multi-walled carbon nanotubes under high temperature treatments, Carbon **40**, 1765–1773 (2002)

8.147 P.G. Collins, P. Avouris: Nanotubes for electronics, Sci. Am. **283**, 38–45 (2000)

8.148 Q.-H. Yang, P.X. Hou, S. Bai, M.Z. Wang, H.M. Cheng: Adsorption and capillarity of nitrogen in aggregated multi-walled carbon nanotubes, Chem. Phys. Lett. **345**, 18–24 (2001)

8.149 S. Inoue, N. Ichikuni, T. Suzuki, T. Uematsu, K. Kaneko: Capillary condensation of N_2 on multiwall carbon nanotubes, J. Phys. Chem. **102**, 4689–4692 (1998)

8.150 M.E. Birch, T.A. Ruda-Eberenz, M. Chai, R. Andrews, R.L. Hatfield: Properties that influence the specific surface areas of carbon nanotubes and nanofibers, Ann. Occup. Hyg. **57**, 1148–1166 (2013)

8.151 M. Eswaramoorthy, R. Sen, C.N.R. Rao: A study of micropores in single-walled carbon nanotubes by the adsorption of gases and vapors, Chem. Phys. Lett. **304**, 207–210 (1999)

8.152 A. Peigney, C. Laurent, E. Flahaut, R.R. Bacsa, A. Rousset: Specific surface area of carbon nanotubes and bundles of carbon nanotubes, Carbon **39**, 507–514 (2001)

8.153 K.A. Williams, P.C. Eklund: Monte Carlo simulation of H_2 physisorption in finite diameter carbon nanotube ropes, Chem. Phys. Lett. **320**, 352–358 (2000)

8.154 M. Cinke, J. Li, B. Chen, A. Cassell, L. Delzeit, J. Han, M. Meyyappan: Pore structure of raw and purified HiPco single-walled carbon nanotubes, Chem. Phys. Lett. **365**, 69–74 (2002)

8.155 E. Frackowiak, S. Delpeux, K. Jurewicz, K. Szostak, D. Cazorla-Amoros, F. Béguin: Enhanced capacitance of carbon nanotubes through chemical activation, Chem. Phys. Lett. **336**, 35–41 (2002)

8.156 E. Raymundo-Pinhero, P. Azaïs, T. Cacciaguerra, D. Cazorla-Amorós, A. Linares-Solano, F. Béguin: KOH and NaOH activation mechanisms of multiwalled carbon nanotubes with different structural organisation, Carbon **43**, 786–795 (2005)

8.157 R. Axet, R. Bacsa, B. Machado, P. Serp: Adsorption on and reactivity of carbon nanotubes and graphene. In: *Handbook of Carbon Nano Materials*, Vol. 5, ed. by F. D'Souza, K.M. Kadish (World Scientific, Singapore 2014) pp. 39–184

8.158 Z. Chen, W. Thiel, A. Hirsch: Reactivity of the convex and concave surfaces of single-walled carbon nanotubes (SWCNTs) towards addition reactions: Dependence on the carbon-atom pyramidalization, Chem. Phys. Chem. **1**, 93–97 (2003)

8.159 K. Azizi, S. Majid Hashemianzadeh, S. Bahramifar: Density functional theory study of carbon monoxide adsorption on the inside and outside of the armchair single-walled carbon nanotubes, Curr. Appl. Phys. **11**, 776–782 (2011)

8.160 M. Muris, N. Dupont-Pavlosky, M. Bienfait, P. Zeppenfeld: Where are the molecules adsorbed on single-walled nanotubes?, Surf. Sci. **492**, 67–74 (2001)

8.161 R.B. Hallock, Y.H. Yang: Adsorption of helium and other gases to carbon nanotubes and nanotubes bundles, J. Low Temp. Phys. **134**, 21–30 (2004)

8.162 J. Zhu, Y. Wang, W. Li, F. Wei, Y. Yu: Density functional study of nitrogen adsorption in single-wall carbon nanotubes, Nanotechnology **18**, 095707 (2007)

8.163 J. Zhao, A. Buldum, J. Han, J.P. Lu: Gas molecule adsorption in carbon nanotubes and nanotube bundles, Nanotechnology **13**, 195–200 (2002)

8.164 C. Matranga, B. Bockrath: Hydrogen-bonded and physisorbed CO in single-walled carbon nanotubes bundles, J. Phys. Chem. B **109**, 4853–4864 (2005)

8.165 D.V. Kazachkin, Y. Nishimura, S. Irle, X. Feng, R. Vidic, E. Borguet: Temperature and pressure dependence of molecular adsorption on single wall carbon nanotubes and the existence of an adsorption/desorption pressure gap, Carbon **48**, 1867–1875 (2010)

8.166 A. Fujiwara, K. Ishii, H. Suematsu, H. Kataura, Y. Maniwa, S. Suzuki, Y. Achiba: Gas adsorption in the inside and outside of single-walled carbon nanotubes, Chem. Phys. Lett. **336**, 205–211 (2001)

8.167 C.M. Yang, H. Kanoh, K. Kaneko, M. Yudasaka, S. Iijima: Adsorption behaviors of HiPco single-walled carbon nanotubes aggregates for alcohol vapors, J. Phys. Chem. **106**, 8994–8999 (2002)

8.168 H. Ulbricht, G. Moos, T. Hertel: Physisorption of molecular oxygen on single-wall carbon nanotube bundles and graphite, Phys. Rev. B **66**, 075404–1–075404–7 (2002)

8.169 H. Ulbricht, J. Kriebel, G. Moos, T. Hertel: Desorption kinetics and interaction of Xe with single-wall carbon nanotube bundles, Chem. Phys. Lett. **363**, 252–260 (2002)

8.170 J. Hilding, E.A. Grulke, S.B. Sinnott, D. Qian, R. Andrews, M. Jagtoyen: Sorption of butane on carbon multiwall nanotubes at room temperature, Langmuir **17**, 7540–7544 (2001)

8.171 A. Bilic, J.D. Gale: Chemisorption of molecular hydrogen on carbon nanotubes: A route to effective hydrogen storage?, J. Phys. Chem. C **112**, 12568–12575 (2008)

8.172 M. Hatami, A. Farmany, R. Sahraei: Physisorption and chemisorption of oxygen molecules on single- and multi-walled carbon nanotubes, Fuller. Nanot. Carbon Nanostruct. **22**, 434–453 (2014)

8.173 E. Durgun, S. Dag, V.M.K. Bagci, O. Gülseren, T. Yildirim, S. Ciraci: Systematic study of adsorption of single atoms on a carbon nanotube, Phys. Rev. B **67**, 201401 (2003)

8.174 N. Park, D. Sung, S. Lim, S. Moon, S. Hong: Realistic adsorption geometries and binding affinities of metal nanoparticles onto the surface of carbon nanotubes, Appl. Phys. Lett. **94**, 073105 (2009)

8.175 J.-C. Charlier, X. Blase, S. Roche: Electronic and transport properties of carbon nanotubes, Rev. Mod. Phys. **79**, 677–732 (2007)

8.176 R. Saito, G. Dresselhaus, M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College, London 1998)

8.177 A. Charlier, E. McRae, R. Heyd, M.F. Charlier, D. Moretti: Classification for double-walled carbon nanotubes, Carbon **37**, 1779–1783 (1999)

8.178 A. Charlier, E. McRae, R. Heyd, M.F. Charlier: Metal semi-conductor transitions under uniaxial stress for single- and double-walled carbon nanotubes, J. Phys. Chem. Solids **62**, 439–444 (2001)

8.179 P. Delanay, H.J. Choi, J. Ihm, S.G. Louie, M.L. Cohen: Broken symmetry and pseudogaps in ropes of carbon nanotubes, Nature **391**, 466–468 (1998)

8.180 P. Puech, H. Hubel, D. Dunstan, R.R. Bacsa, C. Laurent, W.S. Bacsa: Discontinuous tangential stress in double wall carbon nanotubes, Phys. Rev. Lett. **93**, 095506 (2004)

8.181 P.M. Ajayan, M. Terrrones, A. de la Guardia, V. Hue, N. Grobert, B.Q. Wei, H. Lezec, G. Ramanath, T.W. Ebbesen: Nanotubes in a flash – ignition and reconstruction, Science **296**, 705 (2002)

8.182 H. Ajiki, T. Ando: Electronic states of carbon nanotubes, J. Phys. Soc. Jpn. **62**, 1255–1266 (1993)

8.183 T. Ando: Excitons in carbon nanotubes, J. Phys. Soc. Jpn. **66**, 1066 (1997)

8.184 S.M. Bachilo, M.S. Strano, C. Kittrell, R.H. Hauge, R.E. Smalley, R.B. Weisman: Structure-assigned optical spectra of single-walled carbon nanotubes, Science **298**, 2361 (2002)

8.185 M. Bockrath, D.H. Cobden, J. Lu, A.G. Rinzler, R.E. Smalley, L. Balents, P.L. McEuen: Luttinger liquid behaviour in carbon nanotubes, Nature **397**, 598–601 (1999)

8.186 C.T. White, T.N. Todorov: Carbon nanotubes as long ballistic conductors, Nature **393**, 240–242 (1998)

8.187 S. Frank, P. Poncharal, Z.L. Wang, W.A. de Heer: Carbon nanotube quantum resistors, Science **280**, 1744–1746 (1998)

8.188 W. Liang, M. Bockrath, D. Bozovic, J.H. Hafner, M. Tinkham, H. Park: Fabry–Perot interference in a nanotube electron waveguide, Nature **411**, 665–669 (2001)

8.189 L. Langer, V. Bayot, E. Grivei, J.-P. Issi, J.-P. Heremans, C.H. Olk, L. Stockman, C. van Haesendonck, Y. Buynseraeder: Quantum transport in a multiwalled carbon nanotube, Phys. Rev. Lett. **76**, 479–482 (1996)

8.190 K. Liu, S. Roth, G.S. Duesberg, G.T. Kim, D. Popa, K. Mukhopadhyay, R. Doome, J. B'Nagy: Antilocalization in multiwalled carbon nanotubes, Phys. Rev. B **61**, 2375–2379 (2000)

8.191 G. Fedorov, B. Lassagne, M. Sagnes, B. Raquet, J.M. Broto, F. Triozon, S. Roche, E. Flahaut: Gate-dependent magnetoresistance phenomena in carbon nanotubes, Phys. Rev. Lett. **94**, 66801–66804 (2005)

8.192 A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai: Ballistic carbon nanotube field-effect transistors, Nature **424**, 654–657 (2003)

8.193 Y.A. Kasumov, R. Deblock, M. Kociak, B. Reulet, H. Bouchiat, I.I. Khodos, Y.B. Gorbatov, V.T. Volkov, C. Journet, M. Burghard: Supercurrents through single-walled carbon nanotubes, Science **284**, 1508–1511 (1999)

8.194 B.W. Alphenaar, K. Tsukagoshi, M. Wagner: Magnetoresistance of ferromagnetically contacted carbon nanotubes, Phys. Eng. **10**, 499–504 (2001)

8.195 S. Berber, Y. Kwon, D. Tomanek: Unusually high thermal conductivity of carbon nanotubes, Phys. Rev. Lett. **84**, 4613–4616 (2000)

8.196 M.-F. Yu, O. Lourie, M.J. Dyer, K. Moloni, T.F. Kelley, R.S. Ruoff: Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load, Science **287**, 637–640 (2000)

8.197 D.A. Walters, L.M. Ericson, M.J. Casavant, J. Liu, D.T. Colbert, K.A. Smith, R.E. Smalley: Elastic strain of freely suspended single-wall carbon nanotube ropes, Appl. Phys. Lett. **74**, 3803–3805 (1999)

8.198 B.G. Demczyk, Y.M. Wang, J. Cumings, M. Hetman, W. Han, A. Zettl, R.O. Ritchie: Direct mechanical measurement of the tensile strength and elastic modulus of multiwalled carbon nanotubes, Mater. Sci. Eng. A **334**, 173–178 (2002)

8.199 R.P. Gao, Z.L. Wang, Z.G. Bai, W.A. De Heer, L.M. Dai, M. Gao: Nanomechanics of individual carbon nanotubes from pyrolytically grown arrays, Phys. Rev. Lett. **85**, 622–625 (2000)

8.200 M.M.J. Treacy, T.W. Ebbesen, J.M. Gibson: Exceptionally high Young's modulus observed for individual carbon nanotubes, Nature **381**, 678–680 (1996)

8.201 N. Yao, V. Lordie: Young's modulus of single-wall carbon nanotubes, J. Appl. Phys. **84**, 1939–1943 (1998)

8.202 O. Lourie, H.D. Wagner: Transmission electron microscopy observations of fracture of single-wall carbon nanotubes under axial tension, Appl. Phys. Lett. **73**, 3527–3529 (1998)

8.203 J. Zhang, M. Terrones, C.-R. Park, R. Mukherjee, M. Monthioux, N. Koratkar, Y.S. Kim, R. Hurt, E. Frackowiak, T. Enoki, Y. Chen, Y. Chen, A. Bianco: Carbon science in 2016: Status, challenges and perspectives, Carbon **98**, 708–732 (2016)

8.204 S.C. Tsang, Y.K. Chen, P.J.F. Harris, M.L.H. Green: A simple chemical method of opening and filling carbon nanotubes, Nature **372**, 159–162 (1994)

8.205 M. Monthioux, J.-C. Charlier: Giving credit where credit is due: the Stone-(Thrower)-Wales designation revisited, Carbon **75**, 1–4 (2014)

8.206 S. Agnihotri, J.P. Mota, M. Rostam-Abadi, M.J. Rood: Structural characterization of single-walled carbon nanotube bundles by experiment and molecular simulation, Langmuir **21**, 896–904 (2005)

8.207 M. Monthioux: Filling single-wall carbon nanotubes, Carbon **40**, 1809–1823 (2002)

8.208 M. Monthioux: Introduction to the meta-nanotube book. In: *Carbon Meta-Nanotubes: Synthesis, Properties, Applications*, ed. by M. Monthioux (Wiley-Blackwell, Chichester 2012) pp. 1–5

8.209 M. Monthioux (Ed.): *Carbon Meta-Nanotubes: Synthesis, Properties, Applications* (Wiley-Blackwell, Chichester 2012)

8.210 D. Golberg, M. Terrones: Heterogeneous Nanotubes (X*CNTs, X*BNNTs). In: *Carbon Meta-Nanotubes: Synthesis, Properties, Applications*, ed. by M. Monthioux (Wiley-Blackwell, Chichester 2012) pp. 323–409

8.211 V. Krstic, C.P. Ewels, T. Wågberg, M.S. Ferreira, A.M. Janssens, O. Stéphan, M. Glerup: Indirect magnetic coupling in light-element-doped single-walled carbon nanotubes, ACS Nano **9**, 5081–5086 (2010)

8.212 W.K. Hsu, S.Y. Chu, E. Munoz-Picone, J.L. Boldu, S. Firth, P. Franchi, B.P. Roberts, A. Shilder, H. Terrones, N. Grobert, Y.Q. Zhu, M. Terrones, M.E. McHenry, H.W. Kroto, D.R.M. Walton: Metallic behaviour of boron-containing carbon nanotubes, Chem. Phys. Lett. **323**, 572–579 (2000)

8.213 R. Czerw, M. Terrones, J.C. Charlier, X. Blasé, B. Foley, R. Kamalakaran, N. Grobert, H. Terrones, D. Tekleab, P.M. Ajayan, W. Blau, M. Rühle, D.L. Carroll: Identification of electron donor states, in N-doped carbon nanotubes, Nano Lett. **1**, 457–460 (2001)

8.214 O. Stephan, P.M. Ajayan, C. Colliex, P. Redlich, J.M. Lambert, P. Bernier, P. Lefin: Doping graphitic and carbon nanotube structures with boron and nitrogen, Science **266**, 1683–1685 (1994)

8.215 A. Loiseau, F. Willaime, N. Demoncy, N. Schramchenko, G. Hug, C. Colliex, H. Pascard: Boron nitride nanotubes, Carbon **36**, 743–752 (1998)

8.216 C.C. Tang, L.M. de la Chapelle, P. Li, Y.M. Liu, H.Y. Dang, S.S. Fan: Catalytic growth of nanotube and nanobamboo structures of boron nitride, Chem. Phys. Lett. **342**, 492–496 (2001)

8.217 K. Suenaga, C. Colliex, N. Demoncy, A. Loiseau, H. Pascard, F. Willaime: Synthesis of nanoparticles and nanotubes with well separated layers of boron-nitride and carbon, Science **278**, 653–655 (1997)

8.218 D. Golberg, Y. Bando, L. Bourgeois, K. Kurashima, T. Sato: Large-scale synthesis and HRTEM analysis of single-walled B- and N-doped carbon nanotube bundles, Carbon **38**, 2017–2027 (2000)

8.219 D.E. Gourari, M. Razafinimanana, M. Monthioux, R. Arenal, F. Valensi, S. Joulié, V. Serin: Synthesis of (B-C-N) nanomaterials by arc discharge using heterogeneous anodes, Plasma Sci. Technol. **18**, 465–468 (2016)

8.220 R.S. Lee, J. Gavillet, M.L. de la Chapelle, A. Loiseau, J.-L. Cochon, D. Pigache, J. Thibault, F. Willaime: Catalyst-free synthesis of boron nitride single-wall nanotubes with a preferred zig-zag configuration, Phys. Rev. B **64**, 121405.1–121405.4 (2001)

8.221 M. Monthioux, E. Flahaut, J.-P. Cleuziou: Hybrid carbon nanotubes: Strategy, progress, and perspectives, J. Mater. Res. **21**, 2774–2793 (2006)

8.222 B. Burteaux, A. Claye, B.W. Smith, M. Monthioux, D.E. Luzzi, J.E. Fischer: Abundance of encapsulated C_{60} in single-wall carbon nanotubes, Chem. Phys. Lett. **310**, 21–24 (1999)

8.223 N. Behabtu, C.C. Young, D.E. Tsentalovich, O. Kleinerman, X. Wang, A.W.K. Ma, E.A. Bengio, R.F. ter Waarbeek, J.J. de Jong, R.E. Hoogerwerf, S.B. Fairchild, J.B. Ferguson, B. Maruyama, J. Kono, Y. Talmon, Y. Cohen, M.J. Otto, M. Pasquali: Strong, light, multifunctional fibers of carbon nanotubes with ultrahigh conductivity, Science **339**, 182–186 (2013)

8.224 D. Ugarte, A. Châtelain, W.A. de Heer: Nanocapillarity and chemistry in carbon nanotubes, Science **274**, 1897–1899 (1996)

8.225 J. Cook, J. Sloan, M.L.H. Green: Opening and filling carbon nanotubes, Fuller. Sci. Technol. **5**, 695–704 (1997)

8.226 P.M. Ajayan, S. Iijima: Capillarity-induced filling of carbon nanotubes, Nature **361**, 333–334 (1993)

8.227 P.M. Ajayan, T.W. Ebbesen, T. Ichihashi, S. Iijima, K. Tanigaki, H. Hiura: Opening carbon nanotubes with oxygen and implications for filling, Nature **362**, 522–525 (1993)

8.228 S. Seraphin, D. Zhou, J. Jiao, J.C. Withers, R. Loufty: Yttrium carbide in nanotubes, Nature **362**, 503 (1993)

8.229 S. Seraphin, D. Zhou, J. Jiao, J.C. Withers, R. Loufty: Selective encapsulation of the carbides of yttrium and titanium into carbon nanoclusters, Appl. Phys. Lett. **63**, 2073–2075 (1993)

8.230 R.S. Ruoff, D.C. Lorents, B. Chan, R. Malhotra, S. Subramoney: Single-crystal metals encapsulated in carbon nanoparticles, Science **259**, 346–348 (1993)

8.231 A. Loiseau, H. Pascard: Synthesis of long carbon nanotubes filled with Se, S, Sb, and Ge by the arc method, Chem. Phys. Lett. **256**, 246–252 (1996)

8.232 N. Demoncy, O. Stephan, N. Brun, C. Colliex, A. Loiseau, H. Pascard: Filling carbon nanotubes with metals by the arc discharge method: The key role of sulfur, Eur. Phys. J. B **4**, 147–157 (1998)

8.233 C.H. Kiang, J.S. Choi, T.T. Tran, A.D. Bacher: Molecular nanowires of 1 nm diameter from capillary filling of single-walled carbon nanotubes, J. Phys. Chem. B **103**, 7449–7551 (1999)

8.234 Z.L. Zhang, B. Li, Z.J. Shi, Z.N. Gu, Z.Q. Xue, L.M. Peng: Filling of single-walled carbon nanotubes with silver, J. Mater. Res. **15**, 2658–2661 (2000)

8.235 A. Govindaraj, B.C. Satishkumar, M. Nath, C.N.R. Tao: Metal nanowires and intercalated metal layers in single-walled carbon nanotubes bundles, Chem. Mater. **12**, 202–205 (2000)

8.236 J. Mittal, M. Monthioux, H. Allouche: Room temperature filling of single-wall carbon nanotubes with chromium oxide in open air, Chem. Phys. Lett. **339**, 311–318 (2001)

8.237 E. Dujardin, T.W. Ebbesen, H. Hiura, K. Tanigaki: Capillarity and wetting of carbon nanotubes, Science **265**, 1850–1852 (1994)

8.238 E. Flahaut, J. Sloan, K.S. Coleman, V.C. Williams, S. Friedrichs, N. Hanson, M.L.H. Green: 1-D p-block halide crystals confined into single walled carbon nanotubes, Proc. Mater. Res. Soc. Symp. **633**, A13.15.1–A13.15.6 (2001)

8.239 J. Sloan, A.I. Kirkland, J.L. Hutchison, M.L.H. Green: Integral atomic layer architectures of 1-D crystals inserted into single walled carbon nanotubes, Chem. Commun. (2002) doi:10.1039/B200537A

8.240 J. Sloan, M.C. Novotny, S.R. Bailey, G. Brown, C. Xu, V.C. Williams, S. Friedrichs, E. Flahaut, R.L. Callender, A.P.E. York, K.S. Coleman, M.L.H. Green, R.E. Dunin-Borkowski, J.L. Hutchison: Two layer 4:4 co-ordinated KI crystals grown within single walled carbon nanotubes, Chem. Phys. Lett. **329**, 61–65 (2000)

8.241 G. Brown, S.R. Bailey, J. Sloan, C. Xu, S. Friedrichs, E. Flahaut, K.S. Coleman, J.L. Hutchinson, R.E. Dunin-Borkowski, M.L.H. Green: Electron beam induced in situ clusterisation of 1-D $ZrCl_4$ chains within single-walled carbon nanotubes, Chem. Commun. (2001) doi:10.1039/B1012610

8.242 J. Sloan, D.M. Wright, H.G. Woo, S. Bailey, G. Brown, A.P.E. York, K.S. Coleman, J.L. Hutchison, M.L.H. Green: Capillarity and silver nanowire formation observed in single walled carbon nanotubes, Chem. Commun. (1999) doi:10.1039/A901572H

8.243 X. Fan, E.C. Dickey, P.C. Eklund, K.A. Williams, L. Grigorian, R. Buczko, S.T. Pantelides, S.J. Pennycook: Atomic arrangement of iodine atoms inside single-walled carbon nanotubes, Phys. Rev. Lett. **84**, 4621–4624 (2000)

8.244 G. Brown, S.R. Bailey, M. Novotny, R. Carter, E. Flahaut, K.S. Coleman, J.L. Hutchison, M.L.H. Green, J. Sloan: High yield incorporation and washing properties of halides incorporated into single walled carbon nanotubes, Appl. Phys. A **76**, 457–462 (2003)

8.245 J. Sloan, D.E. Luzzi, A.I. Kirkland, J.L. Hutchison, M.L.H. Green: Imaging and characterization of molecules and one-dimensional crystals formed within carbon nanotubes, Mater. Res. Soc. Bull. **29**, 265–271 (2004)

8.246 B.W. Smith, M. Monthioux, D.E. Luzzi: Encapsulated C_{60} in carbon nanotubes, Nature **396**, 323–324 (1998)

8.247 B.W. Smith, D.E. Luzzi, Y. Achiba: Tumbling atoms and evidence for charge transfer in $La_2@C_{80}$@SWNT, Chem. Phys. Lett. **331**, 137–142 (2000)

8.248 K. Suenaga, M. Tence, C. Mory, C. Colliex, H. Kato, T. Okazaki, H. Shinohara, K. Hirahara, S. Bandow, S. Iijima: Element-selective single atom imaging, Science **290**, 2280–2282 (2000)

8.249 D.E. Luzzi, B.W. Smith, R. Russo, B.C. Satishkumar, F. Stercel, N.R.C. Nemes: Encapsulation of metallofullerenes and metallocenes in carbon nanotubes. In: *Proc. Electron. Prop. Nov. Mater. – XVI Int. Winterseh. – AIP Conf.*, ed. by H. Kuzmany, J. Fink, M. Mehring, S. Roth (Springer, Berlin, Heidelberg 2001) pp. 622–626

8.250 J. Chancolon, F. Archaimbault, A. Pineau, S. Bonnamy: Confinement of selenium into carbon nanotubes, Fuller. Nanotub. Carbon Nanostruct. **13**, 189–194 (2005)

8.251 L. Guan, K. Suenaga, Z. Shi, Z. Gu, S. Iijima: Polymorphic structures of iodine and their phase transition in confined nanospace, Nano Lett. **7**, 1532–1535 (2007)

8.252 B.W. Smith, D.E. Luzzi: Formation mechanism of fullerene peapods and coaxial tubes: A path to large scale synthesis, Chem. Phys. Lett. **321**, 169–174 (2000)

8.253 K. Hirahara, K. Suenaga, S. Bandow, H. Kato, T. Okazaki, H. Shinohara, S. Iijima: One-dimensional metallo-fullerene crystal generated inside single-walled carbon nanotubes, Phys. Rev. Lett. **85**, 5384–5387 (2000)

8.254 Y.P. Sun, K. Fu, Y. Lin, W. Huang: Functionalized carbon nanotubes: Properties and applications, Acc. Chem. Res. **35**, 1095–1104 (2002)

8.255 S. Campidelli, S.S. Wong, M. Prato: Functionalized Carbon nanotubes (X-CNTs). In: *Carbon Meta-Nanotubes: Synthesis, Properties, Applications*, ed. by M. Monthioux (Wiley-Blackwell, Chichester 2012) pp. 113–161

8.256 S. Osswald, E. Flahaut, H. Ye, Y. Gogotsi: Elimination of D-band in Raman spectra of double-wall carbon nanotubes by oxidation, Chem. Phys. Lett. **402**, 422–427 (2005)

8.257 W. Xia, C. Jin, S. Kundu, M. Muhler: A highly efficient gas-phase route for the oxygen functionalization of carbon nanotubes based on nitric acid vapor, Carbon **47**, 919–922 (2009)

8.258 D.C. Vennerberg, R.L. Quirino, Y. Chang, M.R. Kessler: Oxidation behavior of multiwalled carbon nanotubes fluidized with ozone, Appl. Mat. Interfaces **6**, 1835–1842 (2014)

8.259 P. Lassègue, N. Coppey, L. Noé, M. Monthioux, B. Caussat: Decoration of carbon nanotubes by semiconducting or metallic nanoparticles using fluidized bed chemical vapor deposition, Kona Powder Part. J. **33**, 322–332 (2016)

8.260 J. Chen, M.A. Hamon, M. Hui, C. Yongsheng, A.M. Rao, P.C. Eklund, R.C. Haddon: Solution properties of single-walled carbon nanotubes, Science **282**, 95–98 (1998)

8.261 D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato: Chemistry of Carbon Nanotubes, Chem. Rev. **106**, 1105–1136 (2006)

8.262 J. Chen, A.M. Rao, S. Lyuksyutov, M.E. Itkis, M.A. Hamon, H. Hu, R.W. Cohn, P.C. Eklund, D.T. Colbert, R.E. Smalley, R.C. Haddon: Dissolution of full-length single-walled carbon nanotubes, J. Phys. Chem. B **105**, 2525–2528 (2001)

8.263 K. Nagaraju, R. Reddy, N. Reddy: A review on protein functionalized carbon nanotubes, J. Appl. Biomater. Funct. Mater. **13**, e301–e312 (2015)

8.264 S. Daniel, T.P. Rao, K.S. Rao, S.U. Rani, G.R.K. Naidu, H.-Y. Lee, T. Kawai: A review of DNA functionalized/grafted carbon nanotubes and their characterization, Sens. Actuat. B **122**, 672–682 (2007)

8.265 Y.P. Sun, W. Huang, Y. Lin, K. Fu, A. Kitaygorodskiy, L.A. Riddle, Y.J. Yu, D.L. Carroll: Soluble dendron-functionalized carbon nanotubes: Preparation, characterization, and properties, Chem. Mater. **13**, 2864–2869 (2001)

8.266 K. Fu, W. Huang, Y. Lin, L.A. Riddle, D.L. Carroll, Y.P. Sun: Defunctionalization of functionalized carbon nanotubes, Nano Lett. **1**, 439–441 (2001)

8.267 P.W. Chiu, G.S. Duesberg, U. Dettlaff-Weglikowska, S. Roth: Interconnection of carbon nanotubes by chemical functionalization, Appl. Phys. Lett. **80**, 3811–3813 (2002)

8.268 M. Tunckol, J. Durand, P. Serp: Carbon nanomaterial–ionic liquid hybrids, Carbon **50**, 4303–4334 (2012)

8.269 A. Di Crescenzo, V. Ettorre, A. Fontana: Non-covalent and reversible functionalization of carbon nanotubes, Beilstein J. Nanotechnol. **5**, 1675–1690 (2014)

8.270 E.T. Mickelson, C.B. Huffman, A.G. Rinzler, R.E. Smalley, R.H. Hauge, J.L. Margrave: Fluorination of single-wall carbon nanotubes, Chem. Phys. Lett. **296**, 188–194 (1998)

8.271 V.N. Khabashesku, W.E. Billups, J.L. Margrave: Fluorination of single-wall carbon nanotubes and subsequent derivatization reactions, Acc. Chem. Res. **35**, 1087–1095 (2002)

8.272 P.-C. Ma, N.A. Siddiqui, G. Marom, J.-K. Kim: Dispersion and functionalization of carbon nanotubes for polymer-based nanocomposites: A review, Composites: Part A **41**, 1345–1367 (2010)

8.273 L. Meng, C. Fu, Q. Lu: Advanced technology for functionalization of carbon nanotubes, Progr. Nat. Sci. **19**, 801–810 (2009)

8.274 J. Han, C. Gao: Functionalization of carbon nanotubes and other nanocarbons by azide chemistry, Nano-Micro Lett. **2**, 213–226 (2010)

8.275 R.R. Bacsa, P. Serp: Decorated (coated) nanotubes. In: *Carbon Meta-Nanotubes: Synthesis, Properties, Applications*, ed. by M. Monthioux (Wiley-Blackwell, Chichester 2012) pp. 163–221

8.276 K.N. Chaudhari, S. Chaudhari, J.S. Yu: Synthesis and supercapacitor performance of Au-nanoparticle decorated MWCNT, J. Electroanal. Chem. **761**, 98–105 (2016)

8.277 G.G. Wildgoose, C.E. Banks, R.G. Compton: Metal nanoparticles and related materials supported on carbon nanotubes: Methods and applications, Small **2**, 182–193 (2006)

8.278 Q. Kuang, S.F. Li, Z.X. Xie, S.C. Lin, X.H. Zhang, S.Y. Xie, R.B. Huang, L.S. Zheng: Controllable fabrication of SnO_2-coated multiwalled carbon nanotubes by chemical vapour deposition, Carbon **44**, 1166–1172 (2006)

8.279 A. Goulas, J.R. van Ommen: Scalable production of nanostructured particles using atomic layer deposition, Kona Powder Part. J. **31**, 234–246 (2014)

8.280 S. Banerjee, K. Dasgupta, A. Kumar, P. Ruz, B. Vishwanadh, J.B. Joshi, V. Sudarsan: Comparative evaluation of hydrogen storage behavior of Pd doped carbon nanotubes prepared by wet impregnation and polyol methods, Int. J. Hydrogen En. **40**(8), 3268–3276 (2015)

8.281 K. Fu, O. Yildiz, H. Bhanushali, Y. Wang, K. Stano, L. Xue, X. Zhang, P.D. Bradford: Aligned carbon nanotube–Silicon sheets: A novel nano-architecture for flexible lithium ion battery electrodes, Adv. Mater. **25**, 5109–5114 (2013)

8.282 F. Mendoza, D.M. Hernandez, V. Makarov, E. Febus, B.R. Weiner, G. Morell: Room temperature gas sensor based on tin dioxide–carbon nanotubes composite films, Sensors Actuat. B **190**, 227–233 (2014)

8.283 M. Feng, R.J. Puddephatt: Chemical vapour deposition of nickel-group metals on multiwall carbon nanotubes, Can. J. Chem. **85**, 645–650 (2007)

8.284 N. Coppey, L. Noé, M. Monthioux, B. Caussat: Decorated carbon nanotubes by silicon deposition in fluidized bed for li-ion battery anodes, Chem. Eng. Res. Design **91**, 2491–2496 (2013)

8.285 S.R. Bakshi, D. Lahiri, A. Agarwal: Carbon nanotube reinforced metal matrix composites – A review, Int. Mater. Rev. **55**, 41–64 (2010)

8.286 N. Saheb, Z. Iqbal, A. Khalil, A.S. Hakeem, N. Al Aqeeli, T. Laoui, A. Al-Qutub, R. Kirchner: Spark plasma sintering of metals and metal matrix nanocomposites: A review, J. Nanomater. **2012**, 983470 (2012)

8.287 S.C. Tjong: Recent progress in the development and properties of novel metal matrix nanocomposites reinforced with carbon nanotubes and graphene nanosheets, Mater. Sci. Eng. R-Rep. **74**, 281–350 (2013)

8.288 C.L. Xu, B.Q. Wei, R.Z. Ma, J. Liang, X.K. Ma, D.H. Wu: Fabrication of aluminum-carbon nanotube composites and their electrical properties, Carbon **37**, 855–858 (1999)

8.289 T. Kuzumaki, K. Miyazawa, H. Ichinose, K. Ito: Processing of carbon nanotube reinforced aluminum composite, J. Mater. Res. **13**, 2445–2449 (1998)

8.290 H. Kwon, D.H. Park, J.F. Silvain, A. Kawasaki: Investigation of carbon nanotube reinforced aluminum matrix composite materials, Compos. Sci. Technol. **70**, 546–550 (2010)

8.291 H.J. Choi, J.Y. Shin, B.H. Min, J.S. Park, D.H. Bae: Reinforcing effects of carbon nanotubes in structural aluminum matrix nanocomposites, J. Mater. Res. **24**, 2610–2616 (2009)

8.292 C.F. Deng, Y.X. Ma, P. Zhang, X.X. Zhang, D.Z. Wang: Thermal expansion behaviors of aluminum composite reinforced with carbon nanotubes, Mater. Lett. **62**, 2301–2303 (2008)

8.293 K.T. Kim, S.I. Cha, S.H. Hong, S.H. Hong: Microstructures and tensile behavior of carbon nanotube reinforced Cu matrix nanocomposites, Mater. Sci. Eng. A **430**, 27–33 (2006)

8.294 W.M. Daoush, B.K. Lim, C.B. Mo, D.H. Nam, S.H. Hong: Electrical and mechanical properties of carbon nanotube reinforced copper nanocomposites fabricated by electroless deposition process, Mater. Sci. Eng. A **513–514**, 247–253 (2009)

8.295 C. Arnaud, F. Lecouturier, D. Mesguich, N. Ferreira, G. Chevallier, C. Estournès, A. Weibel, C. Laurent: High strength-high conductivity double-walled carbon nanotube – Copper composite wires, Carbon **96**, 212–215 (2016)

8.296 C. Guiderdoni, E. Pavlenko, V. Turq, A. Weibel, P. Puech, C. Estournès, A. Peigney, W. Bacsa, C. Laurent: The preparation of carbon nanotube (CNT)/copper composites and the effect of the number of CNT walls on their hardness, friction and wear properties, Carbon **58**, 185–197 (2013)

8.297 C.S. Goh, J. Wei, L.C. Lee, M. Gupta: Simultaneous enhancement in strength and ductility reinforcing magnesium with carbon nanotubes, Mater. Sci. Eng. A **423**, 153–156 (2006)

8.298 F.J. Sun, C.S. Shi, K.Y. Rhee, N.Q. Zhao: In situ synthesis of CNTs in Mg powder at low temperature for fabricating reinforced Mg composites, J. Alloys Compounds **551**, 496–501 (2013)

8.299 T. Kuzumaki, O. Ujiie, H. Ichinose, K. Ito: Mechanical characteristics and preparation of carbon nanotube fiber-reinforced Ti composite, Adv. Eng. Mater. **2**, 416–418 (2000)

8.300 K. Kondoh, T. Threrujirapapong, H. Imai, J. Umeda, B. Fugetsu: Characteristics of powder metallurgy pure titanium matrix composite reinforced with multi-wall carbon nanotubes, Compos. Sci. Technol. **69**, 1077–1081 (2009)

8.301 E. Carreno-Morelli, J. Yang, E. Couteau, K. Hernadi, J.W. Seo, C. Bonjour, L. Forro, R. Schaller: Carbon nanotube/magnesium composites, Phys. Status Solidi (a) **201**, R53–R55 (2004)

8.302 Q. Ngo, B.A. Cruden, A.M. Cassell, M.D. Walker, Q. Ye, J.E. Koehne, M. Meyyappan, J. Li, C.Y. Yang: Thermal conductivity of carbon nanotube composite films, Mater. Res. Soc. Symp. Proc. **812**, 179–184 (2004)

8.303 X.H. Chen, C.S. Chen, H.N. Xiao, F.Q. Cheng, G. Zhang, G.J. Yi: Corrosion behavior of carbon nanotubes-Ni composite coating, Surf. Coat. Technol. **191**, 351–356 (2005)

8.304 X.H. Chen, C.S. Chen, H.N. Xiao, H.B. Liu, L.P. Zhou, S.L. Li, G. Zhang: Dry friction and wear characteristics of nickel/carbon nanotube electroless composite deposits, Tribol. Int. **39**, 22–28 (2006)

8.305 A. Peigney, A. Weibel, C. Laurent: Carbon nanotubes in ceramic-matrix nanocomposites. In: *Encyclopedia of Nanoscience and Nanotechnology*, ed. by H.S. Nalwa (American Scientific, Valencia 2011) pp. 179–196

8.306 J. Cho, A.R. Boccaccini, M.S.P. Shaffer: Ceramic matrix composites containing carbon nanotubes, J. Mater. Sci. **44**, 1934–1951 (2009)

8.307 E. Zapata-Solvas, D. Gómez-García, A. Domínguez-Rodríguez: Towards physical properties tailoring of carbon nanotubes-reinforced ceramic matrix composites, J. Eur. Ceram. Soc. **32**, 3001–3020 (2012)

8.308 S.I. Cha, K.T. Kim, K.H. Lee, C.B. Mo, S.H. Hong: Strengthening and toughening of carbon nanotube reinforced alumina nanocomposite fabricated by molecular level mixing process, Scr. Mater. **53**, 793–797 (2005)

8.309 M. Estili, A. Kawasaki, Y. Sakka: Highly concentrated 3-D macrostructure of individual carbon nanotubes in a ceramic environment, Adv. Mater. **24**, 4322–4326 (2012)

8.310 X. Wang, N.P. Padture, H. Tanaka: Contact-damage-resistant ceramic/single-wall carbon nanotubes and ceramic/graphite composites, Nat. Mater. **3**, 539–544 (2004)

8.311 W.A. Curtin, B.W. Sheldon: CNT-reinforced ceramics and metals, Mater. Today **7**, 44–49 (2004)

8.312 Z. Xia, L. Riester, W.A. Curtin, H. Li, B.W. Sheldon, J. Liang, B. Chang, J.M. Xu: Direct observation of toughening mechanisms in carbon nanotube ceramic matrix composites, Acta Mater. **52**, 931–944 (2004)

8.313 M. Estili, Y. Sakka: Recent advances in understanding the reinforcing ability and mechanism of carbon nanotubes in ceramic matrix composites, Sci. Technol. Adv. Mater. **15**, 064902 (2014)

8.314 A. Peigney, F. Legorreta Garcia, C. Estournès, A. Weibel, C. Laurent: Toughening and hardening in double-walled carbon nanotube/nanostructured magnesia composites, Carbon **48**, 1952–1960 (2010)

8.315 D.S. Lim, J.W. An, H.J. Lee: Effect of carbon nanotube addition on the tribological behavior of carbon/carbon composites, Wear **252**, 512–517 (2002)

8.316 D.-S. Lim, D.-H. You, H.-J. Choi, S.-H. Lim, H. Jang: Effect of CNT distribution on tribological behavior of alumina-CNT composites, Wear **259**, 539–544 (2005)

8.317 Z.H. Xia, J. Lou, W.A. Curtin: A multiscale experiment on the tribological of aligned carbon nanotube/ceramic composites, Scr. Mater. **58**, 223–226 (2008)

8.318 A. Kasperski, A. Weibel, D. Alkattan, C. Estournès, V. Turq, C. Laurent, A. Peigney: Microhardness and friction coefficient of multi-walled carbon nanotube-yttria-stabilized ZrO_2 composites prepared by spark plasma sintering, Scr. Mater. **69**, 338–341 (2013)

8.319 G.-D. Zhan, J.D. Kuntz, H. Wang, C.-M. Wang, A.K. Mukherjee: Anisotropic thermal properties of single-wall-carbon-nanotube-reinforced nanoceramics, Philos. Mag. Lett. **84**, 419–423 (2004)

8.320 Q. Huang, L. Gao, Y. Liu, J. Sun: Sintering and thermal properties of multiwalled carbon nanotube-$BaTiO_3$ composites, J. Mater. Chem. **15**, 1995–2001 (2005)

8.321 G.-D. Zhan, J.D. Kuntz, J.E. Garay, A.K. Mukherjee: Electrical properties of nanoceramics reinforced with ropes of single-walled carbon nanotubes, Appl. Phys. Lett. **83**, 1228–1230 (2003)

8.322 S. Rul, F. Lefevre-Schlick, E. Capria, C. Laurent, A. Peigney: Percolation of single-walled carbon nanotubes in ceramic matrix nanocomposites, Acta Mater. **52**, 1061–1067 (2004)

8.323 S.-L. Shi, J. Liang: Electronic transport properties of multiwall carbon nanotubes/yttria-stabilized zirconia composites, J. Appl. Phys. **101**, 023708–5 (2007)

8.324 M.J. de Andrade, A. Weibel, C. Laurent, S. Roth, C.P. Bergmann, C. Estournès, A. Peigney: Electrical conductive double-walled carbon nanotubes-silica glass nanocomposites prepared by the sol-gel process and spark plasma sintering, Scr. Mater. **61**, 988–991 (2009)

8.325 A. Peigney, E. Flahaut, C. Laurent, F. Chastel, A. Rousset: Aligned carbon nanotubes in ceramic-matrix nanocomposites prepared by high-temperature extrusion, Chem. Phys. Lett. **352**, 20–25 (2002)

8.326 P.M. Ajayan, O. Stephan, C. Colliex, D. Trauth: Aligned carbon nanotube arrays formed by cutting a polymer resin-nanotube composite, Science **265**, 1212–1214 (1994)

8.327 Z. Spitalskya, D. Tasisb, K. Papagelisb, C. Galiotis: Carbon nanotube–polymer composites: Chemistry, processing, mechanical and electrical properties, Prog. Polym. Sci. **35**, 357–401 (2010)

8.328 R. Haggenmueller, H.H. Gommans, A.G. Rinzler, J.E. Fischer, K.I. Winey: Aligned single-wall carbon nanotubes in composites by melt processing methods, Chem. Phys. Lett. **330**, 219–225 (2000)

8.329 L.S. Schadler, S.C. Giannaris, P.M. Ajayan: Load transfer in carbon nanotube epoxy composites, Appl. Phys. Lett. **73**, 3842–3844 (1998)

8.330 S.J.V. Frankland, A. Caglar, D.W. Brenner, M. Griebel: Molecular simulation of the influence of chemical cross-links on the shear strength of carbon nanotube-polymer interfaces, J. Phys. Chem. B **106**, 3046–3048 (2002)

8.331 H.D. Wagner: Nanotube-polymer adhesion: A mechanics approach, Chem. Phys. Lett. **361**, 57–61 (2002)

8.332 P.M. Ajayan, L.S. Schadler, C. Giannaris, A. Rubio: Single-walled carbon nanotube-polymer composites: Strength and weakness, Adv. Mater. **12**, 750–753 (2000)

8.333 X. Gong, J. Liu, S. Baskaran, R.D. Voise, J.S. Young: Surfactant-assisted processing of carbon nanotube/polymer composites, Chem. Mater. **12**, 1049–1052 (2000)

8.334 E.T. Thostenson, W.Z. Li, D.Z. Wang, Z.F. Ren, T.W. Chou: Carbon nanotube/carbon fiber hybrid multiscale composites, J. Appl. Phys. **91**, 6034–6037 (2002)

8.335 F.H. Gojny, M.H.G. Wichmann, U. Kopke, B. Fiedler, K. Schulte: Carbon nanotube-reinforced epoxy-composites: enhanced stiffness and fracture toughness at low nanotube content, Compos. Sci. Technol. **64**, 2363–2371 (2004)

8.336 F.H. Gojny, M.H.G. Wichmann, B. Fiedler, K. Schulte: Influence of different carbon nanotubes on the mechanical properties of epoxy matrix composites – A comparative study, Compos. Sci. Tech. **65**, 2300–2313 (2005)

8.337 H. Rajoria, N. Jalili: Passive vibration damping enhancement using carbon nanotube-epoxy reinforced composites, Compos. Sci. Tech. **65**, 2079–

2093 (2005)
8.338 A.B. Dalton, S. Collins, E. Munoz, J.M. Razal, V.H. Ebron, J.P. Ferraris, J.N. Coleman, B.G. Kim, R.H. Baughman: Super-tough carbon-nanotube fibers, Nature **23**, 703 (2003)
8.339 P. Miaudet, S. Badaire, M. Maugey, A. Derre, V. Pichot, P. Launois, P. Poulin, C. Zakri: Hot-drawing of single and multiwall carbon nanotube fibers for high toughness and alignment, Nano Lett. **5**, 2212–2215 (2005)
8.340 B. Vigolo, A. Pénicaud, C. Coulon, C. Sauder, R. Pailler, C. Journet, P. Bernier, P. Poulin: Macroscopic fibers and ribbons of oriented carbon nanotubes, Science **290**, 1331–1334 (2000)
8.341 B. Vigolo, P. Poulin, M. Lucas, P. Launois, P. Bernier: Improved structure and properties of single-wall carbon nanotube spun fibers, Appl. Phys. Lett. **11**, 1210–1212 (2002)
8.342 G. Mittal, V. Dhand, K.Y. Rhee, S.-J. Park, W.R. Lee: A review on carbon nanotubes and graphene as fillers in reinforced polymer nanocomposites, J. Ind. Eng. Chem. **21**, 11–25 (2015)
8.343 J.K.W. Sandler, J.E. Kirk, I.A. Kinloch, M.S.P. Shaffer, A.H. Winlde: Ultra-low electrical percolation threshold in carbon-nanotube-epoxy composites, Polymer **44**, 5893–5899 (2003)
8.344 Z. Hana, A. Fina: Thermal conductivity of carbon nanotubes and their polymer nanocomposites: A review, Prog. Polym. Sci. **36**, 914–944 (2011)
8.345 M.S.P. Shaffer, A.H. Windle: Fabrication and characterization of carbon nanotube/poly(vinyl alcohol) composites, Adv. Mater. **11**, 937–941 (1999)
8.346 L. Jin, C. Bower, O. Zhou: Alignment of carbon nanotubes in a polymer matrix by mechanical stretching, Appl. Phys. Lett. **73**, 1197–1199 (1998)
8.347 H.D. Wagner, O. Lourie, Y. Feldman, R. Tenne: Stress-induced fragmentation of multiwall carbon nanotubes in a polymer matrix, Appl. Phys. Lett. **72**, 188–190 (1998)
8.348 H.D. Wagner, O. Lourie, X.F. Zhou: Macrofragmentation and microfragmentation phenomena in composite materials, Compos. Part A **30**, 59–66 (1998)
8.349 J.R. Wood, Q. Zhao, H.D. Wagner: Orientation of carbon nanotubes in polymers and its detection by Raman spectroscopy, Compos. Part A **32**, 391–399 (2001)
8.350 Q. Zhao, J.R. Wood, H.D. Wagner: Using carbon nanotubes to detect polymer transitions, J. Polym. Sci. B **39**, 1492–1495 (2001)
8.351 M. Cochet, W.K. Maser, A.M. Benito, M.A. Callejas, M.T. Martinesz, J.M. Benoit, J. Schreiber, O. Chauvet: Synthesis of a new polyaniline/nanotube composite: In-situ polymerisation and charge transfer through site-selective interaction, Chem. Commun. (2001) doi:10.1039/B104009J
8.352 D. Qian, E.C. Dickey, R. Andrews, T. Rantell: Load transfer and deformation mechanisms in carbon nanotube-polystyrene composites, Appl. Phys. Lett. **76**, 2868–2870 (2000)
8.353 V. Datsyuk, C. Guerret-Piecourt, S. Dagreou, L. Billon, J.-C. Dupin, E. Flahaut, A. Peigney, C. Laurent: Double walled carbon nanotube/polymer composites via in-situ nitroxide mediated polymerisation of amphiphilic block copolymers, Carbon **43**, 873–876 (2005)
8.354 R. Blake, Y.K. Gun'ko, J. Coleman, M. Cadek, A. Fonseca, J.B. Nagy, W.J. Blau: A generic organometallic approach toward ultra-strong carbon nanotube polymer composites, J. Am. Chem. Soc. **126**, 10226–10227 (2004)
8.355 T. Kashiwagi, E. Grulke, J. Hilding, K. Groth, R. Harris, K. Butler, J. Shields, S. Kharchenko, J. Douglas: Thermal and flammability properties of polypropylene/carbon nanotube nanocomposites, Polymer **45**, 4227–4239 (2004)
8.356 C. Wei, D. Srivastava, K. Cho: Thermal expansion and diffusion coefficients of carbon nanotube-polymer composites, preprint arXiv:cond-mat/0203349, 1–11 (2002)
8.357 J.C. Grunlan, M.V. Bannon, A.R. Mehrabi: Latex-based, single-walled nanotube composites: Processing and electrical conductivity, Polym. Prepr. **45**, 154–155 (2004)
8.358 J.C. Grunlan, A.R. Mehrabi, M.V. Bannon, J.L. Bahr: Water-based single-walled-nanotube-filled polymer composite with an exceptionally low percolation threshold, Adv. Mater. **16**, 150–153 (2004)
8.359 C. Pirlot, I. Willems, A. Fonseca, J.B. Nagy, J. Delhalle: Preparation and characterization of carbon nanotube/polyacrylonitrile composites, Adv. Eng. Mater. **4**, 109–114 (2002)
8.360 H. Lam, H. Ye, Y. Gogotsi, F. Ko: Structure and properties of electrospun single-walled carbon nanotubes reinforced nanocomposite fibrils by co-electrospinning, Polym. Prepr. **45**, 124–125 (2004)
8.361 L. Cao, H. Chen, M. Wang, J. Sun, X. Zhang, F. Kong: Photoconductivity study of modified carbon nanotube/oxotitanium phthalocyanine composites, J. Phys. Chem. B **106**, 8971–8975 (2002)
8.362 I. Musa, M. Baxendale, G.A.J. Amaratunga, W. Eccleston: Properties of regular poly(3-octylthiophene)/multi-wall carbon nanotube composites, Synth. Met. **102**, 1250 (1999)
8.363 E. Kymakis, I. Alexandou, G.A.J. Amaratunga: Single-walled carbon nanotube-polymer composites: Electrical, optical and structural investigation, Synth. Met. **127**, 59–62 (2002)
8.364 K. Yoshino, H. Kajii, H. Araki, T. Sonoda, H. Take, S. Lee: Electrical and optical properties of conducting polymer-fullerene and conducting polymer-carbon nanotube composites, Fuller. Sci. Technol. **7**, 695–711 (1999)
8.365 S.A. Curran, P.M. Ajayan, W.J. Blau, D.L. Carroll, J.N. Coleman, A.B. Dalton, A.P. Davey, A. Drury, B. McCarthy, S. Maier, A. Strevens: A composite from poly(*m*-phenylenevinylene-co-2,5-dioctoxy-*p*-phenylenevinylene) and carbon nanotubes. A novel material for molecular optoelectronics, Adv. Mater. **10**, 1091–1093 (1998)
8.366 P. Fournet, D.F. O'Brien, J.N. Coleman, H.H. Horhold, W.J. Blau: A carbon nanotube composite as an electron transport layer for M3EH-PPV based light-emitting diodes, Synth. Met. **121**,

1683–1684 (2001)
8.367 H.S. Woo, R. Czerw, S. Webster, D.L. Carroll, J. Ballato, A.E. Strevens, D. O'Brien, W.J. Blau: Hole blocking in carbon nanotube-polymer composite organic light-emitting diodes based on poly(*m*-phenylene-vinylene-co-2,5-dioctoxy-*p*-phenylene vinylene), Appl. Phys. Lett. **77**, 1393–1395 (2000)
8.368 H.S. Woo, R. Czerw, S. Webster, D.L. Carroll, J.W. Park, J.H. Lee: Organic light emitting diodes fabricated with single wall carbon nanotubes dispersed in a hole conducting buffer: The role of carbon nanotubes in a hole conducting polymer, Synth. Met. **116**, 369–372 (2001)
8.369 H. Ago, K. Petritsch, M.S.P. Shaffer, A.H. Windle, R.H. Friend: Composites of carbon nanotubes and conjugated polymers for photovoltaic devices, Adv. Mater. **11**, 1281–1285 (1999)
8.370 P. Poulin, B. Vigolo, P. Launois: Films and fibers of oriented single wall nanotubes, Carbon **40**, 1741–1749 (2002)
8.371 K. Jiang, Q. Li, S. Fan: Spinning continuous carbon nanotube yarn, Nature **419**, 801 (2002)
8.372 M. Zhang, K.R. Atkinson, R.H. Baughman: Multifunctional carbon nanotube yarns by downsizing an ancient technology, Science **306**, 1356–1361 (2004)
8.373 J. Steinmetz, M. Glerup, M. Paillet, P. Bernier, M. Holzinger: Production of pure nanotube fibers using a modified wet-spinning method, Carbon **43**, 2397–2400 (2005)
8.374 B. Maruyama, K. Alam: Carbon nanotubes and nanofibers in composite materials, SAMPE J. **38**, 59–70 (2002)
8.375 M. Monthioux: Applications of carbon nanotubes. In: *Strained Hydrocarbons*, ed. by H. Dodziuk (Wiley, Weinheim 2009) pp. 356–373
8.376 M.F.L. De Volder, S.H. Tawfick, R.H. Baughman, A.J. Hart: Carbon nanotubes: Present and future commercial applications, Science **339**, 535–539 (2013)
8.377 H.J. Dai, J.H. Hafner, A.G. Rinzler, D.T. Colbert, R.E. Smalley: Nanotubes as nanoprobes in scanning probe microscopy, Nature **384**, 147–150 (1996)
8.378 J.H. Hafner, C.L. Cheung, A.T. Wooley, C.M. Lieber: Structural and functional imaging with carbon nanotube AFM probes, Progr. Biophys. Mol. Biol. **77**, 73–110 (2001)
8.379 S.S. Wong, E. Joselevich, A.T. Woodley, C.L. Cheung, C.M. Lieber: Covalently functionalized nanotubes as nanometre-size probes in chemistry and biology, Nature **394**, 52–55 (1998)
8.380 C.L. Cheung, J.H. Hafner, C.M. Lieber: Carbon nanotube atomic force microscopy tips: Direct growth by chemical vapor deposition and application to high-resolution imaging, Proc. Natl. Acad. Sci. USA **97**, 3809–3813 (2000)
8.381 W.A. de Heer, A. Châtelain, D. Ugarte: A carbon nanotube field-emission electron source, Science **270**, 1179–1180 (1995)
8.382 J.M. Bonard, J.P. Salvetat, T. Stockli, W.A. de Heer, L. Forro, A. Chatelâin: Field emission from single-wall carbon nanotube films, Appl. Phys. Lett. **73**, 918–920 (1998)
8.383 W. Zhu, C. Bower, O. Zhou, G. Kochanski, S. Jin: Large curent density from carbon nanotube field emitters, Appl. Phys. Lett. **75**, 873–875 (1999)
8.384 Y. Saito, R. Mizushima, T. Tanaka, K. Tohji, K. Uchida, M. Yumura, S. Uemura: Synthesis, structure, and field emission of carbon nanotubes, Fuller. Sci. Technol. **7**, 653–664 (1999)
8.385 F. Houdellier, A. Masseboeuf, M. Monthioux, M.J. Hÿtch: New carbon cone nanotip for use in a highly coherent cold field emission electron microscope, Carbon **50**, 2037–2044 (2012)
8.386 C. Peltola: Weeks, I.A. Levitsky, D.A. Britz, P. Glatkowski, M. Trottier, T. Huang: Carbon-nanotube transparent electrodes for flexible displays, Inf. Disp. **23**, 20–23 (2007)
8.387 C.M. Trottier, P. Glatkowski, P. Wallis, J. Luo: Properties and characterization of carbon-nanotube-based transparent conductive coating, J. Soc. Inf. Disp. **13**, 759–763 (2005)
8.388 D.S. Hecht, L.B. Hu, G. Irvin: Emerging transparent electrodes based on thin films of carbon nanotubes, graphene, and metallic nanostructures, Adv. Mater. **23**, 1482–1513 (2011)
8.389 T. Rueckes, K. Kim, E. Joselevich, G.Y. Tseng, C.-L. Cheung, C.M. Lieber: Carbon nanotube-based nonvolatile random access memory for molecular computing, Science **289**, 94–97 (2000)
8.390 T. Hawkins, S. Drozdz: Cathodic nanocoating technology for corrosion control and steel structures. In: *Waterbone Symp., Proc. Thirty-Ninth Ann. Int. Waterborne, High Solids, Powder Coat. Symp.*, ed. by J.W. Rawlins, R.F. Storey (DEStech, Lancaster 2012) pp. 100–105
8.391 J.R. Dahn: Phase diagram of Li_xC_6, Phys. Rev. B **44**, 9170–9177 (1991)
8.392 C. de las Casas, W. Li: A review of application of carbon nanotubes for lithium ion battery anode material, J. Power Sources **208**, 74–85 (2012)
8.393 G. Maurin, F. Henn: Electrochemical insertion of lithium in carbon nanotubes. In: *Encylopedia of Nanoscience and Nanotechnology*, ed. by H.S. Nalwa (American Scientific, California 2003) pp. 773–792
8.394 B.J. Landi, M.J. Ganter, C.D. Cress, R.A. DiLeo, R.P. Raffaelle: Carbon nanotubes for lithium ion batteries, Energ. Environ. Sci. **2**, 638–654 (2009)
8.395 L. Dai, D.W. Chang, J.-B. Baek, W. Lu: Carbon nanomaterials for advanced energy conversion and storage, Small **8**, 1130–1166 (2012)
8.396 M. Endo, T. Hayashi, Y.-A. Kim: Large-scale production of carbon nanotubes and their applications, Pure Appl. Chem. **78**, 1703–1713 (2006)
8.397 T. Zhang, S. Mubeen, N.V. Myung, M.A. Deshusses: Recent progress in carbon nanotube-based gas sensors, Nanotechnology **19**, 332001 (2008)
8.398 Y. Wang, J.T.W. Yeow: A Review of carbon nanotubes-based gas sensors, J. Sensors **2009**, 493904 (2009)
8.399 P. Bondavalli, P. Legagneux, D. Pribat: Carbon nanotubes based transistors as gas sensors: State of

the art and critical review, Sensors Actuat. B **140**, 304–318 (2009)
8.400 M. Meyyappan: Carbon nanotube-based chemical sensors, Small **12**, 2118–2129 (2016)
8.401 D.W.H. Fam, A. Palaniappana, A.I.Y. Tok, B. Liedberga, S.M. Moochhal: A review on technological aspects influencing commercialization of carbon nanotube sensors, Sensors Actuat. B **157**, 1–7 (2011)
8.402 P. Serp: Carbon. In: *Comprehensive Inorganic Chemistry II*, Vol. 7, ed. by J. Reedijk, K. Poeppelmeier (Elsevier, Oxford 2013) pp. 323–369
8.403 J.M. Planeix, N. Coustel, B. Coq, B. Botrons, P.S. Kumbhar, R. Dutartre, P. Geneste, P. Bernier, P.M. Ajayan: Application of carbon nanotubes as supports in heterogeneous catalysis, J. Am. Chem. Soc. **116**, 7935–7936 (1994)
8.404 P. Serp, B. Machado (Eds.): *Nanostructured Carbon Materials for Catalysis*, RSC Catalysis (Cambridge, Cambridge 2015)
8.405 E. Antolini: Carbon supports for low-temperature fuel cell catalysts, Appl. Catal. B **88**, 1–24 (2009)
8.406 S. Mukherjee, A. Bates, S.C. Lee, D.-H. Lee, S. Park: A Review of the application of CNTs in PEM fuel cells, Int. J. Green Energ. **12**, 787–809 (2015)
8.407 R. Hurt, M. Monthioux, A. Kane: Toxicology of carbon nanomaterials: Status, trends, and perspectives on the special issue, Carbon **44**(6), 1028–1033 (2006)
8.408 C. Salvador-Morales, E. Flahaut, E. Sim, J. Sloan, M.L.H. Green, R.B. Sim: Complement activation and protein adsorption by carbon nanotubes, Mol. Immun. **43**, 193–201 (2006)
8.409 C. Salvador-Morales, P. Townsend, E. Flahaut, C. Vénien-Bryan, A. Vlandas, M.L.H. Green, R.B. Sim: Binding of pulmonary surfactant proteins to carbon nanotubes; potential for damage to lung immune defence mechanisms, Carbon **45**, 607–617 (2007)

9. Nanowires

Mildred S. Dresselhaus, Marcie R. Black, Vincent Meunier, Oded Rabin

This chapter provides an up-to-date overview of research on inorganic nanowires, particularly metallic and semiconducting nanowires. Nanowires are one-dimensional, anisotropic structures, small in diameter, and large in surface-to-volume ratio. Their physical properties are different than those of structures of other scales and dimensionality. While the study of nanowires is particularly challenging, scientists have made immense progress in developing synthetic methodologies for the fabrication of nanowires, developing instrumentation for their characterization, and incorporating nanowires as functional elements in advanced materials and devices. The chapter is divided into three main sections addressing the synthesis, the physical properties, and the applications of nanowires. Yet, the reader will discover many links that make these aspects of nanoscience intimately interdependent.

9.1 **Synthesis** 250
9.1.1 Template-Assisted Synthesis 250
9.1.2 VLS Method for Nanowire Synthesis 255
9.1.3 Etching Methods 257
9.1.4 Other Synthesis Methods 258
9.1.5 Nanowire Alignment and Superstructures of Nanowires 260
9.2 **Characterization and Physical Properties of Nanowires** 262
9.2.1 Structural Characterization 262
9.2.2 Mechanical Properties 266
9.2.3 Transport Properties 267
9.2.4 Optical Properties 277
9.3 **Applications** 282
9.3.1 Electrical Applications 282
9.3.2 Optical Applications 284
9.3.3 Energy Applications 286
9.3.4 Chemical and Biochemical Sensing Devices 288
9.3.5 Magnetic Applications 289
9.4 **Concluding Remarks** 290
References 290

Nanowires are attracting much interest from those seeking to apply nanotechnology and (especially) those investigating nanoscience. Nanowires, unlike other low-dimensional systems, have two quantum-confined directions but one unconfined direction available for electrical conduction. This allows nanowires to be used in applications where electrical conduction, rather than tunneling transport, is required. Because of their unique density of electronic states, in the limit of small diameters nanowires are expected to exhibit significantly different optical, electrical, and magnetic properties to their bulk three-dimensional (3-D) crystalline counterparts. Increased surface area, very high density of electronic states and joint density of states near the energies of their van Hove singularities, enhanced exciton binding energy, diameter-dependent bandgap, and increased surface scattering for electrons and phonons are just some of the ways in which nanowires differ from their corresponding bulk materials. Yet the sizes of nanowires are typically large enough (> 1 nm in the quantum-confined directions) to result in local crystal structures that are closely related to their parent materials, allowing theoretical predictions about their properties to be made based on knowledge of their bulk properties.

Not only do nanowires exhibit many properties that are similar to, and others that are distinctly different from, those of their bulk counterparts, nanowires also have the advantage from an applications standpoint that material parameters critical for certain properties can be independently controlled in nanowires but not in their bulk counterparts. Certain properties can also be enhanced nonlinearly in small-diameter nanowires, by exploiting the singular aspects of the one-dimensional (1-D) electronic density of states.

Furthermore, nanowires have been shown to provide a promising framework for applying the *bottom-up*

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_9

approach [9.1] to the design of nanostructures for nanoscience investigations and for potential nanotechnology applications.

Driven by (1) these new research and development opportunities, (2) the smaller and smaller length scales now being used in the semiconductor, optoelectronics, and magnetics industries, and (3) the dramatic development of the biotechnology industry where the action is also at the nanoscale, the nanowire research field has developed with exceptional speed in the last decade. Therefore, a review of the current status of nanowire research is of significant broad interest at the present time. It is the aim of this review to focus on nanowire properties that differ from those of their parent crystalline bulk materials, with an eye toward possible applications that might emerge from the unique properties of nanowires and from future discoveries in this field.

For quick reference, examples of typical nanowires that have been synthesized and studied are listed in Table 9.1. Also of use to the reader are review articles that focus on a comparison between nanowire and nanotube properties [9.2] and the many reviews that have been written about carbon nanotubes [9.3–5], which can be considered as model one-dimensional systems.

9.1 Synthesis

In this section we survey the most common synthetic approaches that have successfully afforded high-quality nanowires of a large variety of materials (Table 9.1). In Sect. 9.1.1, we discuss methods that make use of various templates with nanochannels to confine the nanowire growth in two dimensions. In Sect. 9.1.2, we present the synthesis of nanowires by the vapor–liquid–solid mechanism and its many variations. In Sect. 9.1.3, etching methods for vertically aligned silicon nanowires are reviewed. In Sect. 9.1.4, examples of other synthetic methods of general applicability are presented. The last part of this section (Sect. 9.1.5) features several approaches that have been developed to organize nanowires into simple architectures.

9.1.1 Template-Assisted Synthesis

The template-assisted synthesis of nanowires is a conceptually simple and intuitive way to fabricate nanostructures [9.67–69]. In this context, templates are synthetic or naturally occurring host materials that contain small-diameter cylindrical pores or voids. These empty volumes are filled with a material of choice, which adopts the pore morphology, to form nanowires. In this section, commonly used templates and techniques for filling the templates to make nanowires are described.

Templates and Fabrication Thereof

In template-assisted synthesis of nanostructures, the chemical stability and mechanical properties of the template, as well as the diameter, uniformity, and density of the pores are important characteristics to consider. Templates frequently used for nanowire synthesis include anodic alumina (Al_2O_3) and ion-track–etched polymer films. Both are commercially available (as filtration membranes). Nanochannel glass, mica films, mesoporous materials, block-copolymers, inorganic nanotubes, viruses, and biopolymers have also been used as nanowire templates.

Porous anodic alumina templates are produced by anodizing pure Al films in selected acids [9.70–72]. Under carefully chosen anodization conditions, the resulting oxide film possesses a regular hexagonal array of parallel and nearly cylindrical channels, as shown in Fig. 9.1a. The self-organization of the pore structure in an anodic alumina template involves two coupled processes: pore formation with uniform diameters and pore ordering. The pores form with uniform diameters because of a delicate balance between electric field–enhanced diffusion, which determines the growth rate of the alumina, and dissolution of the alumina into the acidic electrolyte [9.73]. The pores self-order because of mechanical stress at the aluminum-alumina interface due to volume expansion during the anodization process. This stress produces a repulsive force between the pores, causing them to arrange in a close-packed lattice [9.74]. Depending on the anodization conditions, the pore diameter can be systematically varied from ≤ 10 up to 200 nm with a pore density in the range of 10^9–10^{11} pores/cm^2 [9.22, 34, 70, 71]. It has been shown by many groups that the pore size distribution and the pore ordering of the anodic alumina templates can be significantly improved by a two-step anodization technique [9.75–77], where the aluminum oxide layer is dissolved after the first anodization in an acidic solution followed by a second anodization under the same conditions. Alternatively, the locations of the pores can be controlled by pretexturing the surface of the aluminum [9.78, 79].

Track-etched polymer foils are nanowire templates produced by bombardment of polymer films in a heavy-ion accelerator followed by chemical etching [9.81]. The ion-beam bombardment creates latent tracks of

Table 9.1 Selected syntheses of nanowires by material

Material	Growth Technique	Reference
ABO_4-type	Template[a]	[9.6]
Ag	DNA-template, redox	[9.7]
	Template, pulsed ECD[b]	[9.8]
	Liquid-phase	[9.9–11]
Au	Template, ECD[b]	[9.12, 13]
	Liquid-phase	[9.14]
Bi	Stress-induced	[9.15, 16]
	Template, vapor-phase	[9.17]
	Template, ECD[b]	[9.18–20]
	Template, pressure injection	[9.21–23]
BiSb	Pulsed ECD[b]	[9.24]
Bi_2Te_3	Template, DC ECD[b]	[9.25]
CdS	Liquid-phase (surfactant), recrystallization	[9.26]
	Template, AC ECD[b]	[9.27, 28]
CdSe	Liquid-phase (surfactant), redox	[9.29]
	Template, AC ECD[b]	[9.30, 31]
Cu	Vapor deposition	[9.32]
	Template, ECD[b]	[9.33]
Fe	Template, ECD[b]	[9.34, 35]
	Shadow deposition	[9.36]
GaN	Template, CVD[c]	[9.37, 38]
	VLS[d]	[9.39, 40]
GaAs	Template, liquid/vapor MOCVD[e]	[9.41]
Ge	High-T, high-P liquid-phase, redox	[9.42]
	VLS[d]	[9.43]
	Oxide-assisted	[9.44]
InAs	VLS[d]	[9.45]
MgO	VLS[d]	[9.46]
Mo	Step decoration, ECD[b] + redox	[9.47]
Ni	Template, ECD[b]	[9.20, 48, 49]
Pb	Liquid-phase[f]	[9.50]
PbSe	Liquid-phase	[9.51]
	Self-assembly of nanocrystals[g]	[9.52]
Pd	Step decoration, ECD[b]	[9.53]
	Liquid-phase	[9.54]
Se	Liquid-phase, recrystallization	[9.55]
	Template, pressure injection	[9.56]
Si	VLS[d]	[9.57]
	Laser-ablation VLS[d]	[9.58]
	Oxide-assisted	[9.59]
	Low-T VLS[d]	[9.60]
	MACE[h]	[9.61]
W	Vapor transport	[9.62]
Zn	Template, vapor-phase	[9.63]
	Template, ECD[b]	[9.64]
ZnO	VLS[d]	[9.65]
	Template, ECD[b]	[9.64, 66]

[a] Template synthesis, [b] Electrochemical deposition (ECD), [c] Chemical vapor deposition (CVD), [d] Vapor–liquid–solid (VLS) growth, [e] Metalorganic chemical vapor deposition (MOCVD), [f] Liquid phase synthesis, [g] Self-assembly of nanocrystals (in liquid phase), [h] Metal-assisted chemical etching

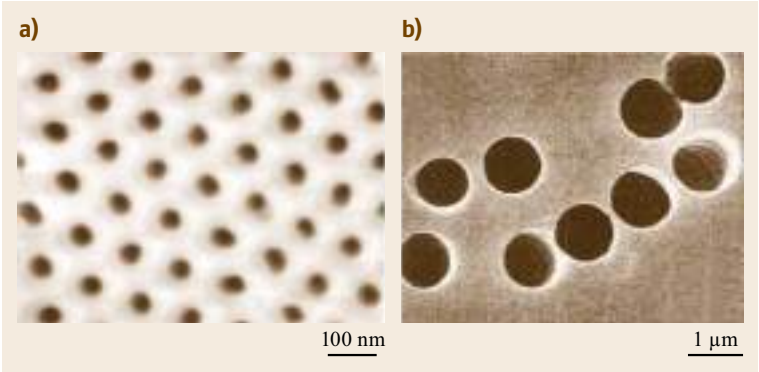


Fig. 9.1 **(a)** Scanning electron microscope (SEM) images of the top surfaces of porous anodic alumina templates anodized with an average pore diameter of 44 nm (after [9.75]). **(b)** SEM image of the particle track-etched polycarbonate membrane, with a pore diameter of 1 μm (after [9.80])

damage in the polymer. These areas are susceptible to chemical etch and can be selectively removed to create hollow channels [9.82]. This technique offers control over pore size, pore aspect ratio, pore shape, and pore density; however, the pores are randomly distributed. Track-etched polycarbonate membranes (Fig. 9.1b) are economical, easy to handle, and dissolve in organic solvents [9.83, 84]. The track-etching technique can be applied also to hard materials: irradiated single-crystal mica films etched with hydrofluoric acid afford nanochannels of diamond-shaped cross-section in a refractory host material [9.48].

Other porous materials can be used as host templates for nanowire growth, as discussed by *Ozin* [9.67]. Nanochannel glass (NCG), for example, contains a regular hexagonal array of glass capillaries similar to the pore structure in anodic alumina with a packing density as high as 3×10^{10} pores/cm^2 [9.68]. Porous Vycor glass that contains an interconnected network of pores less than 10 nm was also employed for the early study of nanostructures [9.85]. The nanochannel array template with the smallest pore diameter may have been identified in mesoporous molecular sieves termed MCM-41 [9.86] that possess hexagonally packed pores with very small channel diameters, which can be varied between 2 and 10 nm. Conducting organic filaments have been fabricated in the nanochannels of MCM-41 [9.87].

Templates for nanowire growth can be produced also using diblock copolymers, polymers that consist of two chain segments of different chemistries covalently linked to each other. When the two components are immiscible in each other, phase segregation occurs, and depending on their volume ratio, spheres, cylinders, or lamellae may self-assemble. To form self-assembled arrays of nanopores:

1. The block copolymer must adopt the cylinder morphology.
2. The cylinders must be parallel and oriented perpendicular to the surface.
3. The component in the cylinders must be etched away selectively.

With polystyrene and polymethylmethacrylate (PMMA) block copolymers (P(S-b-MMA)), by applying an electric field while the copolymer was heated above the glass transition temperature of the two constituent polymers, the self-assembled cylinders of PMMA could be aligned with their main axis perpendicular to the film [9.88]. Selective removal of the PMMA component afforded the preparation of 14 nm-diameter ordered pore arrays with a packing density of 1.9×10^{11} cm^{-3}.

To conclude the list of templates used to fabricate nanowires, we note efforts to utilize biology for the task. Biological systems can produce high-aspect-ratio structures with remarkably consistent dimensions and in high volumes. Coating these filamentous structures with inorganic materials results in useful nanowires. But unlike with the porous templates described above, the nanowire growth is not restricted in any direction; and diameter control is implemented through the process conditions. Additionally, the organic material remains trapped inside the nanowire. For example, the DNA molecule has been used as a template for growing nanometer-sized wires [9.7], as were M13 bacteriophages and the tobacco mosaic virus [9.89–92].

Nanowire Template-Assisted Growth by Pressure Injection

The pressure injection technique can be employed for fabricating highly crystalline nanowires from a low-melting point material in porous templates with robust mechanical strength. In the high-pressure injection method, the nanowires are formed by pressure-injecting the desired material in liquid form into the evacuated pores of the template. Due to the heating and pressurization processes, the templates used for the pressure injection method must be chemically stable and be able to maintain their structural integrity at high tempera-

tures and at high pressures. Anodic aluminum oxide films and nanochannel glass are two typical materials used as templates in conjunction with the pressure-injection filling technique. Metal nanowires (Bi, In, Sn, and Al) and semiconductor nanowires (Se, Te, GaSb, and Bi_2Te_3) have been fabricated in anodic aluminum oxide templates using this method [9.21, 56, 93].

The pressure P required to overcome the surface tension for the liquid material to fill pores with a diameter d_W is determined by the Washburn equation [9.94]

$$d_W = -4\frac{\gamma \cos\theta}{P} ,$$

where γ is the surface tension of the liquid, and θ is the contact angle between the liquid and the template. To reduce the required pressure and to maximize the filling factor, some surfactants are used to decrease the surface tension and the contact angle. For example, the introduction of Cu into the Bi melt can facilitate the filling of the pores in the anodic alumina template with liquid Bi and can increase the number of nanowires that are formed [9.22]. However, some surfactants might cause contamination problems and should therefore be avoided. Nanowires produced by the pressure-injection technique usually possess high crystallinity and a preferred crystal orientation along the wire axis. For example, Fig. 9.2 shows the x-ray diffraction (XRD) patterns of Bi nanowire arrays of three different wire diameters with an injection pressure of ≈ 5000 psi [9.93], showing that the major ($> 80\%$) crystal orientation of the wire axis in the 95 and 40 nm diameter Bi nanowire arrays are, respectively, normal to the (202) and (012) lattice planes, which are denoted by $[10\bar{1}1]$ and $[01\bar{1}2]$ when using a hexagonal unit cell, suggesting a wire diameter–dependent crystal growth direction. On the other hand, 30 and 200 nm Bi nanowires produced using a much higher pressure of $> 20\,000$ psi show a different crystal orientation of (001) along the wire axis [9.23], indicating that the preferred crystal orientation may also depend on applied pressure, with the most dense packing direction along the wire axis for the highest applied pressure.

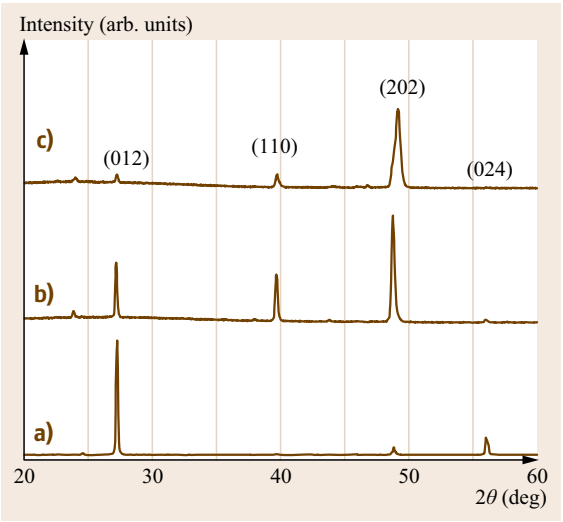


Fig. 9.2a–c XRD patterns of bismuth/anodic alumina nanocomposites with average bismuth wire diameters of **(a)** 40 nm, **(b)** 52 nm, and **(c)** 95 nm [9.93]. The Miller indices corresponding to the lattice planes of bulk Bi are indicated above the individual peaks. The majority of the Bi nanowires are oriented along the $[10\bar{1}1]$ and $[01\bar{1}2]$ directions for $d_W \geq 60$ nm and $d_W \leq 50$ nm respectively (after [9.22, 93]). The existence of more than one dominant orientation in the 52 nm Bi nanowires is attributed to the transitional behavior of *intermediate*-diameter nanowires as the preferential growth orientation is shifted from $[10\bar{1}1]$ to $[01\bar{1}2]$ with decreasing d_W

Electrochemical Deposition

The electrochemical deposition (ECD) technique has attracted increasing attention as a versatile method for fabricating nanowires in templates. Traditionally, electrochemistry has been used to grow thin films on conducting surfaces. Since electrochemical growth is usually controllable in the direction normal to the substrate surface, this method can be readily extended to fabricate 1-D or 0-D nanostructures, if the deposition is confined within the pores of an appropriate template. In the electrochemical methods, a thin conducting metal film is first coated on one side of the porous membrane to serve as the cathode for electroplating. The length of the deposited nanowires can be controlled by varying the duration of the electroplating process. The ECD method has been used to synthesize a wide variety of nanowires, such as metals (Bi [9.18, 84]; Co [9.95, 96]; Fe [9.34, 97]; Cu [9.83, 98]; Ni [9.48, 95]; Ag [9.99]; Au [9.12]); conducting polymers [9.18, 80]; superconductors (Pb [9.100]); compound semiconductors (CdS [9.28]; Bi_2Te_3 [9.25]); and even superlattice nanowires with A/B constituents (such as Cu/Co [9.83, 98]) (Table 9.1). The ECD method offers a low-cost, low-temperature, and rapid approach to template-assisted nanowire synthesis.

In the ECD process, the chosen template has to be chemically stable in the electrolyte during the electroplating process. Cracks and defects in the templates are detrimental to nanowire growth, because deposition rates are higher in the more accessible cracks, leaving most of the nanopores unfilled. Particle-track–etched polymer membranes or mica films are typical templates used in ECD employing a direct current (DC). To use

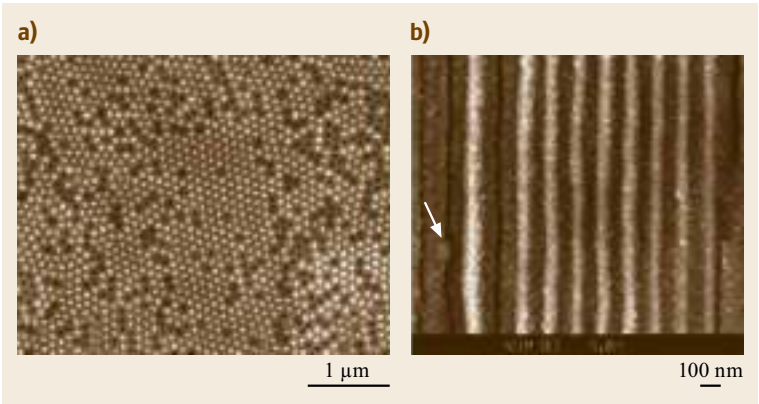


Fig. 9.3 (a) SEM image of a Bi_2Te_3 nanowire array in cross-section showing a relatively high pore filling factor. (b) SEM image of a Bi_2Te_3 nanowire array cross-sectioned along the wire axis (after [9.25])

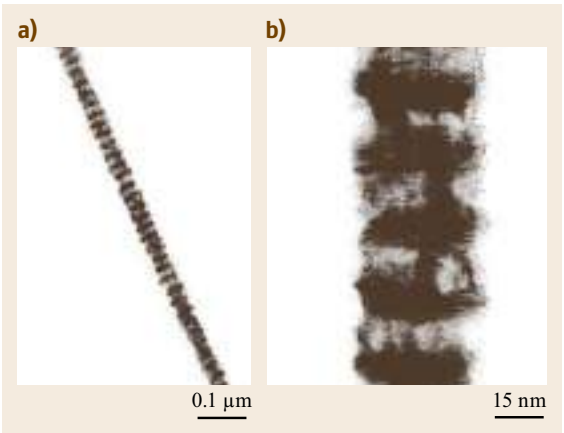


Fig. 9.4 (a) TEM image of a single Co(10 nm)/Cu(10 nm) multilayered nanowire. (b) A selected region of the sample at high magnification (after [9.98])

anodic aluminum oxide films in DC ECD, the insulating barrier layer, separating the pores from the bottom aluminum substrate, has to be removed and a metal film is then evaporated onto the back of the template membrane [9.101]. At the end of the ECD in anodic alumina templates, the surfaces can be polished to even out the variations in nanowire length and reveal the ends of the nanowires. Figure 9.3 presents SEM images of cross-sectional views of a Bi_2Te_3 nanowire array electrodeposited in porous anodic alumina [9.25]. The light areas are associated with Bi_2Te_3 nanowires, the dark regions denote empty pores, and the surrounding gray matrix is alumina. Optimization of the electrochemical conditions and good laboratory practices result in high pore filling factors of close to 100% for ECD methods.

It is possible to employ an alternating current (AC) electrodeposition method in porous anodic alumina templates without the removal of the barrier layer, by utilizing the rectifying properties of the oxide barrier. In AC electrochemical deposition, although the applied voltage is sinusoidal and symmetric, the current is greater during the cathodic half-cycles (deposition current) than in the anodic half-cycles (stripping current). Since no current rectification occurs at defect sites lacking the barrier layer, the deposition and stripping rates are similar and material does not effectively deposit in cracks in the template. In this fashion, metals such as Co [9.96] and Fe [9.34, 97], and semiconductors such as CdS [9.28], have been deposited into the pores of anodic aluminum oxide templates without removing the barrier layer.

In contrast to nanowires synthesized by the pressure injection method, nanowires fabricated by the ECD process are usually polycrystalline, with no preferred crystal orientation, as observed by XRD studies. However, some exceptions exist. For example, anisotropy in crystal growth rates may result in polycrystalline nanowires having a preferred wire growth orientation [9.28]. In addition, *Xu* et al. have prepared a number of single-crystal II–VI semiconductor nanowire arrays, including CdS, CdSe and CdTe, by DC ECD in anodic alumina templates using nonaqueous electrolytes [9.27, 31]. Pulsed electrodeposition techniques can achieve or improve single-crystal nanowire array growth and orientation. The use of pulsed currents is believed to be advantageous for the growth of crystalline wires because the concentration of ions in the electrolyte inside the pores can recover in the period between the electrical pulses and therefore uniform deposition conditions can be produced for each deposition pulse. Single-crystal Ag and Pb nanowires were fabricated by pulse electrodeposition [9.8, 100].

One advantage of the ECD technique is the possibility of fabricating multilayered structures within nanowires. With electrolytes that contain two or more different metal ions, by varying the cathodic potential with time, different metal layers can be controllably

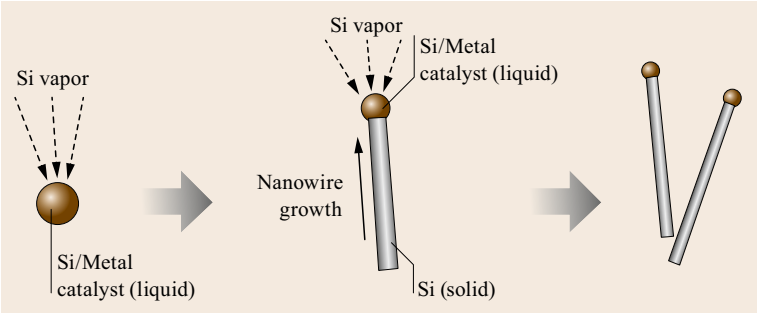


Fig. 9.5 Schematic diagram illustrating the growth of silicon nanowires by the VLS mechanism

deposited in sequence. Co/Cu multilayered nanowires have been synthesized in this way [9.83, 98]. Figure 9.4 shows transmission electron microscopy (TEM) images of a single Co/Cu nanowire about 40 nm in diameter [9.98]. The light bands correspond to Co-rich layers and the dark bands correspond to Cu-rich layers. This electrodeposition method provides a low-cost approach to preparing multilayered 1-D nanostructures.

Vapor Deposition

Vapor deposition of nanowires includes physical vapor deposition (PVD) [9.17], chemical vapor deposition (CVD) [9.38], metalorganic chemical vapor deposition (MOCVD) [9.41] and atomic layer deposition (ALD) [9.102]. Like electrochemical deposition, vapor deposition is usually capable of preparing smaller-diameter (≤ 20 nm) nanowires than pressure injection methods, since surface tension is typically not a barrier for inserting the material into the pores. The main challenge with these techniques is the deposition of material on the surface of the template blocking the pore openings before the pores are filled.

In the PVD process, the material for the nanowires is first vaporized by heat or other means, and is then introduced through the pores of the template and cooled to solidify. Using a specially designed experimental setup [9.17], nearly single-crystal Bi nanowires in anodic aluminum templates with pore diameters as small as 7 nm have been synthesized, and these Bi nanowires were found to possess a preferred crystal growth orientation along the wire axis, similar to the Bi nanowires prepared by pressure injection [9.17, 22].

In the CVD and ALD processes, reacting gas(es) decompose on the surface of the template to form elemental or compound materials. The decomposition is promoted by the CVD reactor temperature and/or a catalyst on the surface. For example, single-crystal GaN nanowires have been synthesized in anodic alumina templates through a gas reaction of Ga_2O vapor with a flowing ammonia atmosphere [9.37, 38]. Templates were filled with TiO_2 by ALD to produce an array of nanowires [9.102], but with high-aspect-ratio pores the ALD product is typically a conformal coating, that is to say, an array of nanotubes [9.103, 104]. A hybrid liquid-/gas-phase CVD approach has been used to prepare polycrystalline GaAs and InAs nanowires in a nanochannel glass array [9.41]. In this method, the nanochannels are filled with one liquid precursor (such as Me_3Ga or Et_3In) via a capillary effect and the nanowires are formed within the template by reactions between the liquid precursor and the other gas reactant (such as AsH_3).

9.1.2 VLS Method for Nanowire Synthesis

A significant number of notable modern syntheses of semiconductor nanowires are based on the so-called vapor–liquid–solid (VLS) mechanism of anisotropic crystal growth. This mechanism was first proposed for the growth of single-crystal silicon whiskers, 100 nm to hundreds of micrometers in diameter [9.105]. The proposed growth mechanism (Fig. 9.5) involves the absorption of source material from the gas phase into a liquid droplet of catalyst (a molten particle of gold on a silicon substrate in the original work [9.105]). Upon supersaturation of the liquid alloy, a nucleation event generates a solid precipitate of the source material. This seed serves as a preferred site for further deposition of material at the interface of the liquid droplet, promoting the elongation of the seed into a nanowire or a whisker, and suppressing further nucleation events on the same catalyst. Since the liquid droplet catalyzes the incorporation of material from the gas source to the growing crystal, the deposit grows anisotropically as a whisker whose diameter is dictated by the diameter of the liquid alloy droplet. The nanowires thus obtained are of high purity, except for the end containing the solidified catalyst as an alloy particle (Figs. 9.5 and 9.6a). Real-time observations of the alloying, nucleation, and elongation steps in the growth of germanium nanowires from gold nanoclusters by the VLS method were recorded by in-situ TEM [9.106].

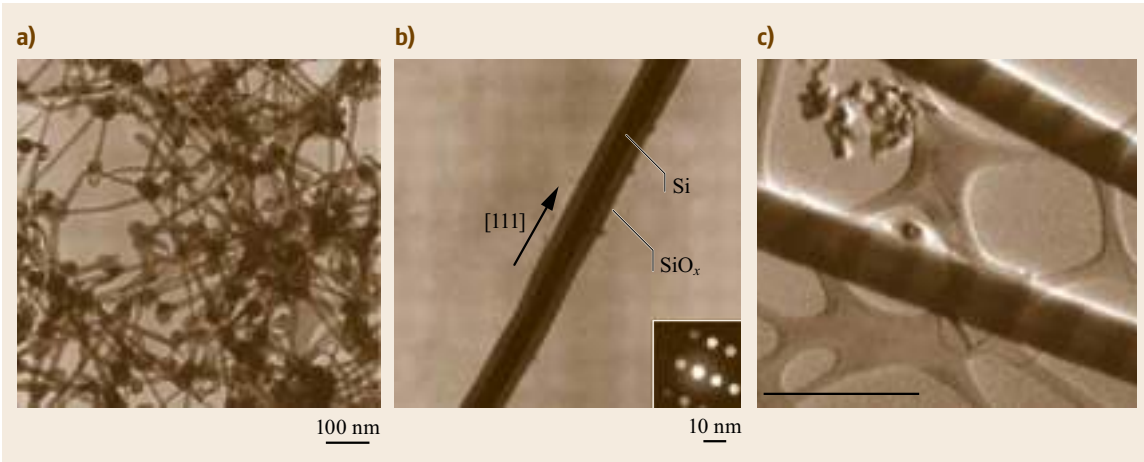


Fig. 9.6 **(a)** TEM images of Si nanowires produced after laser-ablating a $Si_{0.9}Fe_{0.1}$ target. The *dark spheres* with a slightly larger diameter than the wires are solidified catalyst clusters (after [9.58]). **(b)** Diffraction contrast TEM image of a Si nanowire. The crystalline Si core appears *darker* than the amorphous oxide surface layer. The *inset* shows the convergent beam electron diffraction pattern recorded perpendicular to the wire axis, confirming the nanowire crystallinity (after [9.58]). **(c)** Scanning transmission electron microscope (STEM) image of $Si/Si_{1-x}Ge_x$ superlattice nanowires in the bright field mode. The scale bar is 500 nm (after [9.107])

Reduction of the average wire diameter to the nanometer scale requires the generation of nanosized catalyst droplets. However, due to the balance between the liquid-vapor surface free energy and the free energy of condensation, the size of a liquid droplet, in equilibrium with its vapor, is usually limited to the micrometer range. This obstacle has been overcome by several modern methodologies:

1. Advances in the synthesis of metal nanoclusters have made monodispersed nanoparticles commercially available. These can be dispersed on a solid substrate in high dilution so that when the temperature is raised above the melting point, the liquid clusters do not aggregate [9.57].
2. Alternatively, metal islands of nanoscale sizes can self-form when a strained thin layer is grown or heat-treated on a non-epitaxial substrate [9.43].
3. Laser-assisted catalytic VLS growth is a method used to generate nanowires under non-equilibrium conditions. Using laser ablation of a target containing both the catalyst and the source materials, a plasma is generated from which catalyst nanoclusters nucleate as the plasma cools down. Single-crystal nanowires grow as long as the particle remains liquid [9.58].
4. Interestingly, by optimizing the material properties of the catalyst-nanowire system, conditions can be achieved for which nanocrystals nucleate in a liquid catalyst pool supersaturated with the nanowire material, migrate to the surface due to a large surface tension, and continue growing as nanowires perpendicular to the liquid surface [9.60]. In this case, supersaturated nanodroplets are sustained on the outer end of the nanowire due to the low solubility of the nanowire material in the liquid [9.108].

A wide variety of elemental, binary, and compound semiconductor nanowires have been synthesized via the VLS method, and relatively good control over the nanowire diameter and diameter distribution has been achieved. The nanowires are often single crystals with a preferred crystallographic orientation along the wire axis, and may be oriented with respect to the substrate by epitaxy relations [9.109, 110]. Researchers are currently focusing their attention on the controlled variation of the material properties along the nanowire axis. In this context, researchers have modified the VLS synthesis apparatus to generate compositionally modulated nanowires. GaAs/GaP–modulated nanowires have been synthesized by alternately ablating targets of the corresponding materials in the presence of gold nanoparticles [9.111]. p-Si/n-Si nanowires were grown by chemical vapor deposition from alternating gaseous mixtures containing the appropriate dopant [9.111]. $Si/Si_{1-x}Ge_x$ nanowires were grown by combining silicon from a gaseous source with germanium from a periodically ablated target (Fig. 9.6c) [9.107]. NiSi-Si nanowires have been successfully synthesized that directly incorporate a nanowire metal contact into active nanowire devices [9.112]. Finally, using an ultra-high vacuum chamber and molecular beams, InAs/InP

nanowires with atomically sharp interfaces were obtained [9.113]. These compositionally modulated nanowires are expected to exhibit exciting electronic, photonic, and thermoelectric properties.

Interestingly, silicon and germanium nanowires grown by the VLS method consist of a crystalline core coated with a relatively thick amorphous oxide layer (2−3 nm) (Fig. 9.6b). These layers are too thick to be the result of ambient oxidation, and it has been shown that these oxides play an important role in the nanowire growth process [9.59, 114]. Silicon oxides were found to serve as a special and highly selective catalyst that significantly enhances the yield of Si nanowires without the need for metal catalyst particles [9.59, 114, 115]. A similar yield enhancement was also found in the synthesis of Ge nanowires from the laser ablation of Ge powder mixed with GeO_2 [9.44]. The Si and Ge nanowires produced from these metal-free targets generally grow along the [112] crystal direction [9.116], and have the benefit that no catalyst material is found on either ends of the nanowires. Based on these observations and other TEM studies [9.44, 114, 116], an oxide-enhanced nanowire growth mechanism different from the classical VLS mechanism was proposed, where no metal catalyst is required during the laser ablation-assisted synthesis [9.114]. It is postulated that the nanowire growth is dependent on the presence of SiO (or GeO) vapor, which decomposes in the nanowire tip region into both Si (or Ge), which is incorporated into the crystalline phase, and SiO_2 (or GeO_2), which contributes to the outer coating. The initial nucleation events generate oxide-coated spherical nanocrystals. The (112) crystal faces have the fastest growth rate, and therefore the nanocrystals soon begin elongating in the perpendicular direction to form one-dimensional structures. The Si_mO or Ge_mO ($m > 1$) layers on the nanowire tips may be in or near their molten states, catalyzing the incorporation of gas molecules in a directional fashion [9.116]. Besides nanowires with smooth walls, a second morphology of chains of unoriented nanocrystals linked by oxide necks is frequently observed (indicated by an arrow in Fig. 9.7). In addition, scanning tunneling microscope (STM) studies revealed that about 1% of the wires consist of a third morphology: zigzag nanowires 3 nm in diameter and microns in length [9.117]. The recurring structure of two alternating straight wire segments, 10 and 5 nm in length respectively, whose junctions form an angle of 30°, is probably a result of alternating growth along different crystallographic orientations [9.117].

Other variations of the VLS mechanism of nanowire growth include synthesis at low temperatures where the catalysts are solid nanoparticles (i. e., vapor-solid-solid or VSS) [9.109, 118, 119] and synthesis in a liquid phase where the catalyst is suspended in a liquid solution in which the reactants are dissolved (i. e., solution–liquid–solid or SLS) [9.120].



Fig. 9.7 TEM image showing the two major morphologies of Si nanowires prepared by the oxide-assisted growth method (after [9.114]). Notice the absence of metal particles when compared to Fig. 9.6a. The *arrow* points at an oxide-linked chain of Si nanoparticles

9.1.3 Etching Methods

For applications that require uniform nanowires, specific crystallographic orientation, control of the nanowire array dimensions, and/or high-density vertical nanowires, etching methods offer a practical *top-down* alternative method to fabricate nanowire arrays. The two most common methods of making etched nanowires are reactive ion etching (RIE) and metal-assisted chemical etching (MACE). Both methods result in vertically aligned nanowire arrays with high densities. In addition, in versions of both methods, the diameter and density of the nanowires can be controlled. Furthermore, MACE and RIE etched nanowires can result in large arrays of nanowires with uniform heights. Silicon, due to its favorable properties for electronic and photonic applications and its prevalence in the semiconductor industry, is the dominant material used to make etched nanowires.

Reactive ion etching is a vacuum technique used in the semiconductor industry to etch silicon and III–V semiconductors. Energetic charged particles bombarding the substrate remove material by a combination of sputtering and chemical attack. Process engineers have long known that when process conditions are not set a certain way unpatterned regions in the RIE-etched silicon wafers will turn into black silicon [9.121]. The black silicon is in fact a dense array of vertical nanowires, also known as silicon grass. An example of silicon grass is shown in Fig. 9.8. While unwanted by the complementary metal-oxide-semiconductor (CMOS)



Fig. 9.8 A SEM image of silicon grass formed by RIE. ©IOP Publishing. Reproduced with permission. All rights reserved (after [9.121])

industry, black silicon has proven useful for making and studying nanowire arrays. Ultimately, RIE methods have been used to form patterned silicon nanowire arrays in a controllable fashion. RIE and often deep RIE (DRIE) are used in combination with a mask to pattern where nanowires form [9.122]. The ordered arrays of vertical nanowires shown in Fig. 9.9 are examples of wire arrays formed by using e-beam lithography in combination with deep RIE.

Metal-assisted chemical etching to make silicon nanowire arrays is a well explored research topic. The resulting silicon structure made from MACE is often called *black silicon*, *silicon whiskers*, *nanotexture*, *porous silicon,* or *nanowires*. Recently several reviews have covered the process and the capabilities of MACE [9.123–125]. In MACE a thin metal is deposited onto a silicon surface and then immersed into an etch solution containing hydrofluoric acid and an oxidizer. The metal catalyzes the etching of silicon. Thus silicon is etched where the metal is in contact with the silicon and is not etched where the metal is absent. As the etching progresses, the metal moves down the silicon with the etching front. In the case where the metal is put down such that there are holes in the metal, nanowires form as the metal etches the silicon around the nanowires. MACE results in very anisotropic etched structures. See Fig. 9.10 for an example of a MACE-etched array of nanowires. In this particular example, the metal was masked using a silica colloidal crystal template such that the metal was blocked from touching the silicon wherever a silica sphere was present, thus forming a nanowire [9.61].

9.1.4 Other Synthesis Methods

In this section we review several other general procedures available for the synthesis of a variety of nanowires. We focus on *bottom-up* approaches, which afford many kinds of nanowires in large numbers, and do not require highly sophisticated equipment (such as scanning microscopy or lithography-based methods), and exclude cases for which the nanowires are not self-sustained (such as in the case of atomic rows on the surface of crystals).

A solution-phase synthesis of nanowires with controllable diameters has been demonstrated [9.55, 126], without the use of templates, catalysts, or surfactants.

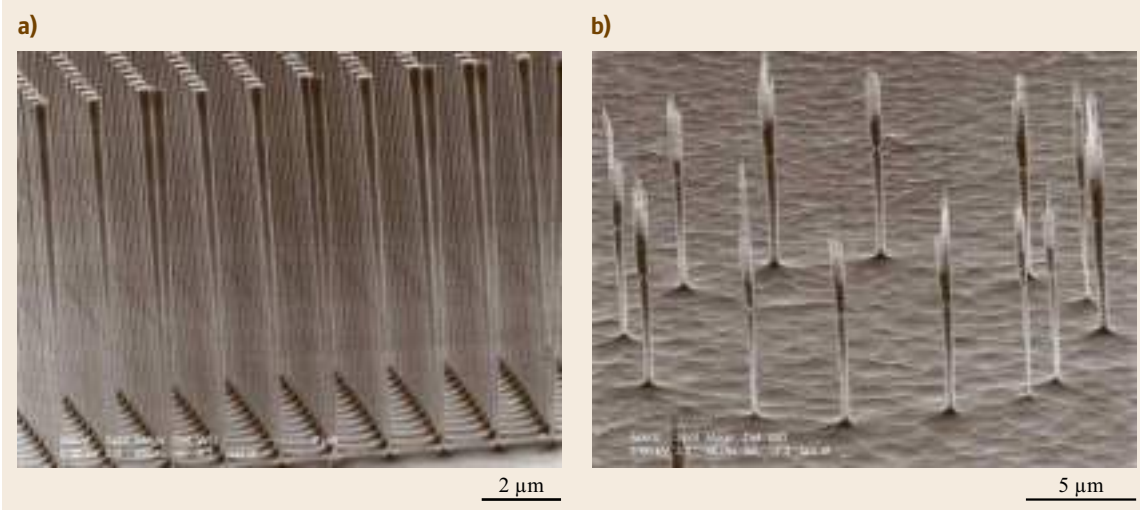


Fig. 9.9 Examples of patterned RIE etched nanowires. In these examples the RIE was patterned by using e-beam lithography. Copyright 2009, American Vacuum Society. Reprinted with permission from [9.122]

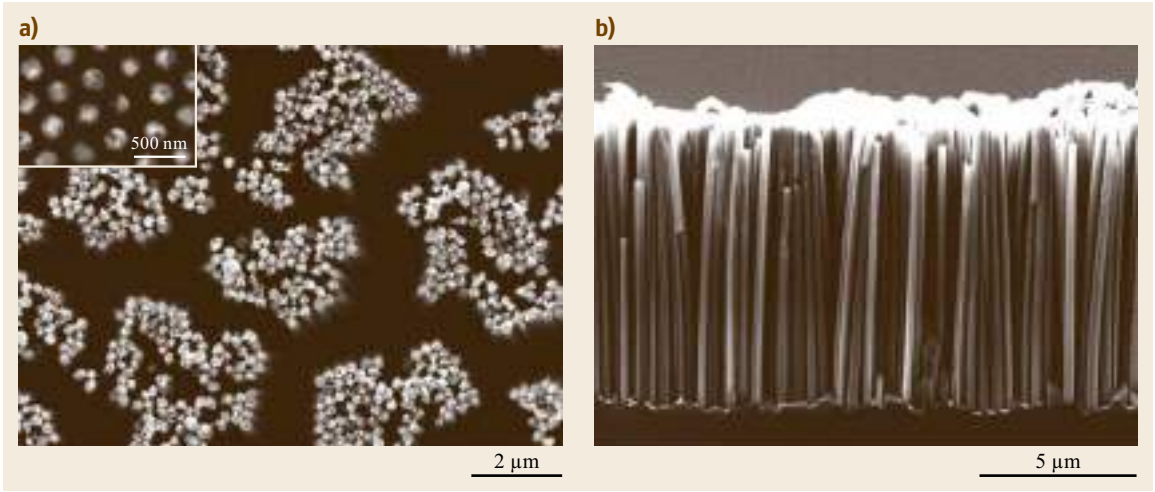


Fig. 9.10a,b MACE silicon nanowires formed by masking the metal with colloidal silica spheres. The silicon is etched wherever the metal touches the silicon. The nanowires form where the metal is blocked from touching the silicon. **(a)** Top view and **(b)** cross-section view of the array. The long nanowires bundle at the tips during drying, while short nanowires (up to 1 µm) retain the arrangement of the silica spheres (*inset*). (Reprinted from [9.61], with permission of AIP Publishing)

Instead, Gates et al. make use of the anisotropy of the crystal structure of trigonal selenium and tellurium, which can be viewed as rows of 1-D helical atomic chains. Their approach is based on the mass transfer of atoms during an aging step from a high free-energy solid phase (e.g., amorphous selenium) to a seed (e.g., trigonal selenium nanocrystal), which grows preferentially along one crystallographic axis. The lateral dimension of the seed, which dictates the diameter of the nanowire, can be controlled by the temperature of the nucleation step. Furthermore, Se/Te alloy nanowires were synthesized by this method, and Ag_2Se compound nanowires were obtained by treating selenium nanowires with $AgNO_3$ [9.127–129]. In a separate work, tellurium nanowires were transformed into Bi_2Te_3 nanowires by their reaction with $BiPh_3$ [9.130].

More often, however, the use of surfactants is necessary to promote the anisotropic 1-D growth of nanocrystals in solution. Solution-phase synthetic routes have been optimized to produce monodispersed quantum dots (zero-dimensional isotropic nanocrystals) [9.131]. Surfactants are necessary in this case to stabilize the interfaces of the nanoparticles and to retard oxidation and aggregation processes. Detailed studies on the effect of growth conditions revealed that they can be manipulated to induce a directional growth of the nanocrystals, usually generating nanorods (aspect ratio of ≈ 10), and in favorable cases, nanowires with high aspect ratios. *Heath* and *LeGoues* synthesized germanium nanowires by reducing a mixture of $GeCl_4$ and phenyl-$GeCl_3$ at high temperature and high pressure. The phenyl ligand was essential for the formation of high-aspect-ratio nanowires [9.42]. In growing CdSe nanorods [9.29], *Alivisatos* et al. used a mixture of two surfactants, whose concentration ratio influenced the shape of the nanocrystal. It is believed that different surfactants have different affinities, and different absorption rates, for the different crystal faces of CdSe, thereby regulating the growth rates of these faces. In the liquid-phase synthesis of Bi nanowires, the additive $NaN(SiMe_3)_2$ induces the growth of nanowires oriented along the [110] crystal direction from small bismuth seed clusters, while water solely retarded the growth along the [001] direction, inducing the growth of hexagonal-plate particles [9.130]. A coordinating alkyl-diamine solvent was used to grow polycrystalline PbSe nanowires at low temperatures [9.51]. Here, the surfactant-induced directional growth is believed to occur through the formation of organometallic complexes in which the bidentate ligand assumes the equatorial positions, thus hindering the ions from approaching each other in this plane. Additionally, the alkyl-diamine molecules coat the external surface of the wire, preventing lateral growth. The aspect ratio of the wires increased as the temperature was lowered in the range $10\,°C < T < 117\,°C$. Noble metal nanowires were grown by the polyol-reduction method utilizing polyvinylpyrrolidone as a surfactant [9.9–11, 54]. These nanowires have a pentagonal cross-section. The 1-D growth is a consequence of extending the twin boundaries in the dodecahedral seed nanoparticles that

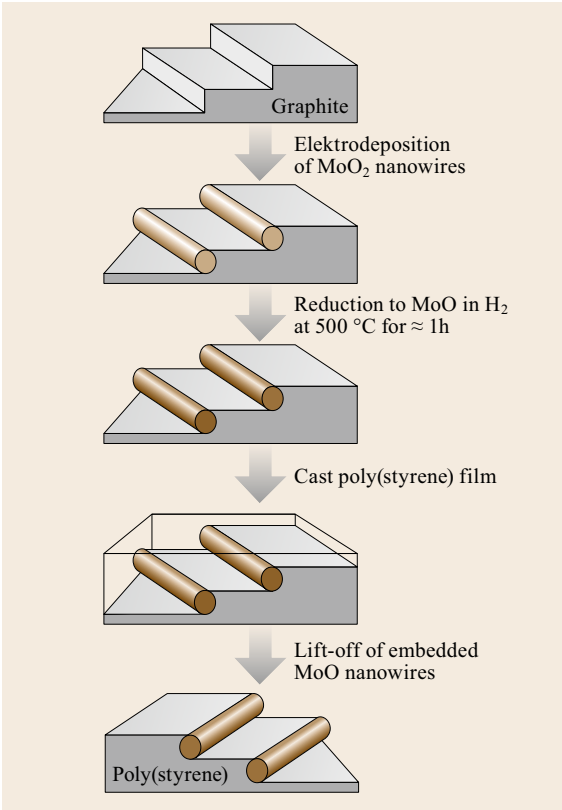


Fig. 9.11 Schematic of the electrodeposition step edge decoration of HOPG (highly oriented pyrolytic graphite) for the synthesis of molybdenum nanowires (after [9.47, 132])

form upon nucleation and passivation of the (100) surfaces by the surfactant.

Stress-induced crystalline bismuth nanowires have been grown from sputtered films of layers of Bi and CrN. The nanowires presumably grow from defects and cleavage fractures in the film, and are up to several millimeters in length with diameters ranging from 30 to 200 nm [9.15]. Other elemental and metal chalcogenide nanowires of soft, low-melting point metals (Bi, Pb, Ga) have been observed to grow from polycrystalline films under stress. This approach offers the benefits of being low-cost and free of contaminants [9.16, 133, 134].

Selective electrodeposition along the step edges in highly oriented pyrolytic graphite (HOPG) was used to obtain MoO_2 nanowires as shown in Fig. 9.11. The site selectivity was achieved by applying a low overpotential to the electrochemical cell in which the HOPG served as cathode, thus minimizing the nucleation events on less favorable sites (i. e., plateaus). While these nanowires cannot be removed from the substrate, they can be reduced to metallic molybdenum

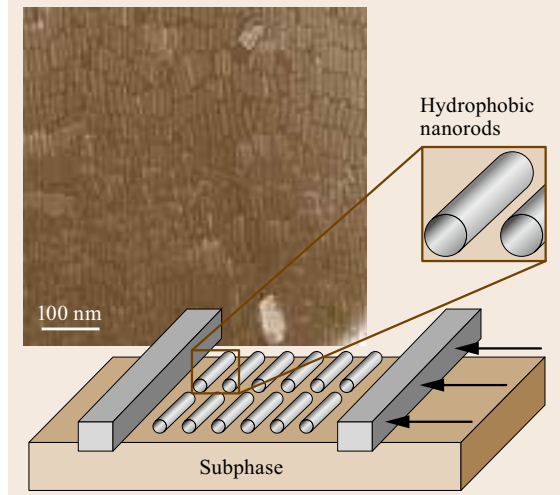


Fig. 9.12 A TEM image of a smectic phase of a $BaCrO_4$ nanorod film (*left inset*) achieved by the Langmuir–Blodgett technique, as depicted by the illustration (after [9.136])

nanowires, which can then be released as free-standing nanowires. Other metallic nanowires were also obtained by this method [9.47, 132]. In contrast to the template synthesis approaches described above, in this method the substrate only defines the position and orientation of the nanowire, not its diameter. In this context, other surface morphologies, such as self-assembled grooves in etched crystal planes, have been used to generate nanowire arrays via gas-phase shadow deposition (for example, Fe nanowires on (110)NaCl [9.36]). The cross-section of artificially prepared superlattice structures has also been used for site-selective deposition of parallel and closely spaced nanowires [9.135]. Nanowires prepared on the above-mentioned substrates have semicircular, rectangular, or other unconventional cross-sectional shapes.

9.1.5 Nanowire Alignment and Superstructures of Nanowires

Controlling the position of a nanowire in the growth process is important for preparing devices or test structures containing nanowires, especially when it involves a large array of nanowires. Post-synthesis methods to align and position nanowires include microfluidic channels [9.137], Langmuir–Blodgett assemblies [9.136], and electric-field–assisted assembly [9.138]. The first method involves the orientation of the nanowires by the liquid flow direction when a nanowire solution is injected into a microfluidic channel assembly and by the interaction of the nanowires with the side walls of

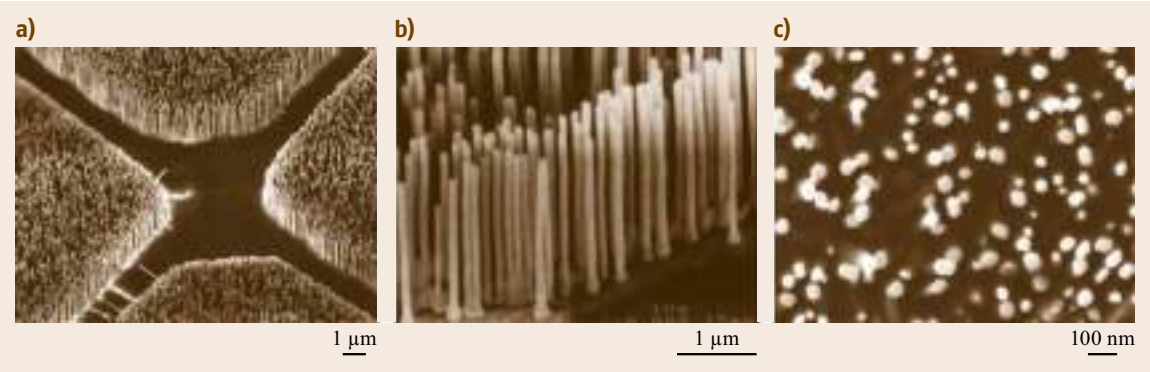


Fig. 9.13a–c SEM images of ZnO nanowire arrays grown on a sapphire substrate, where **(a)** shows patterned growth, **(b)** shows a higher-resolution image of the parallel alignment of the nanowires, and **(c)** shows the faceted side-walls and the hexagonal cross-section of the nanowires. For nanowire growth, the sapphire substrates were coated with a 1.0–3.5 nm-thick patterned layer of Au as the catalyst, using a TEM grid as the shadow mask. These nanowires have been used for nanowire laser applications (after [9.110])

the channel. The second method involves the alignment of nanowires at a liquid-gas or liquid-liquid interface by the application of compressive forces on the interface (Fig. 9.12). The aligned nanowire films can then be transferred onto a substrate and lithography methods can be used to define interconnects. This allows the nanowires to be organized with a controlled alignment and spacing over large areas. Using this method, centimeter-scale arrays containing thousands of single silicon nanowire field-effect transistors with high performance could be assembled to make large-scale nanowire circuits and devices [9.139, 140]. The third technique is based on dielectrophoretic forces that pull polarizable nanowires toward regions of high field strength [9.141, 142]. The nanowires align between two isolated electrodes that are capacitively coupled to a pair of buried electrodes biased with an AC voltage. Once a nanowire shorts the electrodes, the electric field is eliminated, preventing more nanowires from depositing. The above techniques have been successfully used to prepare electronic circuitry and optical devices out of nanowires (Sects. 9.3.1 and 9.3.2). Alternatively, alignment and positioning of the nanowires can be specified and controlled during their growth by the proper design of the synthesis method. For example, ZnO nanowires prepared by the VLS method were grown into an array in which both their position on the substrate and their growth direction and orientation were controlled [9.65]. The nanowire growth region was defined by patterning the gold film, which serves as a catalyst for the ZnO nanowire growth, employing soft lithography, e-beam lithography, or photolithography. The orientation of the nanowires was achieved by selecting a substrate with a lattice structure matching that of the nanowire material to facilitate the epitaxial growth. These conditions result in an array of nanowire posts at predetermined positions, all vertically aligned with the same crystal growth orientation (Fig. 9.13). Similarly, GaN nanowire arrays have been synthesized epitaxially on (100)$LiAlO_2$ and (111)MgO single-crystal substrates. The lattice matching conditions enabled control over the crystallographic growth directions of the nanowires. For example, GaN nanowires on (100)$LiAlO_2$ substrates grow oriented along the [110] direction, whereas (111)MgO substrates result in the growth of GaN nanowires with an [001] orientation, due to the different lattice-matching constraints [9.143]. A similar structure could be obtained by the template-mediated electrochemical synthesis of nanowires (Sect. 9.1.1), particularly if anodic alumina with its parallel and ordered channels is used. The control over the location of the nucleation of nanowires in the electrochemical deposition is determined by the pore positions and the back-electrode geometry. The pore positions can be precisely controlled by imprint lithography [9.78]. By growing the template on a patterned conductive substrate that serves as a back-electrode [9.144–146], different materials can be deposited in the pores at different regions of the template.

9.2 Characterization and Physical Properties of Nanowires

In this section we review the structure and properties of nanowires and their interrelationship. The discovery and investigation of nanostructures were spurred on by advances in various characterization and microscopy techniques that enabled material characterization to take place at smaller and smaller length scales, reaching length scales down to individual atoms. For applications, characterizing the structural properties of nanowires is especially important, so that a reproducible relationship between their desired functionality and their geometrical and structural characteristics can be established. Due to the enhanced surface-to-volume ratio in nanowires, their properties may depend sensitively on their surface conditions and geometrical configurations. Even nanowires made of the same material may possess dissimilar properties due to differences in their crystal phase, crystalline size, surface conditions, and aspect ratios, which depend on the synthesis methods and conditions used in their preparation.

9.2.1 Structural Characterization

Structural and geometric factors play an important role in determining the various attributes of nanowires, such as their electrical, optical, and magnetic properties. Therefore, various novel tools have been developed and employed to obtain this important structural information at the nanoscale. At the micrometer scale, optical techniques are extensively used for imaging structural features. Since the sizes of nanowires are usually comparable to or, in most cases, much smaller than the wavelength of visible light, traditional optical microscopy techniques are usually limited when characterizing the morphology and surface features of nanowires. Therefore, electron microscopy techniques play a more dominant role at the nanoscale. Since electrons interact with matter more strongly than photons, electron microscopy is particularly sensitive relative to x-rays for the analysis of tiny samples.

In this section we review and give examples of how scanning electron microscopy, transmission electron microscopy, scanning probe spectroscopies, and diffraction techniques are used to characterize the structure of nanowires. To provide the necessary basis for developing reliable structure-property relations, multiple characterization tools are applied to the same samples.

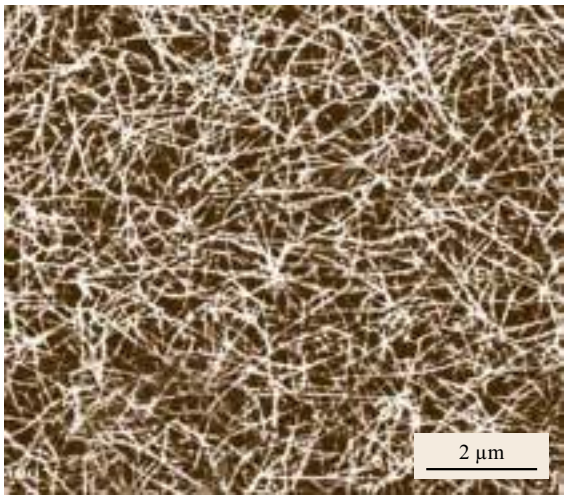


Fig. 9.14 SEM image of GaN nanowires in a mat arrangement synthesized by laser-assisted catalytic growth. The nanowires have diameters and lengths on the order of 10 nm and 10 μm respectively (after [9.39])

Scanning Electron Microscopy

SEM usually produces images down to length scales of ≈ 10 nm and provides valuable information regarding the structural arrangement, spatial distribution, wire density, and geometrical features of the nanowires. The examples of SEM micrographs shown in Figs. 9.1 and 9.3 indicate that structural features at the 10 nm to 10 μm length scales can be probed, providing information on the size, size distribution, shapes, spatial distributions, density, alignment, fill-factors, granularity, etc. As another example, Fig. 9.13a shows an SEM image of ZnO nanowire arrays grown on a sapphire substrate [9.110], which provides evidence for the non-uniform spatial distribution of the nanowires on the substrate, which was attained by patterning the catalyst film to define high-density growth regions and nanowire-free regions. Figure 9.13b, showing a higher magnification of the same system, indicates that these ZnO nanowires grow perpendicular to the substrate, are well aligned with approximately equal wire lengths, and have wire diameters in the range 20–150 nm. The SEM micrograph in Fig. 9.13c provides further information about the surface of the nanowires, showing it to be well faceted, forming a hexagonal cross-section, indicative of nanowire growth normal to the (0001) plane. Both the uniformity of the nanowires' size, their alignment perpendicular to the substrate, and their uniform growth direction, as suggested by the SEM data, are linked to the good epitaxial interface between the (0001) plane of the ZnO nanowire and the (110) plane of the sapphire substrate. (The crystal structures of ZnO and sapphire are essentially incommensurate, with the exception that the *a*-axis of

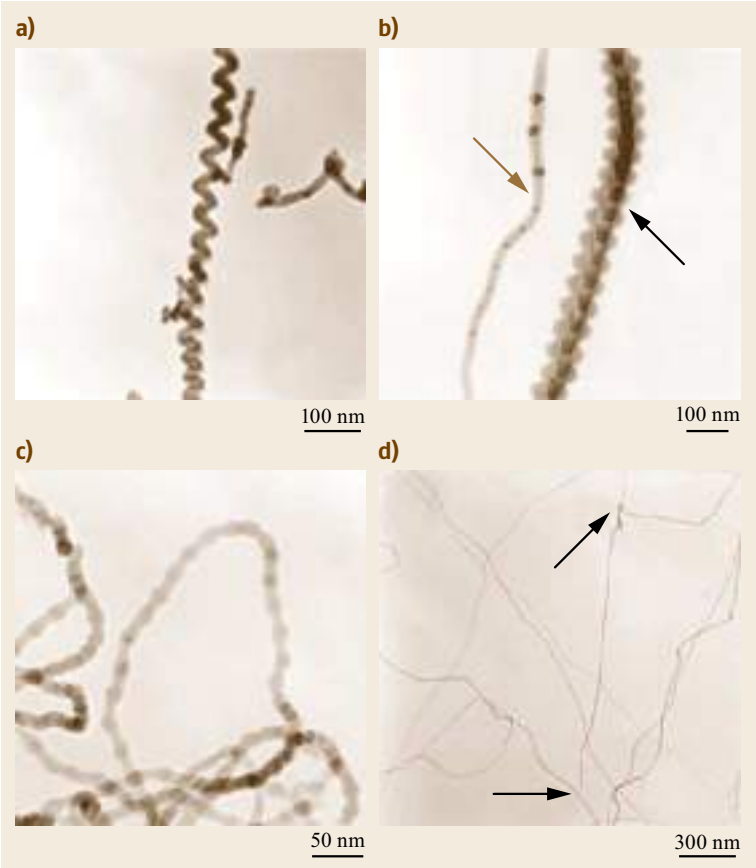


Fig. 9.15a–d TEM morphologies of four special forms of Si nanowires synthesized by the laser ablation of a Si powder target. **(a)** A spring-shaped Si nanowire; **(b)** fishbone-shaped (indicated by a *black arrow*) and frogs' egg-shaped (indicated by a *brown arrow*) Si nanowires; and **(c)** pearl-shaped nanowires. **(d)** shows polysites for the nucleation of silicon nanowires (indicated by *arrows*) (after [9.147])

ZnO and the *c*-axis of sapphire are related almost exactly by a factor of four, with a mismatch of less than 0.08% at room temperature [9.110].) The well-faceted nature of these nanowires has important implications for their lasing action (Sect. 9.3.2). The SEM image of GaN nanowires synthesized by a laser-assisted catalytic growth method [9.39] featured in Fig. 9.14 revals that random spatial orientation of the nanowire axes and a wide diameter distribution for these nanowires, in contrast to the ZnO wires in Fig. 9.13 and to arrays of well-aligned nanowires prepared by template-assisted growth (Fig. 9.3).

Transmission Electron Microscopy

TEM and high-resolution transmission electron microscopy (HRTEM) are powerful imaging tools for studying nanowires at the atomic scale, and they usually provide more detailed geometrical features than are seen in SEM images (SAED). As an example, the TEM images in Fig. 9.15a–c show four different morphologies for Si nanowires prepared by the laser ablation of a Si target: spring-shaped, fishbone-shaped, frogs' egg-shaped, and pearl-shaped nanowires [9.147].

TEM studies also yield information regarding the crystal structure, crystal quality, grain size, and crystal orientation of the nanowire axis. The crystal quality of nanowires is revealed from high-resolution TEM images with atomic resolution, along with selected area electron diffraction (SAED) patterns. For example, Fig. 9.16 shows a TEM image of one of the GaN nanowires from Fig. 9.14, indicating single crystallinity and showing (100) lattice planes, thus indicating the growth direction of the nanowire. This information is supplemented by the corresponding electron diffraction pattern in the upper right. A more comprehensive review of the application of TEM for growth orientation indexing and crystal defect characterization in nanowires is available elsewhere [9.148].

The high resolution of the TEM also permits the surface structures of the nanowires to be studied. In many cases, the nanowires are sheathed with native oxides, an amorphous oxide layer that forms during the growth process. This can be seen in Fig. 9.6b for silicon nanowires and in Fig. 9.17 for germanium nanowires [9.44], showing a mass-thickness contrast TEM image and a selected area electron diffraction pattern of a Ge nano-

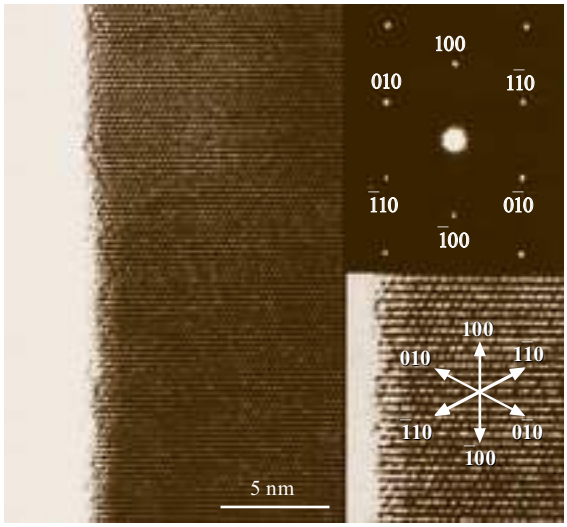


Fig. 9.16 Lattice-resolved high-resolution TEM image of one GaN nanowire (*left*) showing that (100) lattice planes are visible perpendicular to the wire axis. The electron diffraction pattern (*top right*) was recorded along the [001] zone axis. A lattice-resolved TEM image (*lower right*) highlights the continuity of the lattice up to the nanowire edge, where a thin native oxide layer is found. The directions of various crystallographic planes are indicated in the *lower right figure* (after [9.39])

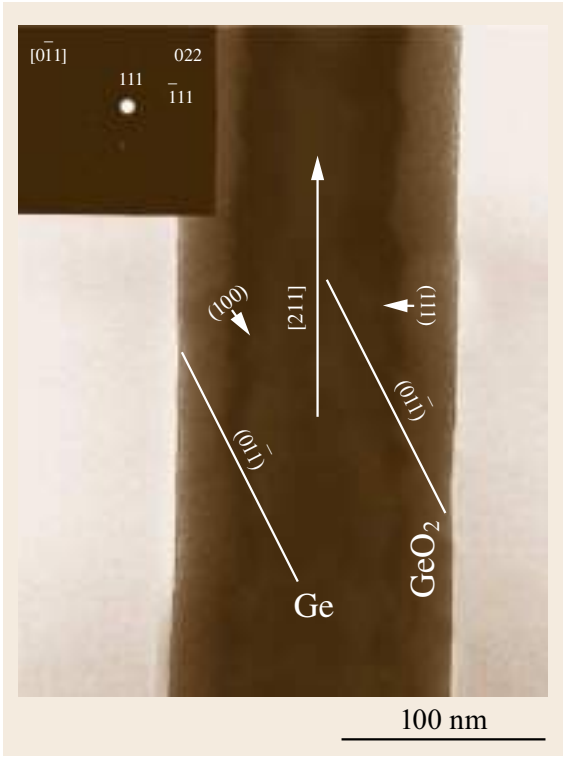


Fig. 9.17 A mass-thickness contrast TEM image of a Ge nanowire taken along the $(0\bar{1}1)$ zone axis and a selected area electron diffraction pattern (*upper left inset*) (after [9.44]). The Ge nanowires were synthesized by laser ablation of a mixture of Ge and GeO_2 powder. The core of the Ge nanowire is crystalline, while the surface GeO_2 is amorphous

wire. The main TEM image shows that these Ge nanowires possess an amorphous GeO_2 sheath with a crystalline Ge core that is oriented in the [211] direction.

Dynamical processes of the surface layer of nanowires can be studied in situ using an environmental TEM chamber, which allows TEM observations to be made while different gases are introduced or as the sample is heat treated at various temperatures, as illustrated in Fig. 9.18. The figure shows high-resolution TEM images of a Bi nanowire before and after a dynamic oxide removal process carried out within the environmental chamber of the TEM [9.149]. The amorphous bismuth-oxide layer coating the nanowire (Fig. 9.18a) is removed by exposure to hydrogen gas within the environmental chamber of the TEM, as indicated in Fig. 9.18b.

By coupling the powerful imaging capabilities of TEM with other characterization tools, such as an electron energy loss spectrometer (EELS) or an energy dispersive x-ray spectrometer (EDS) within the TEM instrument, additional properties of the nanowires can be probed with high spatial resolution. With the EELS technique, the energy and momentum of the incident and scattered electrons are measured in an inelastic electron scattering process to provide information on the energy and momentum of the excitations in the nanowire sample. Figure 9.19 shows the dependence on nanowire diameter of the electron energy loss spectra of Bi nanowires. The spectra were taken from the center of the nanowire, and the shift in the energy of the peak position (Fig. 9.19) indicates the effect of the nanowire diameter on the plasmon frequency in the nanowires. The results show that there are changes in the electronic structure of the Bi nanowires as the wire diameter decreases [9.150]. Such changes in electronic structure as a function of nanowire diameter are also observed in their transport (Sect. 9.2.2) and optical (Sect. 9.2.3) properties, and are related to quantum confinement effects.

EDS measures the energy and intensity distribution of x-rays generated by the impact of the electron beam on the surface of the sample. The elemental composition within the probed area can be determined to a high degree of precision. The technique was par-

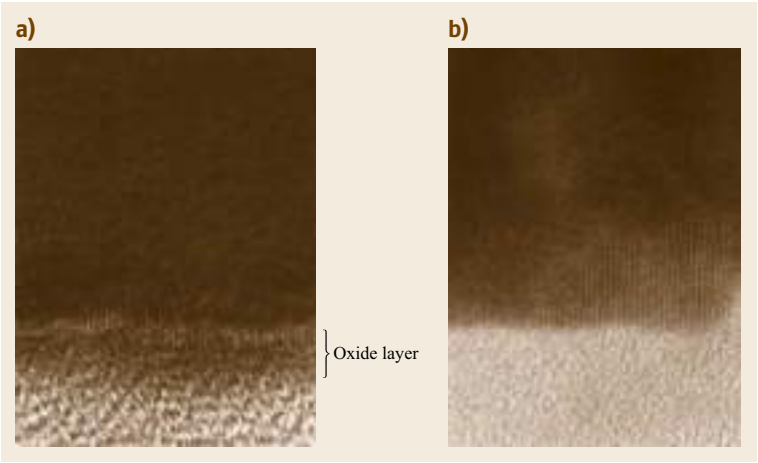


Fig. 9.18 High-resolution transmission electron microscope (HRTEM) image of a Bi nanowire **(a)** before and **(b)** after annealing in hydrogen gas at 130 °C for 6 h within the environmental chamber of the HRTEM instrument to remove the oxide surface layer (after [9.149])

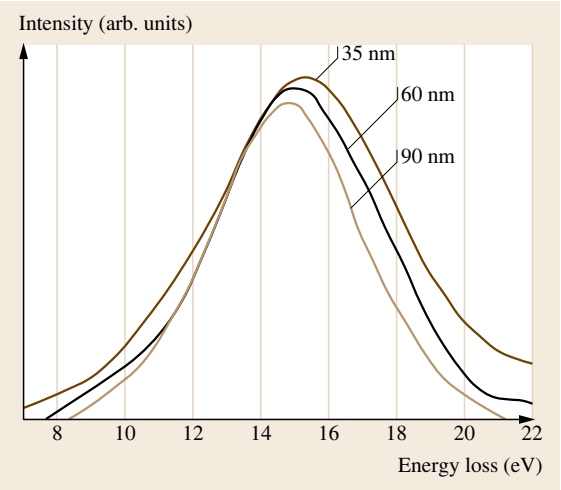


Fig. 9.19 Electron energy loss spectra (EELS) taken from the centers of bismuth nanowires with diameters of 35, 60 and 90 nm. The shift in the volume plasmon peaks is due to the effect of wire diameter on the electronic structure (after [9.150])

ticularly useful for the compositional characterization of superlattice nanowires [9.107] and core-shell nanowires [9.151] (Sect. 9.1.2).

Scanning Probe Techniques

Several scanning probe techniques, such as scanning tunneling microscopy (STM) [9.152], electric field gradient microscopy (EFM) [9.22], magnetic force microscopy (MFM) [9.49], and scanning thermal microscopy (SThM) [9.153], combined with atomic force microscopy (AFM), have been employed to study the structural, electronic, magnetic, and thermal properties of nanowires. A scanning tunneling microscope can be employed to reveal both topographical structural information, such as that illustrated in Fig. 9.20, as well as information on the local electronic density of states of a nanowire, when used in the STS (scanning tunneling spectroscopy) mode. Figure 9.20 shows STM height images (taken in the constant-current STM mode) of MoSe molecular wires deposited from a methanol or acetonitrile solution of $Li_2Mo_6Se_6$ onto Au substrates. The STM image of a single MoSe wire (Fig. 9.20a) exhibits a 0.45 nm lattice repeat distance in a MoSe molecular wire. When both STM and STS measurements are made on the same sample, the electronic and structural properties can be correlated. For example, joint STM/STS studies on Si nanowires [9.117] identified on single nanowires alternating segments grown along the [110] and [112] directions, and different I–V characteristics were measured for the [110] segments as compared with the [112] segments.

Magnetic force microscopy (MFM) has been employed to study magnetic polarization of magnetic nanowires embedded in an insulating template, such as an anodic alumina template. For example, Fig. 9.21a shows the topographic image of an anodic alumina template filled with Ni nanowires recorded in AFM mode, and Fig. 9.21b demonstrates the corresponding magnetic polarization of each nanowire in the template. This micrograph shows that an MFM probe can distinguish between spin-up and spin-down nanowires in the nanowire array, thereby providing a method for measuring magnetic dipolar interactions between wires [9.49].

X-ray Analysis

Other characterization techniques that are commonly used to study the crystal structures and chemical compositions of nanowires include x-ray diffraction and energy-dispersive x-ray spectroscopy (EDS). The peak positions in the x-ray diffraction pattern can used to determine the chemical composition and the crystal phase

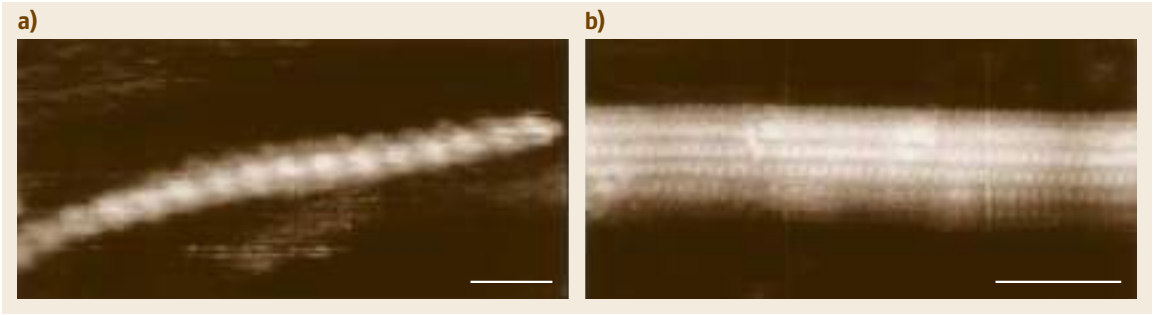


Fig. 9.20a,b STM height images, obtained in the constant-current mode, of MoSe chains deposited on an Au(111) substrate. **(a)** A single chain image, and **(b)** a MoSe wire bundle (after [9.152]). Scale bars: 1 nm

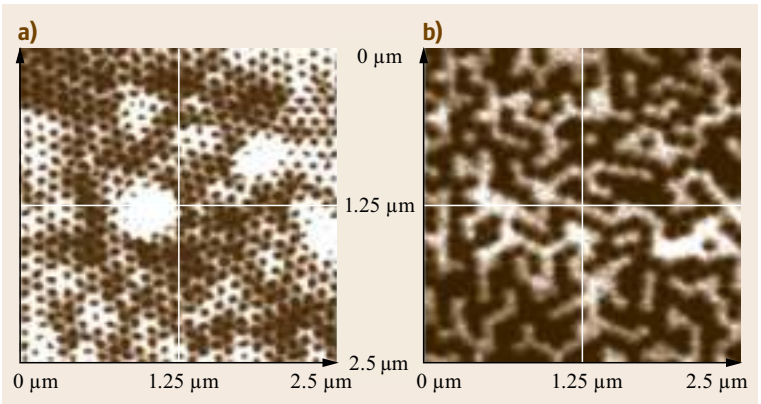


Fig. 9.21 **(a)** Topographic image of a highly ordered porous alumina template with a period of 100 nm filled with 35 nm-diameter nickel nanowires. **(b)** The corresponding MFM (magnetic force microscopy) image of the nano-magnet array, showing that the pillars are magnetized alternately *up* (*white*) and *down* (*black*) (after [9.49])

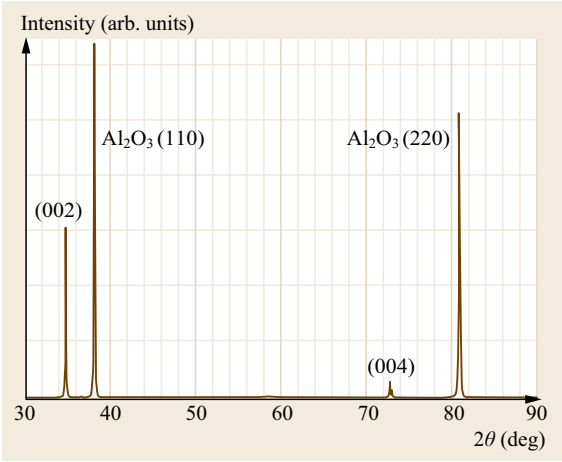


Fig. 9.22 X-ray diffraction pattern of aligned ZnO nanowires (Fig. 9.13) grown on a sapphire substrate. Only (00ℓ) diffraction peaks are observed for the nanowires, owing to their well-oriented growth orientation. Strong diffraction peaks for the sapphire substrate are found (after [9.110])

structure of the nanowires. For example, Fig. 9.2 shows that Bi nanowires have the same crystal structure and lattice constants as bulk bismuth. Both the x-ray diffraction pattern (XRD) for an array of aligned Bi nanowires (Fig. 9.2) and the SAED pattern for individual Bi nanowires [9.22] suggest that the nanowires have a common crystallographic alignment.

As another example of an XRD pattern for an array of aligned nanowires, Fig. 9.22 shows the x-ray diffraction pattern of the ZnO nanowires that are displayed in Fig. 9.13. Only (00ℓ) diffraction peaks are observed for these aligned ZnO nanowires, indicating that their preferred growth direction is [001] along the wire axis. Similarly, XRD was used to confirm the different growth directions of GaN nanowire arrays grown epitaxially on (100)$LiAlO_2$ and (111)MgO substrates [9.143].

EDS has been used to determine the chemical compositions and stoichiometry of compound nanowires or impurity contents in nanowires. However, the results from EDS analysis should be interpreted carefully to avoid systematic errors.

9.2.2 Mechanical Properties

Thermal Stability

Due to the large surface-to-volume ratio in nanowires and other nanoparticles, the thermal stability of na-

nowires is anticipated to differ significantly from that of the bulk material. Theoretical studies of materials in confined geometries show that the melting point of the material is reduced in nanostructures, as is the latent heat of fusion, and that large hysteresis can be observed in melting-freezing cycles. These phenomena have been studied experimentally in three types of nanowire systems: porous matrices impregnated with a plurality of nanowires, individual nanowires sheathed by a thin coating, and individual nanowires.

The melting-freezing of matrix-supported nanowires can be studied by differential scanning calorimetry (DSC), since large volumes of samples can thus be produced. *Huber* et al. investigated the melting of indium in porous silica glasses with mean pore diameters ranging from 6 to 141 nm [9.154]. The melting point of the pore-confined indium shows a linear dependence on inverse pore diameter, with a maximum melting point depression of 50 K. A similar shift was observed in the freezing temperature of pore-confined indium. The melting profile of the pore-confined indium in these samples is broader in temperature than for bulk indium, as expected for the heterogeneity in the pore diameter and in the indium crystal size aspect ratio within the samples.

Sheathed nanowires provide an opportunity to study the melting and recrystallization of individual nanowires. The shell layer surrounding the nanowire provides confinement to keep the liquid phase within the inner cylindrical volume. However, the shell-nanowire surface interaction should be taken into account when analyzing the phase transition thermodynamics and kinetics. *Wu* and *Yang* produced germanium nanowires coated with a thin (1−5 nm) graphite sheath, by pyrolysis of organic molecules over VLS-grown nanowires, and followed the melting and recrystallization of the germanium by variable-temperature TEM imaging [9.155]. The melting of the nanowires was followed by the disappearance of the electronic diffraction pattern. It was found that the nanowires began melting from their ends, with the melting front advancing towards the center of the nanowire as the temperature was increased. During the cool-down part of the cycle, the recrystallization of the nanowire occurred instantaneously following significant supercooling. The authors report both the largest melting point suppression recorded thus far for germanium ($\approx 300\,^{\circ}\mathrm{C}$), and a large melting-recrystallization hysteresis of up to $\approx 300\,^{\circ}\mathrm{C}$. Similarly, carbon nanotubes have been filled with various low-temperature metals [9.156]. A nanothermometer has been demonstrated using a 10 nm liquid gallium-filled carbon nanotube, showing an expansion coefficient that is linear in temperature and identical to the bulk value [9.157].

A different behavior was observed in free-standing copper nanowires [9.158]. In this system, there is little interaction between the nanowire surface and the surroundings, and the nanowire is not confined in its diameter, as in the case of the sheathed nanowires. Thermal treatment of the free-standing nanowires leads to their fragmentation into a linear array of metal spheres. Thinner nanowires are more vulnerable than thicker nanowires to the thermal treatment, showing constrictions and segmentation at lower temperatures. Analysis of the temperature response of the nanowires indicates that the nanowire segmentation is a result of the Rayleigh instability, starting with oscillatory perturbations of the nanowire diameter, leading to long cylindrical segments that become more separated and more spherical at higher temperatures. These observations indicate that annealing and melting are dominated by the surface diffusion of atoms on the entire surface of the nanowire (versus tip-initiated melting).

9.2.3 Transport Properties

The study of electrical transport properties of nanowires is important for nanowire characterization, electronic device applications, and the investigation of unusual transport phenomena arising from one-dimensional quantum effects. Important factors that determine the transport properties of nanowires include the wire diameter, material composition, surface conditions, crystal quality, and the crystallographic orientation along the wire axis for materials with anisotropic material parameters, such as the effective mass tensor, the Fermi surface, or the carrier mobility.

Electronic transport phenomena in low-dimensional systems can be roughly divided into two categories: ballistic transport and diffusive transport. Ballistic transport phenomena occur when the electrons can travel across the nanowire without any scattering. In this case, the conduction is mainly determined by the contacts between the nanowire and the external circuit, and the conductance is quantized into an integral number of universal conductance units $G_0 = 2\mathrm{e}^2/h$ [9.160, 161]. Ballistic transport phenomena are usually observed in very short quantum wires, such as those produced using mechanically-controlled break junctions (MCBJs) [9.162, 163] where the electron mean free path is much longer than the wire length and the conduction is a pure quantum phenomenon. To observe ballistic transport, the thermal energy must also obey the relation $k_\mathrm{B}T \ll \varepsilon_j - \varepsilon_{j-1}$, where $\varepsilon_j - \varepsilon_{j-1}$ is the energy separation between subband levels j and $j-1$. On the other hand, for nanowires with lengths much larger than the carrier mean free path, the electrons (or holes) undergo numerous scattering events when they travel

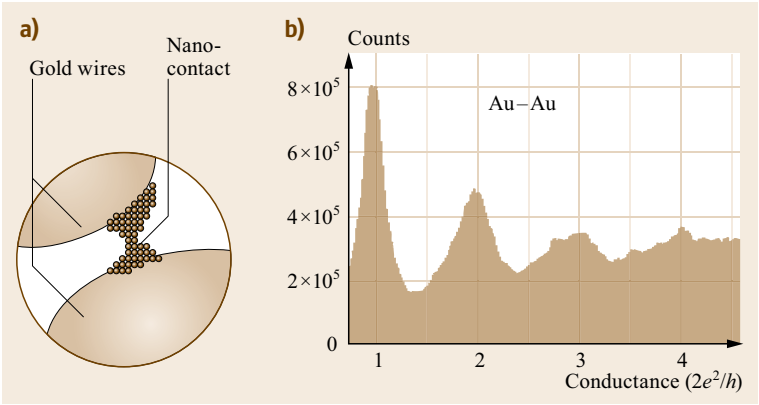


Fig. 9.23 **(a)** Schematic representation of the last stages of the contact breakage process (after [9.159]). **(b)** Histogram of conductance values built with 18 000 gold contact breakage experiments in air at room temperature, showing conductance peaks at integral values of G_0. In this experiment the gold electrodes approach and separate at 89 000 Å/s (after [9.159])

along the wire. In this case, the transport is in the diffusive regime, and the conduction is dominated by carrier scattering within the wires, due to phonons (lattice vibrations), boundary scattering, lattice and other structural defects, and impurity atoms.

Conductance Quantization in Metallic Nanowires

The ballistic transport of 1-D systems has been extensively studied since the discovery of quantized conductance in 1-D systems in 1988 [9.160, 161]. The phenomena of conductance quantization occurs when the diameter of the nanowire is comparable to the electron Fermi wavelength, which is on the order of 0.5 nm for most metals [9.164]. Most conductance quantization experiments up to the present were performed by bringing together and separating two metal electrodes. As the two metal electrodes are slowly separated, a nano-contact is formed before it breaks completely (Fig. 9.23a), and conductance in integral multiple values of G_0 is observed through these nano-contacts. Figure 9.23b shows the conductance histogram built with 18 000 contact breakage curves between two gold electrodes at room temperature [9.159], with the electrode separation up to ≈ 1.8 nm. The conductance quantization behavior is found to be independent of the contact material, and has been observed in various metals, such as Au [9.159], Ag, Na, Cu [9.165], and Hg [9.166]. For semimetals such as Bi, conductance quantization has also been observed for electrode separations as long as 100 nm at 4 K because of the long Fermi wavelength (≈ 26 nm) [9.164], indicating that the conductance quantization may be due to the existence of well-defined quantum states localized at a constriction instead of resulting from the atom rearrangement as the electrodes separate. Since conductance quantization is only observed in breaking contacts, or for very narrow and very short nanowires, most nanowires of practical interest (possessing lengths of several micrometers) lie in the diffusive transport regime, where the carrier scattering is significant and should be considered.

I–V Characterization of Semiconducting Nanowires

The electronic transport behavior of nanowires may be categorized based on the relative magnitudes of three length scales: carrier mean free path ℓ_W, the de Broglie wavelength of electrons λ_e, and the wire diameter d_W. For wire diameters much larger than the carrier mean free path ($d_W \gg \ell_W$), the nanowires exhibit transport properties similar to bulk materials, which are independent of the wire diameter, since the scattering due to the wire boundary is negligible compared to other scattering mechanisms. For wire diameters comparable to or smaller than the carrier mean free path ($d_W \approx \ell_W$ or $d_W < \ell_W$), but still much larger than the de Broglie wavelength of the electrons ($d_W \gg \lambda_e$), the transport in nanowires is in the classical finite size regime, where the band structure of the nanowire is still similar to that of bulk, while the scattering events at the wire boundary alter their transport behavior. For wire diameters comparable to the electronic wavelength $d_W \approx \lambda_e$, the electronic density of states is altered dramatically and discrete subbands are formed due to the quantum confinement effect at the wire boundary. In this regime, the transport properties are further influenced by the change in the band structure. Therefore, transport properties for nanowires in the classical finite size and quantum size regimes are highly diameter-dependent.

Researchers have investigated the transport properties of various semiconducting nanowires and have demonstrated their potential for diverse electronic devices, such as for p-n diodes [9.167, 168], field effect transistors [9.167], memory cells, and switches [9.169] (Sect. 9.3.1). So far, the nanowires studied in this context have usually been made from conventional semiconducting materials, such as group IV and III–V

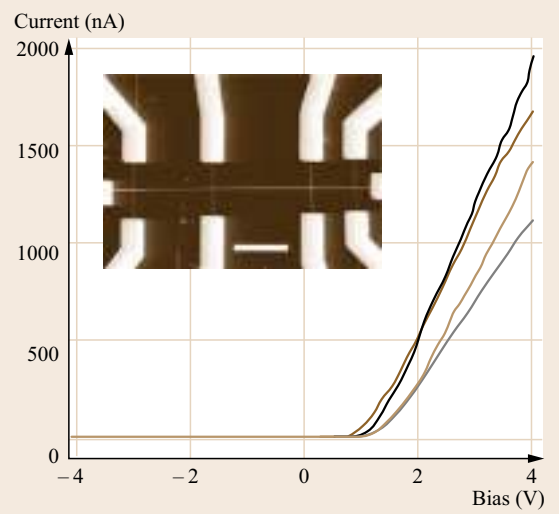


Fig. 9.24 *I–V* behavior for a 4(p) by 1(n) crossed p-Si/n-GaN junction array shown in the *inset*. The *four curves* represent the *I–V* response for each of the four junctions, showing similar current rectifying characteristics in each case. The length scale bar between the two middle junctions is 2 μm (after [9.167]). The p-Si and n-GaN nanowires are 10−25 and 10−30 nm in diameter respectively

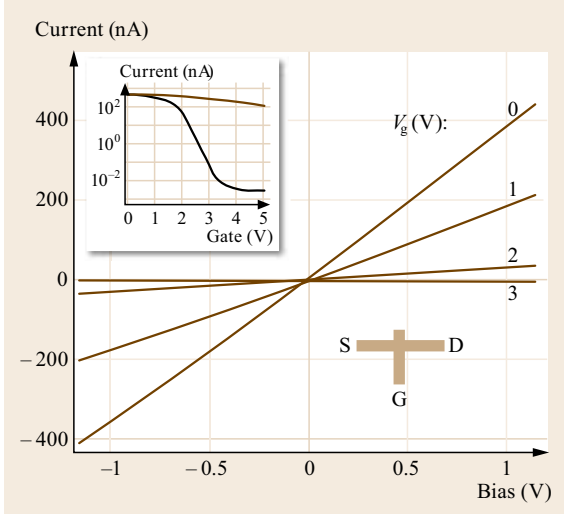


Fig. 9.25 Gate-dependent *I–V* characteristics of a crossed nanowire field-effect transistor (FET). The n-GaN nanowire is used as the nanogate, with the gate voltage indicated (0, 1, 2, and 3 V). The *inset* shows the current versus V_{gate} for a nanowire gate (*lower curve*) and for a global back-gate (*top curve*) when the bias voltage is set to 1 V (after [9.167])

compound semiconductors, via the VLS growth method (Sect. 9.1.2), and their nanowire properties have been compared to their well-established bulk properties. Interestingly, the physical principles for describing bulk semiconductor devices also hold for devices based on these semiconducting nanowires with wire diameters of tens of nanometers. For example, Fig. 9.24 shows the current-voltage (*I–V*) behavior of a four-by-one crossed p-Si/n-GaN junction array at room temperature [9.167]. The long horizontal wire in the figure is a p-Si nanowire (10−25 nm in diameter) and the four short vertical wires are n-GaN nanowires (10−30 nm in diameter). Each of the four nanoscale cross-points independently forms a p-n junction with current rectification behavior, as shown by the *I–V* curves in Fig. 9.24. The junction behavior (for example, the turn-on voltage) can be controlled by varying the oxide coating on these nanowires [9.167].

Huang et al. have demonstrated nanowire junction diodes with a high turn-on voltage (≈ 5 V) by increasing the oxide thickness at the junctions. The high turn-on voltage enables the use of the junction in a nanoscale field-effect transistor (FET), as shown in Fig. 9.25 [9.167] where *I–V* data for a p-Si nanowire are presented, for which the n-GaN nanowire with a thick oxide coating is used as a nanogate. By varying the nanogate voltage, the conductance of the p-Si nanowire can be changed by more than a factor of 10^5 (lower curve in the inset), whereas the conductance changes by only a factor of ten when a global back-gate is used (top curve in the inset of Fig. 9.25). This behavior may be due to the thin gate dielectric between the crossed nanowires and the better control of the local carrier density through a nanogate. Based on the gate-dependent *I–V* data from these p-Si nanowires, it is found that the mobility of the holes in the p-Si nanowires may be higher than that for bulk p-Si, although further investigation is required for complete understanding.

Because of the enhanced surface-to-volume ratios of nanowires, their transport behavior may be modified by changing their surface conditions. For example, researchers have found that by coating n-InP nanowires with a layer of redox molecules, such as cobalt phthalocyanine, the conductance of the InP nanowires may change by orders of magnitude upon altering the charge state of the redox molecules to provide bistable nanoscale switches [9.169]. The resistance (or conductance) of some nanowires (such as Pd nanowires) is also very sensitive to the presence of certain gases (e.g., H_2) [9.170, 171], and this property may be utilized for sensor applications to provide improved sensitivity compared to conventional sensors based on bulk material (Sect. 9.3.4).

Although it remains unclear how the size effect may influence the transport properties and device performance of semiconducting nanowires, many of the

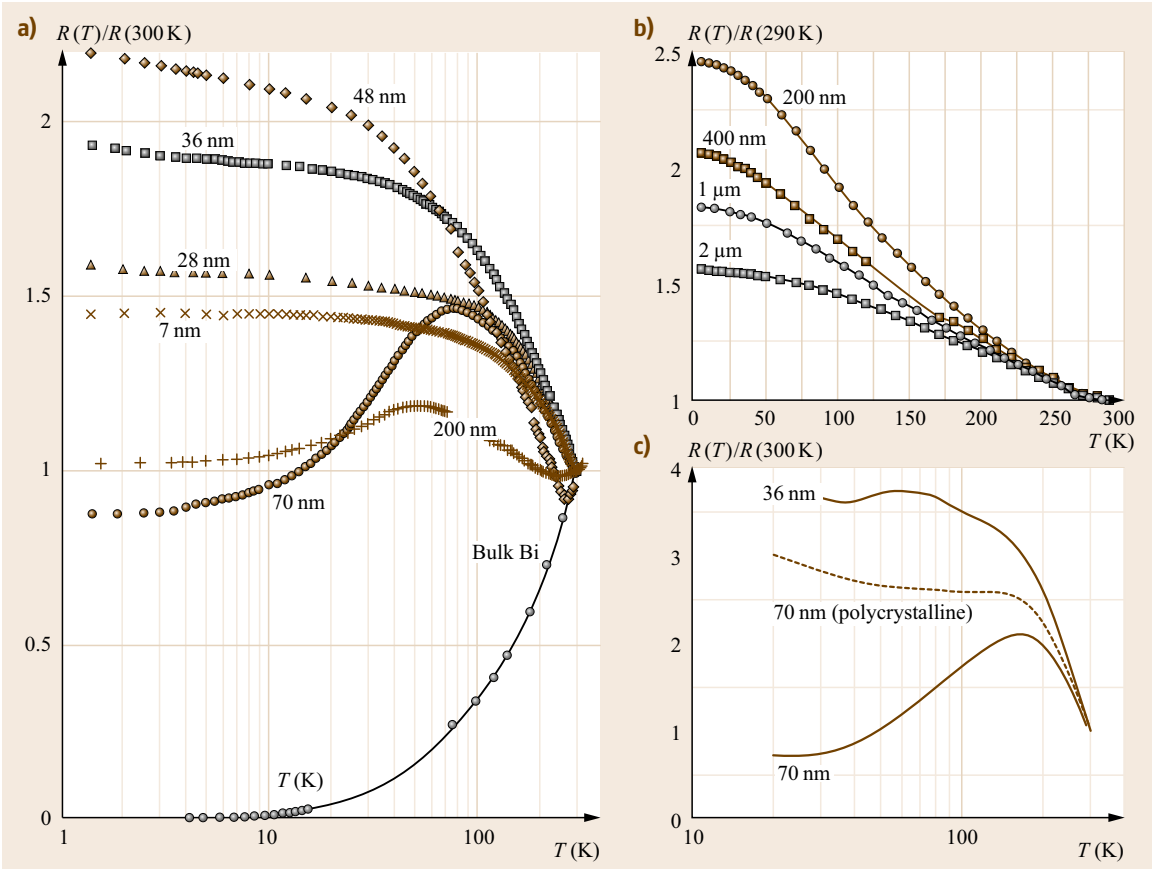


Fig. 9.26 **(a)** Measured temperature dependence of the resistance $R(T)$ normalized to the room temperature (300 K) resistance for bismuth nanowire arrays of various wire diameters d_W (after [9.17]). **(b)** $R(T)/R(290\,K)$ for bismuth wires of larger d_W and lower mobility (after [9.19]). **(c)** Calculated $R(T)/R(300\,K)$ of 36 and 70 nm bismuth nanowires. The *dashed curve* refers to a 70 nm polycrystalline wire with increased boundary scattering (after [9.93])

larger-diameter semiconducting nanowires are expected to be described by classical physics, since their quantization energies $\hbar^2/(2m_e d_W^2)$ are usually smaller than the thermal energy $k_B T$. By comparing the quantization energy with the thermal energy, the critical wire diameter below which quantum confinement effects become significant is estimated to be 1 nm for Si nanowires at room temperature, which is much smaller than the sizes of many of the semiconducting nanowires that have been investigated so far. By using material systems with much smaller effective carrier masses m_e (such as bismuth), the critical diameter for which such quantum effects can be observed increases, thereby facilitating the study of quantum confinement effects. It is for this reason that the bismuth nanowire system has been studied so extensively. Furthermore, since the crystal structure and lattice constants of bismuth nanowires are the same as for 3-D crystalline bismuth, it is possible to carry out detailed model calculations to guide and to interpret transport and optical experiments on bismuth nanowires. For these reasons, bismuth can be considered a model system for studying 1-D effects in nanowires.

Temperature-Dependent Resistance Measurements

Although nanowires with electronic properties similar to their bulk counterparts are promising for constructing nanodevices based on well-established knowledge of their bulk counterparts, it is expected that quantum size effects in nanowires will likely be utilized to generate new phenomena absent in bulk materials, and thus provide enhanced performance and novel functionality for certain applications. In this context, the transport properties of bismuth (Bi) nanowires have been extensively studied, both theoretically [9.172] and experimentally [9.17, 19, 93, 173–175] because of their promise for enhanced thermoelectric performance.

Transport studies of ferromagnetic nanowire arrays, such as Ni or Fe, have also received much attention because of their potential for high-density magnetic storage applications [9.176].

As mentioned in the previous section, the very small electron effective mass components and the long carrier mean free paths in Bi facilitate the study of quantum size effects in the transport properties of nanowires. Quantum size effects are expected to become significant in bismuth nanowires with diameters smaller than 50 nm [9.172], and the fabrication of crystalline nanowires with this diameter range is relatively easy.

Figure 9.26a shows the T dependence of the resistance $R(T)$ for Bi nanowires ($7\,\text{nm} \leq d_W < 200\,\text{nm}$) synthesized by vapor deposition and pressure injection [9.17], illustrating quantum effects in their temperature-dependent resistance. In Fig. 9.26a, the $R(T)$ behavior of Bi nanowires is dramatically different from that of bulk Bi, and is highly sensitive to the wire diameter. Interestingly, the $R(T)$ curves in Fig. 9.26a show a non-monotonic trend for large-diameter (70 and 200 nm) nanowires, although $R(T)$ becomes monotonic with T for small-diameter ($\leq 48\,\text{nm}$) nanowires. This dramatic change in the behavior of $R(T)$ as a function of d_W is attributed to a unique semimetal-semiconductor transition phenomena in Bi [9.93], induced by quantum size effects. Bi is a semimetal in bulk form, in which the T-point valence band overlaps with the L-point conduction band by 38 meV at 77 K. As the wire diameter decreases, the lowest conduction subband increases in energy and the highest valence subband decreases in energy. Model calculations predict that the band overlap should vanish in Bi nanowires (with their wire axes along the trigonal crystallographic direction) at a wire diameter $\approx 50\,\text{nm}$ [9.172].

The resistance of Bi nanowires is determined by two competing factors: the carrier density that increases with T, and the carrier mobility that decreases with T. The non-monotonic $R(T)$ for large-diameter Bi nanowires is due to a smaller carrier concentration variation at low temperature ($\leq 100\,\text{K}$) in semimetals, so that the electrical resistance is dominated by the mobility factor in this temperature range. Based on the semiclassical transport model and the established band structure of Bi nanowires, the calculated $R(T)/R(300\,\text{K})$ for 36 and 70 nm Bi nanowires is shown by the solid curves in Fig. 9.26c to illustrate different $R(T)$ trends for semiconducting and semimetallic nanowires respectively [9.93]. The curves in Fig. 9.26c exhibit trends consistent with experimental results. The condition for the semimetal-semiconductor transition in Bi nanowires can be experimentally determined, as shown by the measured resistance ratio $R(10\,\text{K})/R(100\,\text{K})$ of Bi nanowires as a function of wire diameter [9.177] in Fig. 9.27. The maximum in the resistance ratio $R(10\,\text{K})/R(100\,\text{K})$ at $d_W \approx 48\,\text{nm}$ indicates the wire diameter for the transition of Bi nanowires from a semimetallic phase to a semiconducting phase. The semimetal-semiconductor transition and the semiconducting phase in Bi nanowires are examples of new transport phenomena resulting from low dimensionality that are absent in the bulk 3-D phase. These phenomena further increase the possible benefits of exploiting properties of nanowires for desired applications.

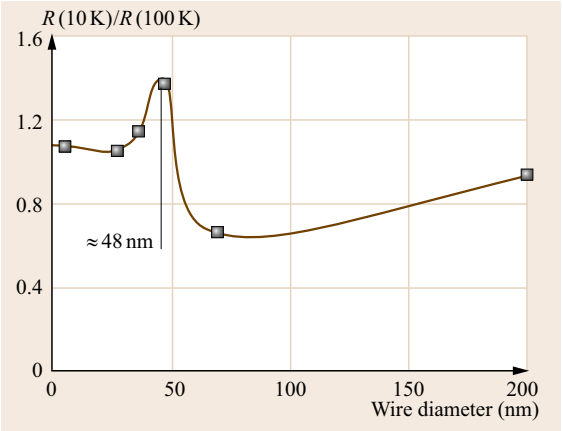


Fig. 9.27 Measured resistance ratio $R(10\,\text{K})/R(100\,\text{K})$ of a Bi nanowire array as a function of diameter. The *peak* indicates the transition from a semimetallic phase to a semiconducting phase as the wire diameter decreases (after [9.178])

It should be noted that good crystal quality is essential for observing the quantum size effect in nanowires, as shown by the $R(T)$ plots in Fig. 9.26a. Fig. 9.26b shows the normalized $R(T)$ measurements of Bi nanowires with larger diameters (200 nm–2 μm) prepared by electrochemical deposition [9.19]. These nanowires possess monotonic $R(T)$ behaviors, quite different from those of the corresponding nanowire diameters shown in Fig. 9.26a. The absence of the resistance maximum in Fig. 9.26b is due to the lower crystalline quality for nanowires prepared by electrochemical deposition, which tends to produce polycrystalline nanowires with a much lower carrier mobility. This monotonic $R(T)$ curve for semimetallic Bi nanowires with a higher defect level is also confirmed by theoretical calculations, as shown by the dashed curve in Fig. 9.26c for 70 nm wires with increased grain boundary scattering [9.179].

The theoretical model developed for Bi nanowires not only provides good agreement with experimental results, but it also plays an essential role in understanding the influence of the quantum size effect, the boundary scattering, and the crystal quality on their electrical

properties. While the electronic density of states may be significantly altered due to quantum confinement effects, various scattering mechanisms related to the transport properties of nanowires can be accounted for by Matthiessen's rule. Furthermore, the transport model has also been generalized to predict the transport properties of Te-doped Bi nanowires [9.93], Sb nanowires [9.180], and BiSb alloy nanowires [9.181], and good agreement between experiment and theory has also been obtained for these cases.

For nanowires with diameters comparable to the phase-breaking length, their transport properties may be further influenced by localization effects. It has been predicted that in disordered systems, the extended electronic wavefunctions become localized near defect sites, resulting in the trapping of carriers and giving rise to different transport behavior. Localization effects are also expected to be more pronounced as the dimensionality and sample size are reduced. Localization effects on the transport properties of nanowire systems have been studied on Bi nanowires [9.182] and Zn nanowires [9.63]. Figure 9.28 shows the measured $R(T)/R(300\,\mathrm{K})$ of Zn nanowires fabricated by vapor deposition in porous silica or alumina [9.63]. While 15 nm Zn nanowires exhibit an $R(T)$ behavior with a T^1 dependence as expected for a metallic wire, the $R(T)$ of 4 nm Zn nanowires exhibits a temperature dependence of $T^{-1/2}$ at low temperatures, consistent with 1-D localization theory. Thus, due to this localization effect, the use of nanowires with very small diameters for transport applications may be limited.

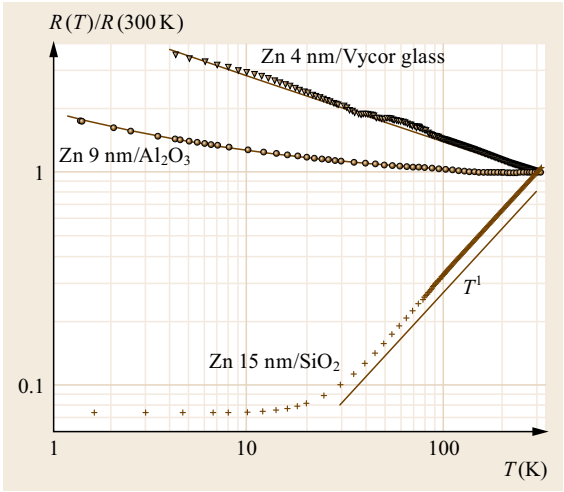


Fig. 9.28 Temperature dependence of the resistance of Zn nanowires synthesized by vapor deposition in various porous templates (after [9.63]). The data are given as *points*, the *full lines* are fits to a T^1 law for 15 nm diameter Zn nanowires in an SiO_2 template, denoted by Zn/SiO_2. Fits to a combined T^1 and $T^{-1/2}$ law were made for the smaller nanowire diameter composite samples denoted by Zn (9 nm)/Al_2O_3 and Zn 4 nm/Vycor glass

Magnetoresistance

Magnetoresistance (MR) measurements for characterizing nanowires yield a great deal of information about the effects of wire boundary, doping and annealing on electron scattering, and about localization effects in nanowires [9.175]. For example, at low fields the MR data show a quadratic dependence on the B field from which carrier mobility estimates can be made (Fig. 9.29 at low B field).

Figure 9.29 shows the longitudinal magnetoresistance ($\boldsymbol{B}$ parallel to the wire axis) for 65 and 109 nm Bi nanowire samples (before thermal annealing) at 2 K. The MR maxima in Fig. 9.29a are due to the classical size effect, where the wire boundary scattering is reduced as the cyclotron radius becomes smaller than the wire radius in the high field limit, resulting in a decrease in resistivity. This behavior is typical for the longitudinal MR of Bi nanowires in the diameter range of 45–200 nm [9.17, 174, 175, 183], and the peak position B_m moves to lower B field values as the wire diameter increases, as shown in Fig. 9.29c [9.183], where B_m varies linearly with $1/d_\mathrm{W}$. The condition for the occurrence of B_m is approximately given by $B_\mathrm{m} \approx 2c\hbar k_\mathrm{F}/(ed_\mathrm{W})$, where k_F is the wave vector at the Fermi energy. The peak position B_m is found to increase linearly with increasing temperature in the range of 2–100 K, as shown in Fig. 9.29b [9.183]. As T is increased, phonon scattering becomes increasingly important, and therefore a higher magnetic field is required to reduce the resistivity associated with boundary scattering sufficiently to change the sign of the MR. Likewise, increasing the grain boundary scattering is also expected to increase the value of B_m at a given T and wire diameter.

The presence of the peak in the longitudinal MR of nanowires requires a high crystal quality with long carrier mean free paths along the nanowire axis, so that most scattering events occur at the wire boundary instead of at a grain boundary, at impurity sites, or at defect sites within the nanowire. *Liu* et al. have investigated the MR of 400 nm Bi nanowires synthesized by electrochemical deposition [9.84], and no peak in the longitudinal MR is observed. The absence of a magnetoresistance peak may be attributed to a higher defect level in the nanowires produced electrochemically and to a large wire diameter, much longer than the carrier mean free path. The negative MR observed for the Bi nanowire arrays above B_m (Fig. 9.29) shows that wire boundary scattering is a dominant scattering process for the longitudinal magnetoresistance, thereby estab-

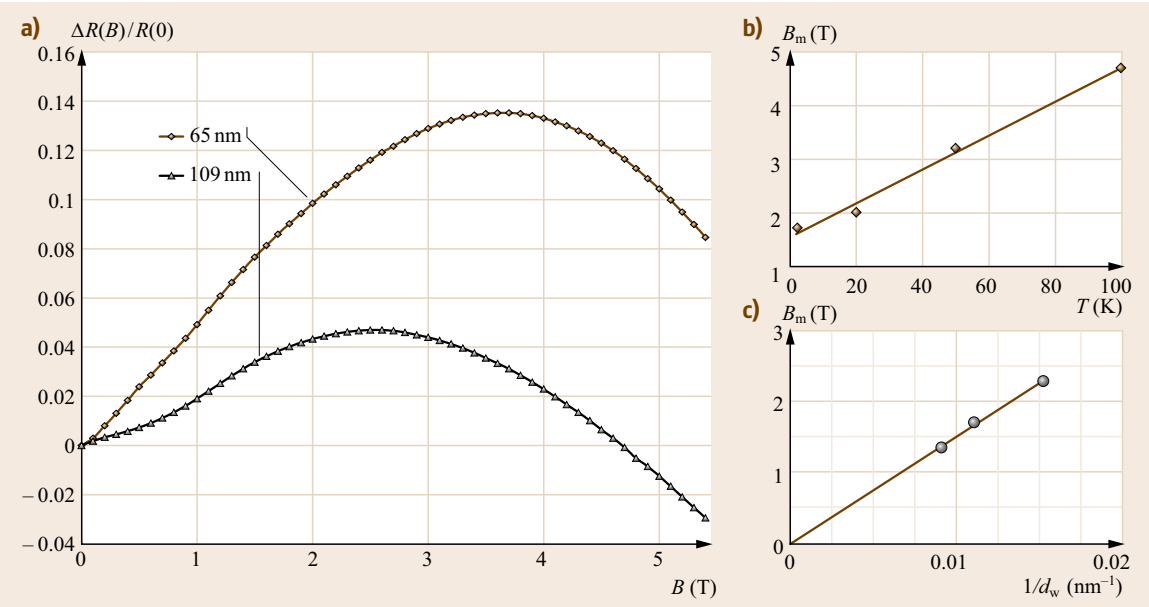


Fig. 9.29 **(a)** Longitudinal magnetoresistance, $\Delta R(B)/R(0)$, at 2 K as a function of B for Bi nanowire arrays with diameters of 65 and 109 nm before thermal annealing. **(b)** The peak position B_m as a function of temperature for the 109 nm diameter Bi nanowire array after thermal annealing. **(c)** The peak position B_m of the longitudinal MR (after thermal annealing) at 2 K as a function of $1/d_W$, the reciprocal of the nanowire diameter (after [9.183])

lishing that the mean free path is larger than the wire diameter.

In addition to the longitudinal magnetoresistance measurements, transverse magnetoresistance measurements (***B*** perpendicular to the wire axis) have also been performed on Bi nanowire array samples [9.17, 175, 183], where a monotonically increasing B^2 dependence over the entire range $0\,\mathrm{T} \leq B \leq 5.5\,\mathrm{T}$ is found for all Bi nanowires studied. This is expected, since the wire boundary scattering cannot be reduced by a magnetic field perpendicular to the wire axis. The transverse magnetoresistance is also found to be always larger than the longitudinal magnetoresistance in nanowire arrays.

The Shubnikov–de Haas (SdH) quantum oscillatory effect, which results from the passage of the quantized Landau levels through the Fermi energy as the field strength varies, should, in principle, provide the most direct measurement of the Fermi energy and carrier density. For example, *Heremans* and *Thrush* have demonstrated that SdH oscillations can be observed in Bi nanowire samples with diameters down to 200 nm [9.184], and they have demonstrated that Te doping can be used to raise the Fermi energy in Bi nanowires. Such information on the Fermi energy is important because, for certain applications based on nanowires, it is necessary to place the Fermi energy near a subband edge where the density of states has a sharp feature. However, due to the unusual 1-D geometry of nanowires, other characterization techniques that are commonly used in bulk materials to determine the Fermi energy and the carrier concentration (such as Hall effect measurement) cannot be applied to nanowire systems.

Thermal Conductivity of Nanowires

The thermal conductivity in nanowires can be significantly reduced as the dimensions of the nanowire become comparable to the phonon mean free path and the phonon wavelength [9.185]. The confinement in the nanowire may modify the phononphonon density of states, while defects in the nanowire enhance scattering. The most dominant suppression of thermal conductivity is achieved when multiple mechanisms operate to impede phonon transport at various length scales: atomic point-defects scatter high-frequency phonons, and interfaces and grain boundaries scatter primarily low-to-mid frequency phonons. Silicon nanowires with smooth and rough surfaces [9.186–188], segmented Si/SiGe nanowires [9.189], and indium gallium zinc oxide nanowires [9.190] have been selected as testbeds for fundamental studies. The enhancement of a new mechanism of phonon scattering in nanowires is evident by the lower thermal conductivity and by the change in its temperature dependence [9.185].

Measurements of the thermal conductivity of nanowires have been carried out on nanowire arrays embed-

ded in a matrix and individual free-standing nanowires. The former provides data of relevance to device performance, while only the latter can provide reliable data that are not distorted by thermal transport via the matrix or substrate and by sample polydispersity. Single-nanowire thermal measurements are very challenging and are now possible thanks to technological developments in the micro- and nanofabrication of miniature thermal sensors [9.191], and the use of nanometer-size thermal scanning probes [9.153, 192, 193]. In a typical fabricated device, shown in Fig. 9.30, a nanowire is suspended across a gap between two SiN_x membranes supported by long beams. On the membranes, metal lines form heating elements, resistive temperature sensors, and electrodes. Using a heat flow model, the thermal conductivity $\kappa(T)$ of the individual suspended nanowires can be extracted from the sensor data [9.194].

The experiments show that the thermal conductivity of small-diameter homogeneous nanowires may be more than one order of magnitude smaller than in the bulk, mainly due to the effect of strong phonon boundary scattering [9.195]. Phonon confinement effects may eventually become important in nanowires with even smaller diameters. The most extensive experimental thermal conductivity measurements have been done on Si nanowires [9.186], where $\kappa(T)$ measurements have been made on nanowires in the diameter range of 22−115 nm. As d_W decreases a large decrease is observed in the peak of $\kappa(T)$, associated with Umklapp processes, indicating a growing importance of boundary scattering and a corresponding decreasing importance of phonon-phonon scattering (Fig. 9.31). At the smallest wire diameter of 22 nm, a linear $\kappa(T)$ dependence is found experimentally, consistent with a linear T dependence of the specific heat for a 1-D system, and a temperature-independent mean free path and sound velocity. Introduction of scattering elements through alloying and composition modulation, such as in Si/SiGe superlattice nanowires, pushed the thermal conductivity values further down [9.189]. Remarkably, surface roughness, observed in some Si nanowires fabricated by the MACE method (Sect. 9.1.3), also reduces $\kappa(T)$ by an order of magnitude, such that 50 nm diameter rough Si nanowires have thermal conductivity values lower than 2 $Wm^{-1}K^{-1}$ at room temperature [9.188].

Model calculations for $\kappa(T)$ based on a radiative heat transfer model have been carried out for Si nanowires [9.196–198]. These results show that the predicted $\kappa(T)$ behavior for Si nanowires is similar to that observed experimentally in the range of 37 nm $\leq d_W \leq$ 115 nm regarding both the functional form of $\kappa(T)$ and the magnitude of the relative decrease in the maximum thermal conductivity as a function of

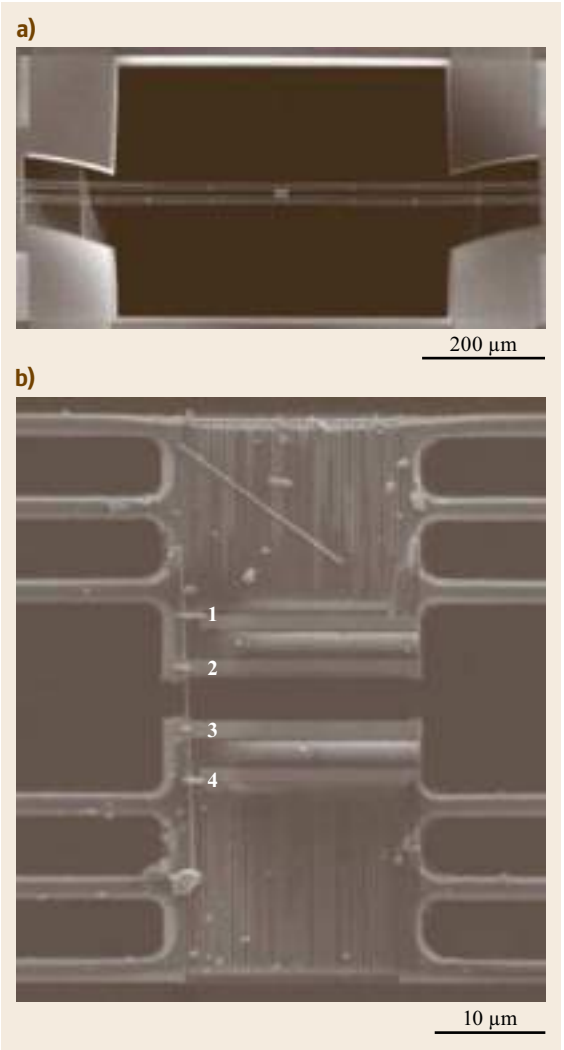


Fig. 9.30a,b Microfabricated device for the measurement of the thermal conductivity and thermoelectric properties of individual nanowires. **(a)** Low magnification view of the full length of the microscale device. **(b)** High magnification view of the probed nanowire and metalization lines on the membranes. IOP Publishing. Reproduced with permission. All right reserved (after [9.191])

d_W. However, the model calculations often overestimate κ_{max}, with agreement improved by using the full phonon dispersion relations. Furthermore, the model calculations (Fig. 9.31) do not reproduce the experimentally observed linear T dependence for the 22 nm nanowires, but rather predict a 3-D behavior for both the density of states and the specific heat in 22 nm nanowires [9.196, 199]. *Kazan* et al. resolved the discrepancies by considering also optical phonon decay and an angle-dependent specularity function for the boundary scattering, and

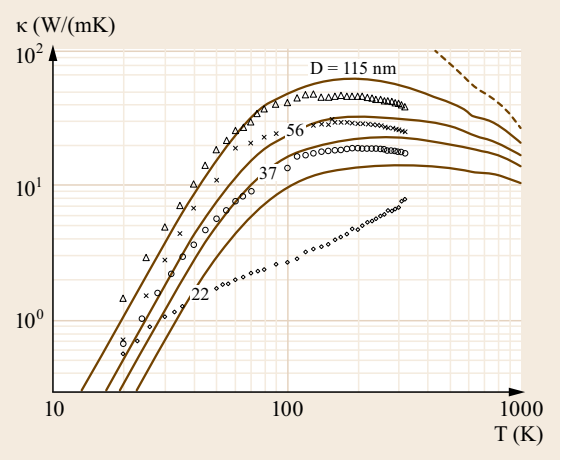


Fig. 9.31 Predicted thermal conductivities of Si nanowires of various diameters (*solid lines*) and the corresponding experimental data (symbols, from [9.186]) (after [9.196])

they were able to model both the diameter and surface roughness effects [9.200].

Thermal conductance measurements on GaAs nanowires below 6 K show a power law dependence, but the T dependence becomes somewhat less pronounced below ≈ 2.5 K [9.192]. This deviation from the power law temperature dependence led to a more detailed study of the quantum limit for the thermal conductance. To carry out more detailed experiments, a mesoscopic phonon resonator and waveguide device were constructed that included four ≈ 200 nm wide and 60 nm thick silicon nitride nanowire-like nano-constrictions (Fig. 9.32a), and this was used to establish the quantized thermal conductance limit of $g_0 = \pi^2 k_B^2 T/(3h)$ (Fig. 9.32b) for ballistic phonon transport [9.201, 202].

For temperatures above 0.8 K, the thermal conductance in Fig. 9.32b follows a T^3 law, but as T is further reduced, a transition to a linear T dependence is observed, consistent with a phonon mean free path of $\approx 1\,\mu$m, and a thermal conductance value approaching $16g_0$, corresponding to four massless phonon modes per channel and four channels in their phonon waveguide structure (Fig. 9.32a). Ballistic phonon transport occurs when the thermal phonon wavelength (380 nm for the experimental structure) is somewhat greater than the width of the phonon waveguide at the waveguide constriction.

Thermoelectric Properties

Nanowires are predicted to hold great promise for thermoelectric applications [9.172, 203], due to their unique band structure compared to their bulk counterparts and the expected reduction in thermal conductivity associated with enhanced boundary scattering (see Sect. 9.2.3, *Thermal Conductivity of Nanowires* above). Due to the sharp density of states function at the 1-D subband edges (where the van Hove singularities occur), nanowires are expected to exhibit enhanced Seebeck coefficients compared to their bulk counterparts. Since the Seebeck coefficient measurement is intrinsically independent of the number of nanowires contributing to the signal, the measurements on nanowire arrays of uniform wire diameter are, in principle, as accurate as single-wire measurements. Individual nanowire measurements, on the other hand, allow for the simultaneous determination of the Seebeck coefficient S, the thermal conductivity κ, and the electrical conductivity σ (and thus, the power factor $S^2\sigma$ and the figure-of-merit $ZT = S^2\sigma T\kappa^{-1}$), as well as electrostatic gating of semiconductor nanowires for detailed stud-

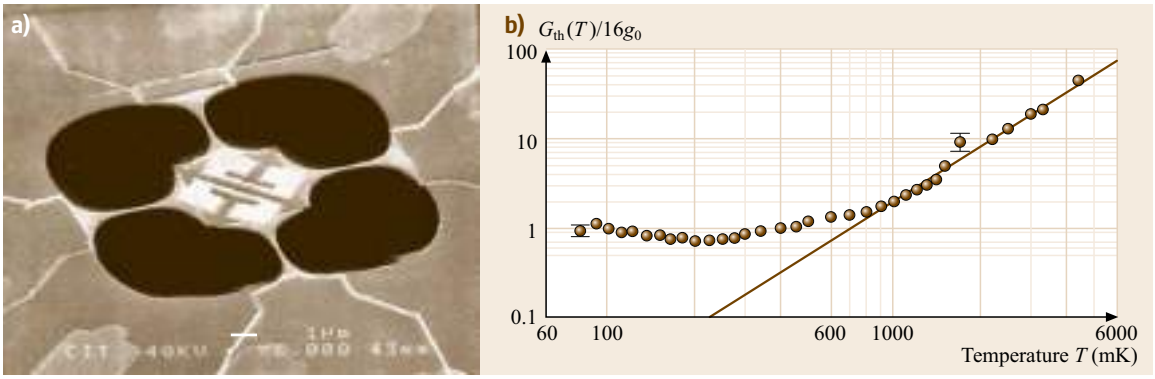


Fig. 9.32 **(a)** Suspended mesoscopic phonon device used to measure ballistic phonon transport. The device consists of a $4\times4\,\mu m^2$ *phonon cavity* (*center*) connected to four Si_3N_4 beams, 60 nm thick and less than 200 nm wide at the narrowest point. The two bright C-shaped objects on the phonon cavity are thin-film heating and sensing Cr/Au resistors, whereas the *dark regions* are empty space. **(b)** Log–log plot of the temperature dependence of the thermal conductance G_0 of the structure in **(a)** normalized to $16g_0$ (see text) (after [9.201])

ies of carrier concentration effects on thermoelectric properties.

Many nanowire systems have demonstrated enhanced Seebeck coefficient values as the diameter of the nanowire is reduced. Measured Seebeck coefficients $S(T)$ may also display a temperature dependence different than the bulk, indicative of carrier localization, enhanced scattering, or quantum confinement effects. In the well-known $Bi_{1-x}Sb_x$ thermoelectric material systems, $S(T)$ values are enhanced in arrays of nanowires with diameters of 40 and 65 nm (but not 200 nm), and by alloying with isoelectronic Sb [9.179]. Thermopower enhancement is attributed to the semimetal-semiconductor transition induced by quantum confinement and to Sb alloying effects in $Bi_{1-x}Sb_x$ nanowires [9.204]. *Heremans* et al. have observed a substantial increase in the thermopower of Bi in tiny nanowires in Bi(15 nm)/silica and Bi(9 nm)/alumina disordered nanocomposites [9.63]. The enhancement is due to the sharp density of states near the Fermi energy in a 1-D system. Measurements on individual or a small number of Si nanowires also showed dramatic enhancement of the Seebeck coefficient near and below room temperature, showing that non-traditional thermoelectric materials can become technologically relevant in nanowire form [9.187, 188]. This is particularly important in light of the poor mechanical properties, low abundance, and toxicity of some of the materials traditionally used for thermoelectrics. Since the Seebeck coefficient is a function of the carrier density, modulation of $S(T, V_G)$ is observed in individual electrostatically-gated nanowires (Sb_2Te_3 [9.205]; PbSe [9.206]; InAs [9.207, 208]), offering an additional strategy for tuning the thermoelectric properties. At cryogenic temperatures, the $S(V_G)$ function shows distinct peaks. The interpretation of these peaks is still under debate; the features may originate from the quantization in the band structure (subband formation) or from nanowire nonuniformities.

For applications in cooling and power generation, enhancing only the thermopower is not sufficient. In fact, in all the nanowire examples quoted above, the enhanced thermopower is accompanied by a significant reduction of the electrical conductivity, detrimental to the power factor $S^2\sigma$. Theoretical work demonstrates that quantum confinement in nanowires effectively reduces that density-of-states for electrons and holes in two-band semiconductors. As a result, as the nanowire diameter decreases, a decrease in the power factor is expected (Fig. 9.33) [9.209, 210]. This trend is reversed in small-diameter nanowires where electrons in the single lowest subband dominate transport ($d < 37$ nm for InSb, see Fig. 9.33). At this strongly confined regime, the power factor goes as d^{-2}; however, electron scattering and localization effects put a cap on the benefits of size reduction. The notable breakthroughs in low-dimensional systems for thermoelectric applications are related to the high figure-of-merit values that were achieved by drastically reducing the thermal conductivity while maintaining the power factor only slightly below bulk values. In smooth- and rough-edge silicon nanowires, reduction of the thermal conductivity by two orders of magnitude led to ZT values of up to 0.6 at 300 K (compare to 0.01 for bulk silicon) [9.187, 188]. Further improvements will likely involve materials with more complex band structures [9.204, 211] and more complex geometries [9.212].

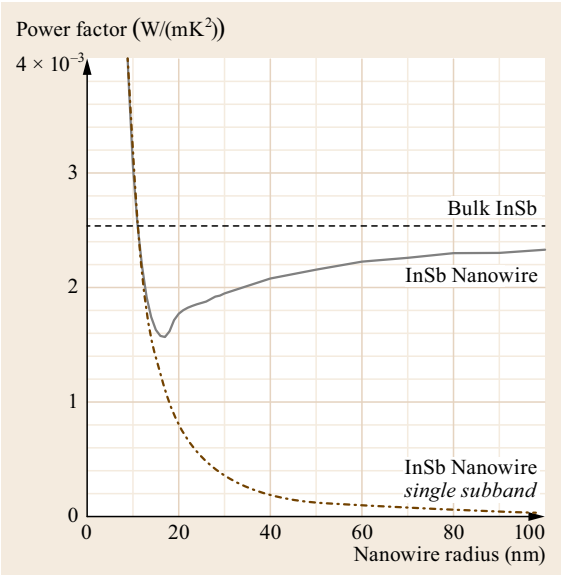


Fig. 9.33 Calculated thermoelectric power factor for n-type InSb nanowires (*solid line*). The power factor surpasses the bulk value (*dashed line*) only for very narrow nanowires, where electrons in one subband dominate transport. The *dash-dotted* line illustrates the size dependence of the power factor in a model considering just one subband (adapted from [9.209])

In 3-D systems, the electronic contribution to κ is proportional to σ in accordance with the Wiedemann–Franz law, and normally materials with high S have a low σ. Hence an increase in the electrical conductivity (for example by electron donor doping) results in an adverse variation in both the Seebeck coefficient (decreasing) and the thermal conductivity (increasing). These two trade-offs set the upper limit for increasing ZT in bulk materials, with the maximum ZT remaining ≈ 1 at room temperature for the 1960–1995 time frame.

The high electronic density of states in quantum-confined structures is proposed as a promising possibil-

ity to bypass the Seebeck/electrical conductivity tradeoff and to control each thermoelectric-related variable independently, thereby allowing for increased electrical conductivity, relatively low thermal conductivity, and a large Seebeck coefficient simultaneously [9.213]. The quantum size effect in nanowires can be combined with other parameters to tailor the band structure and electronic transport behavior (for instance, Sb alloying in Bi) to further optimize ZT. For example, Fig. 9.34 shows the predicted ZT for p-type $Bi_{1-x}Sb_x$ alloy nanowires as a function of wire diameter and Sb content x [9.204]. The occurrence of a local ZT maximum in the vicinity of $x \approx 0.13$ and $d_W \approx 45$ nm is due to the coalescence of the energies of ten valence bandextrema in the nanowire and the resulting unusually high density of states for holes, which is a phenomenon absent in bulk $Bi_{1-x}Sb_x$ alloys. For nanowires with very small diameters, it is speculated that localization effects will eventually limit the enhancement of ZT. However, in bismuth nanowires, localization effects are not significant for wires with diameters larger than 9 nm [9.63]. In addition to 1-D nanowires, ZT values as high as ≈ 2 have also been experimentally demonstrated in macroscopic samples containing PbSe quantum dots (0-D) [9.214] and stacked two-dimensional (2-D) films [9.195].

Quantum Wire Superlattices

The studies on superlattice nanowires, which possess a periodic modulation in their material composition along the wire axis, have attracted much attention recently because of their promise in various applications, such as thermoelectrics [9.107, 215], nanobarcodes (Sect. 9.3.2) [9.216], nanolasers (Sect. 9.3.2) [9.111], one-dimensional waveguides, and resonant tunneling diodes [9.113, 217]. Figure 9.35a shows a schematic structure of a superlattice nanowire consisting of interlaced quantum dots of two different materials, as denoted by A and B. Various techniques have been developed to synthesize superlattice nanowire structures with different interface conditions, as mentioned in Sects. 9.1.1 and 9.1.2.

In this superlattice nanowire structure, the electronic transport along the wire axis is made possible by the tunneling between adjacent quantum dots, while the uniqueness of each quantum dot and its 0-D characteristic behavior is maintained by the energy difference of the conduction or valence bands between quantum dots of different materials (Fig. 9.35b), which provides some amount of quantum confinement. *Björk* et al. have observed interesting nonlinear I–V characteristics with a negative differential resistance in one-dimensional heterogeneous structures made of InAs and InP, where InP serves as the potential barrier [9.113, 217]. The nonlinear I–V behavior is associated with the double barrier resonant tunneling process in one-dimensional structures, demonstrating that transport phenomena occur in superlattice nanowires via tunneling, and the possibility of controlling the electronic band structure of the superlattice nanowires by carefully selecting the constituent materials. These structures are especially attractive for thermoelectric applications, because the interfaces between the nanodots can reduce the lattice thermal conductivity by blocking the phonon conduction along the wire axis, while electrical conduction may be sustained and even benefit from the unusual electronic band structures due to the periodic potential perturbation.

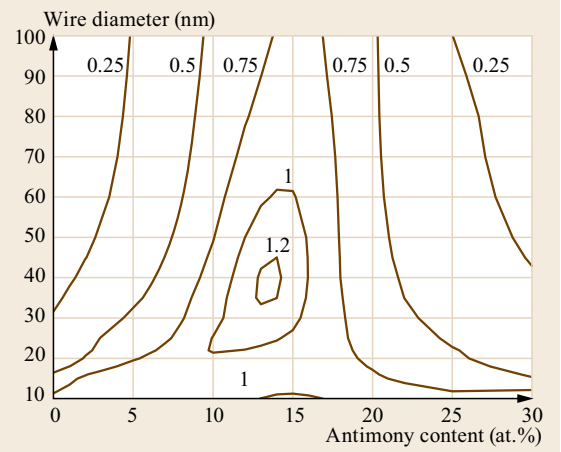


Fig. 9.34 Contour plot of optimal ZT values calculated for p-type $Bi_{1-x}Sb_x$ nanowires versus wire diameter and antimony concentration calculated at 77 K (after [9.204])

9.2.4 Optical Properties

Optical methods provide a convenient tool for measuring the electronic structures of nanowires, since optical

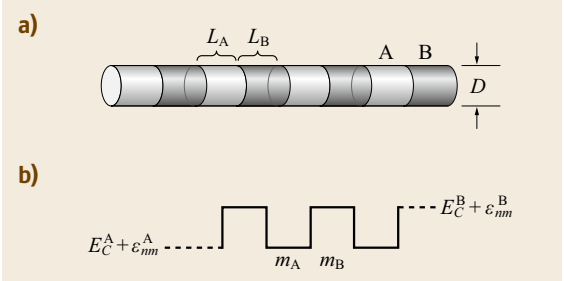


Fig. 9.35 (a) Schematic diagram of superlattice (segmented) nanowires consisting of interlaced nanodots A and B of the indicated length and wire diameter. **(b)** Schematic potential profile of the subbands in the superlattice nanowire (after [9.215])

measurements require minimal sample preparation (for example, contacts are not required) and are sensitive to quantum effects. A variety of optical techniques have shown that the properties of nanowires are different from those of their bulk counterparts. For example, optical spectra of 1-D systems often show intense features at specific energies near singularities in the joint density of states that are formed under strong quantum confinement conditions.

Although optical methods are an extremely important tool for characterizing nanowires, the interpretation of these measurements for nanowires is not always straightforward. The wavelength of light used to probe the sample is usually smaller than the wire length, but larger than the wire diameter. Hence, the probe light used in an optical measurement cannot be focused solely onto the wire, and the wire and the substrate on which the wire rests (or host material, if the wires are embedded in a template) are probed simultaneously.

In this section we discuss the determination of the dielectric function for nanowires from optical measurements. We then discuss various optical techniques with appropriate examples that sensitively differentiate nanowire properties from those also found in the parent bulk material. Finally, phonon confinement effects are reviewed.

The Dielectric Function

The observable optical properties of materials can be related to their complex dielectric function [9.218, 219]. Since nanowire diameters are typically smaller than the wavelength of light, photons at visible or infrared wavelengths *see* a dielectric function for the composite nanowire array/substrate system that is different from that of the nanowire itself. Effective medium theories [9.220, 221] can be applied to model the nanowire and substrate as one continuous composite with a single complex dielectric function $(\epsilon_1 + \mathrm{i}\epsilon_2)$, where the real and imaginary parts of the dielectric function ϵ_1 and ϵ_2 are related to the index of refraction (n) and the absorption coefficient (K) by the relation $\epsilon_1 + \mathrm{i}\epsilon_2 = (n + \mathrm{i}K)^2$. One commonly observed consequence of effective medium theory is the lower (higher) energy of the plasma frequency (when $\epsilon_1(\omega)$ becomes zero), when the magnitude of the dielectric function of the host materials is larger (smaller) than that of the nanowire [9.222].

The dielectric function of the nanowires and the substrate can be separated out by using effective medium theory in reverse. Since the dielectric function of the host material is often known, and the dielectric function of the composite material can be measured (using reflection and transmission measurements in combination with either the Kramer–Kronig relations or Maxwell's equations), the complex dielectric function of the nanowires can be deduced. As an example, this technique was used to determine the frequency dependence of the real and imaginary parts of the dielectric function $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$ for an array of parallel bismuth nanowires filling the pores of an alumina template [9.223].

If both the dielectric permittivity and the magnetic permeability are negative, the material will display a negative index of refraction. A material with a negative index of refraction has interesting physics and can be used to make a lens that is not limited by the diffraction limit [9.224]. Such materials are artificially designed composites with components of subwavelength dimensions, often metallic and magnetic (nano)wires. Negative refraction is also observed in hyperbolic metamaterials, where the permeability is positive and the permittivity tensor has both positive and negative components. Parallel metallic nanowires in a dielectric matrix satisfy this condition [9.225]. Negative refraction at visible light was first realized using a silver nanowire array in porous anodic alumina [9.226].

Characteristic Optical Properties of Nanowires

Reflection and transmission measurements are often used for nanowire characterization. Infrared spectroscopy, for example, can be used to determine the plasma frequency, free carrier density, and donor impurity concentration as a function of temperature. This is especially useful for nanowire research, since Hall effect measurements cannot be made on nanowires. For example, the bandgap of gallium nitride and its temperature dependence were determined by infrared spectroscopy in nanowires in the 10–50 nm range in comparison to bulk values [9.227].

Another common method used to study nanowires is photoluminescence spectrum (PL) or fluorescence spectroscopy. PL has been used to study many properties of nanowires, such as the optical gap behavior, oxygen vacancies in ZnO nanowires [9.66], strain in Si nanowires [9.229], and quantum confinement effects in InP and Si nanowires [9.228, 230]. Figure 9.36 shows the photoluminescence spectrum of InP nanowires as a function of wire diameter, thereby providing direct information on the effective bandgap. As the wire diameter of an InP nanowire is decreased so that it becomes smaller than the bulk exciton diameter of 19 nm, quantum confinement effects set in, and the bandgap is increased. This results in an increase in the PL peak energy. The smaller the effective mass, the larger the quantum confinement effects. When the shift in the peak energy as a function of nanowire diameter (Fig. 9.36) is analyzed using an effective mass model, the reduced effective mass of the ex-

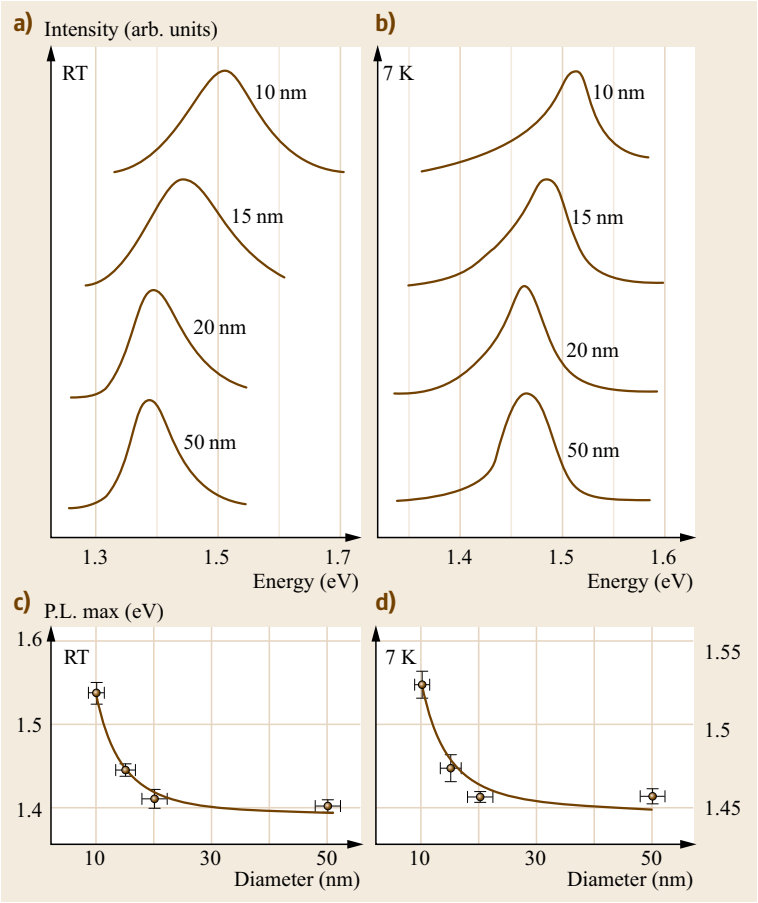


Fig. 9.36 Photoluminescence of InP nanowires of varying diameters at 7 K (**b**,**d**) and room temperature (**a**,**c**) showing quantum confinement effects of the exciton for wire diameters of less than 20 nm (after [9.228])

citon is deduced to be $0.052m_0$, which agrees quite well with the literature value of $0.065m_0$ for bulk InP. Although the line widths of the PL peak for the small-diameter nanowires (10 nm) are smaller at low temperature (7 K), strong quantum confinement and bandgap tunability effects persist at room temperature and thus are significant for photonics applications of nanowires (Sect. 9.3.2).

The resolution of photoluminescence (PL) optical imaging of a nanowire is, in general, limited by the wavelength of light. However, when a sample is placed very close to the detector, the light is not given a chance to diffract, and so samples much smaller than the wavelength of light can be resolved. This technique is known as near-field scanning optical microscopy (NSOM) and has been used to successfully image nanowires [9.231]. For example, Fig. 9.37 shows the topographical (a) and (b) NSOM PL images of a single ZnO nanowire.

Magneto-optics can be used to measure the electronic band structure of nanowires. For example, magneto-optics in conjunction with photoconductance has been proposed as a tool to determine band parameters of nanowires, such as the Fermi energy, electron effective mass, and the number of subbands to be considered [9.232]. Since different nanowire subbands have different electrical transmission properties, the electrical conductivity changes when light is used to excite electrons to higher subbands, thereby providing a method for studying the electronic structure of nanowires optically. Magneto-optics can also be used to study the magnetic properties of nanowires in relation to bulk properties [9.36, 233]. For example, the surface magneto-optical Kerr effect has been used to measure the dependence of the magnetic ordering temperature of Fe-Co alloy nanowires on the relative concentration of Fe and Co [9.233], and it was used to find that, unlike in the case of bulk Fe-Co alloys, cobalt in nanowires inhibits magnetic ordering. Nickel nanowires were found to have a strong increase in their magneto-optical activity with respect to bulk nickel. This increase is attributed to the plasmon resonance in the wires [9.234].

Nonlinear optical properties of nanowires have received particular attention since the nonlinear behavior

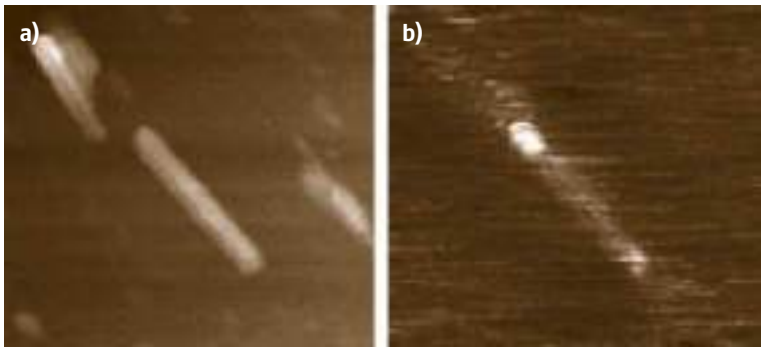


Fig. 9.37 **(a)** Topographical and **(b)** photoluminescence (PL) near-field scanning optical microscopy (NSOM) images of a single ZnO nanowire waveguide (after [9.231])

is often enhanced compared to bulk materials and the nonlinear effects can be utilized for a number of applications. One such study measured the second harmonic generation (SHG) and third harmonic generation (THG) in a single nanowire using near-field optical microscopy [9.235]. ZnO nanowires were shown to have strong SHG and THG effects that are highly polarization sensitive, and this polarization sensitivity can be explained on the basis of optical and geometrical considerations. Some components of the second harmonic polarization tensor are found to be enhanced in nanowires while others are suppressed as the wire diameter is decreased, and such effects could be of interest for device applications. It was also shown that the second-order nonlinearities are mostly wavelength-independent for $\lambda < 400\,\text{nm}$, which is in the transparent regime for ZnO, below the onset of bandgap absorption, and this observation is also of interest for device applications.

Reflectivity and transmission measurements have also been used to study the effects of quantum confinement and surface effects on the low-energy indirect transition in bismuth nanowires [9.236]. *Black* et al. investigated an intense and sharp absorption peak in bismuth nanowires, which is not observed in bulk bismuth. The energy position E_p of this strong absorption peak increases with decreasing diameter. However, the rate of increase in energy with decreasing diameter $|\partial E_\text{p}/\partial d_\text{W}|$ is an order of magnitude less than that predicted for either a direct interband transition or for intersubband transitions in bismuth nanowires. On the other hand, the magnitude of $|\partial E_\text{p}/\partial d_\text{W}|$ agrees well with that predicted for an indirect L-point valence to T-point valence band transition (Fig. 9.38). Since both the initial and final states for the indirect L–T point valence band transition downshift in energy as the wire diameter d_W is decreased, the shift in the absorption peak results from a difference between the effective masses and not from the actual value of either of the masses. Hence the diameter dependence of the absorption peak energy is an order of magnitude less for a valence-to-valence band indirect transition than for a direct interband L-point transition. Furthermore, the band-tracking effect for the indirect transition gives rise to a large value for the joint density of states, thus accounting for the high intensity of this feature. Figure 9.38 shows the experimentally observed transmission spectrum in bismuth nanowires of $\approx 45\,\text{nm}$ diameter (a), and the simulated optical transmission from an indirect transition in bismuth nanowires of $\approx 45\,\text{nm}$ diameter is also shown for comparison in (b). The indirect L–T point valence band transition mechanism [9.237] is also consistent with observations of the effect on the optical spectra of a decrease in the nanowire diameter and of n-type doping of bismuth nanowires with Te.

Phonon Confinement Effects

Phonons in nanowires are spatially confined by the nanowire cross-sectional area, crystalline boundaries, and surface disorder. These finite size effects give rise to phonon confinement, causing an uncertainty in the phonon wavevector, which typically gives rise to a frequency shift and lineshape broadening. Since zone center phonons tend to correspond to maxima in the phonon dispersion curves, the inclusion of contributions from a broader range of phonon wave vectors results in both a downshift in frequency and an asymmetric broadening of the Raman line, which develops a low-frequency tail. These phonon confinement effects have been theoretically predicted [9.238, 239] and experimentally observed in GaN [9.240], as shown in Fig. 9.39 for GaN nanowires with diameters in the range 10–50 nm. The application of these theoretical models indicates that broadening effects should be noticeable as the wire diameter in GaN nanowires decreases to $\approx 20\,\text{nm}$. When the wire diameter decreases further to $\approx 10\,\text{nm}$, the frequency downshift and asymmetric Raman line broadening effects should become observable in the Raman spectra for the GaN nanowires but are not found in the corresponding spectra for bulk GaN.

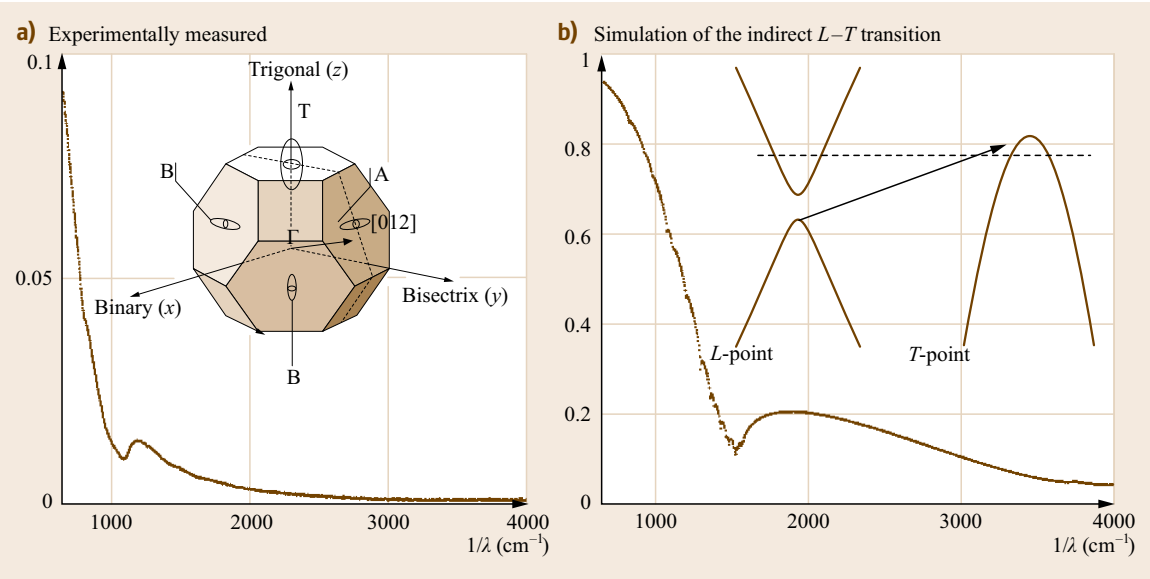


Fig. 9.38 **(a)** The measured optical transmission spectra as a function of wavenumber $(1/\lambda)$ of a $\approx 45\,\text{nm}$ diameter bismuth nanowire array. **(b)** The simulated optical transmission spectrum resulting from an indirect transition of an L-point electron to a T-point valence subband state. The *insert* in **(a)** shows the bismuth Brillouin zone, and the locations of the T-point hole and the three L-point electron pockets, including the nondegenerate A, and the doubly degenerate B pockets. The *insert* in **(b)** shows the indirect L–T point electronic transition induced by a photon with an energy equal to the energy difference between the initial and final states minus the phonon energy (about $100\,\text{cm}^{-1}$) needed to satisfy conservation of energy in a Stokes process (after [9.237])

The experimental spectra in Fig. 9.39 show the four $A_1 + E_1 + 2E_2$ modes expected from symmetry considerations for bulk GaN crystals. Two types of quantum confinement effects are observed. The first type is the observation of the downshift and the asymmetric broadening effects discussed above. Observations of such downshifts and asymmetric broadening have also been reported in 7 nm-diameter Si nanowires [9.241]. A second type of confinement effect found in Fig. 9.39 for GaN nanowires is the appearance of additional Raman features not found in the corresponding bulk spectra and associated with combination modes, and a zone boundary mode. Resonant enhancement effects were also observed for the $A_1(\text{LO})$ phonon at $728\,\text{cm}^{-1}$ at higher laser excitation energies [9.240].

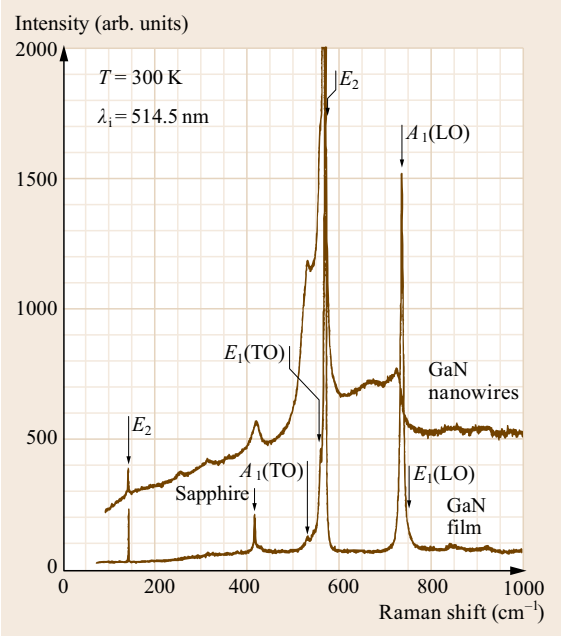


Fig. 9.39 Room-temperature Raman scattering spectra of GaN nanowires and of a 5 μm-thick GaN epilayer film with green (514.5 nm) laser excitation. The Raman scattering response was obtained by dividing the measured spectra by the Bose–Einstein thermal factor (after [9.240]) ►

9.3 Applications

In the preceding sections we have reviewed many of the central characteristics that make nanowires in some cases similar to and in some cases very different from their parent materials. We have also shown that some properties are diameter dependent, and these properties are therefore tunable during synthesis. Thus, nanowires offer a new toolbox of materials for engineers to design systems that could benefit in unprecedented ways from both the unique and tunable properties of nanowires and the small sizes of these nanostructures. As the synthetic methods for the production of nanowires continue to mature (Sect. 9.1) and nanowires are made in reproducible and cost-effective ways, the number of applications utilizing nanowires will continue to expand. This is a timely development, as the semiconductor industry will soon be reaching what seems to be its limit in feature size reduction, and approaching a classical-to-quantum size transition. At the same time, the field of biotechnology is expanding through the availability of tremendous genome information and innovative screening assays. Since nanowires are similar in size to the shrinking electronic components and to cellular biomolecules, it is only natural for nanowires to be good candidates for applications in these fields. Commercialization of nanowire devices, however, will require reliable mass production, effective assembly techniques, and quality control methods.

In this section, applications of nanowires to electronics (Sect. 9.3.1), optics (Sect. 9.3.2), energy (Sect. 9.3.3), chemical and biochemical sensing (Sect. 9.3.4), and magnetic media (Sect. 9.3.5) are discussed.

9.3.1 Electrical Applications

Nanowires are uniquely suited for some electrical applications. They are nano-sized in two dimensions and have one dimension that is not confined, providing a path for electrical conduction without the need for tunneling or carrier hopping. The microelectronics industry continues to face technological and economic challenges as the device feature size is decreased, especially below 20 nm. The self-assembly of nanowires might present a way to construct unconventional devices that do not rely on improvements in photolithography and, therefore, do not necessarily imply increasing fabrication costs. In addition, unlike traditional silicon processing, different semiconductors can be used simultaneously in nanowire devices to produce diverse functionalities. Not only can wires of different materials be combined, but a single wire can be made of different materials. For example, junctions of GaAs and GaP nanowire segments show rectifying behavior [9.111], thus demonstrating that good electronic interfaces between two different semiconductors can be achieved in the synthesis of multicomponent nanowires.

Electronic devices made from nanowires could also hold advantages due to their unique morphology. For example, in bulk field-effect transistors (FETs), the depletion layer formed below the source and drain region results in a source-drain capacitance that limits the operation speed. However, in nanowires, the source-drain contact can be wrapped around the semiconductor wire and thus the depletion layer can be limited to the nanowire volume inside the contact. Thus, depending on the device design, the source-drain capacitance in nanowires could be greatly minimized and possibly eliminated.

Device functionalities common in conventional semiconductor technologies, such as p-n junction diodes [9.167], field-effect transistors [9.169], logic gates [9.167], and light-emitting diodes [9.111, 242], have been demonstrated in nanowires, showing their promise as building blocks that could be used to construct complex integrated circuits by employing the *bottom-up* paradigm. In fact, as of June 2017 there are over 8000 US patent and patent applications containing the words *nanowire* and *transistor*. Several approaches have been investigated to form nanowire diodes (Sect. 9.2.2). For example, Schottky diodes can be formed by contacting a GaN nanowire with Al electrodes [9.168]. Furthermore, p-n junction diodes can be formed at the crossing of two nanowires, such as the crossing of n- and p-type InP nanowires doped by Te and Zn respectively [9.242], or Si nanowires doped by phosphorus (n-type) and boron (p-type) [9.243]. In addition to the crossing of two distinctive nanowires, heterogeneous junctions have also been constructed inside a single wire, either along the wire axis in the form of a nanowire superlattice [9.111], or perpendicular to the wire axis by forming a core-shell structure of silicon and germanium [9.151]. These various nanowire junctions not only possess current rectifying properties (Fig. 9.24) expected of bulk semiconductor devices, but they also exhibit electroluminescence (EL) that may be useful for optoelectronic applications, as shown in Fig. 9.40 for the electroluminescence of a crossed junction of n- and p-type InP nanowires [9.242] (see also Sect. 9.3.2).

The conductance of a semiconductor nanowire can be significantly modified by applying voltage at a third gate terminal, implying the utilization of nanowires in field-effect transistors (FETs). This gate terminal can either be the substrate [9.39, 244–247], a separate metal contact located close to the nanowire [9.248],

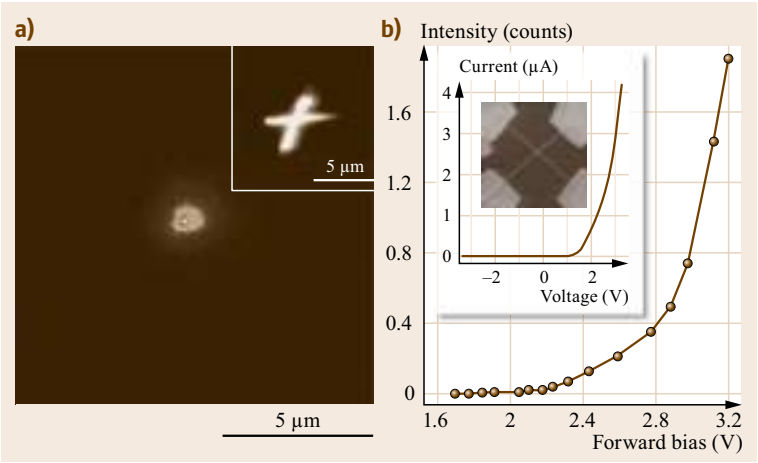


Fig. 9.40a,b Optoelectrical characterization of a crossed nanowire junction formed between 65 nm n-type and 68 nm p-type InP nanowires. **(a)** Electroluminescence (EL) image of the light emitted from a forward-biased nanowire p-n junction at 2.5 V. *Inset*, photoluminescence (PL) image of the junction. **(b)** EL intensity as a function of operation voltage. *Inset*, the SEM image and the *I–V* characteristics of the junction (after [9.242])

or another nanowire with a thick oxide coating in the crossed nanowire junction configuration [9.167]. Various logic devices performing basic logic functions have been demonstrated using nanowire junctions [9.167], as shown in Fig. 9.41 for the OR and AND logic gates constructed from two-by-one and one-by-three nanowire p-n junctions respectively. By functionalizing nanowires with redox-active molecules to store charge, nanowire FETs were demonstrated with two-level [9.169] and with eight-level [9.249] memory effects, which may be used for non-volatile memory or as switches. In another advance, In_2O_3 nanowire FETs with high-k dielectric materials were demonstrated, and substantially enhanced performance was obtained due to the highly efficient coupling of the gate [9.250]. A vertical FET with a surrounding gate geometry has also been demonstrated, which has the potential for high-density nanoscale memory and logic devices [9.251]. In addition, nanowire transistors were made without conventional doping [9.252, 253]. In these devices the gate surrounds the nanowire and fully or partially depletes the nanowire, thus controlling the carrier transport between the source and drain. These nanowire transistors were used in field-programmable logic arrays (FPGA), using an order of magnitude less area and operating power compared to conventional CMOS designs [9.252].

One challenge in making wrap-around transistors is finding a low-cost way to make the gate contact. For the application of tactile imaging, *Wu* et al. demonstrated the use of piezotronic ZnO nanowires that do not require a wrap-around gate [9.254]. The pressure applied to a ZnO nanowire in an array leads to a change in the polarization of immobile ions in the crystal that are present at the interface of the metal contact and the nanowire. This allows pressure to be sensed electrically without needing a gate terminal in each device.

Nanowires have also been proposed for applications associated with electron field emission [9.255, 256], such as flat panel displays, because of their small diameter and large curvature at the nanowire tip, which may reduce the threshold voltage for electron emission [9.257]. In this regard, the demonstration of very high field emission currents from the sharp tip ($\approx$ 10 nm radius) of a Si cone [9.255], from carbon nanotubes [9.258], from Si nanowires inside a carbon nanotube [9.259], and from Co nanowires [9.260], has stimulated interest in this potential area of application for nanowires.

Although the application of nanowires to thermoelectrics appears very promising, these materials are still in the research phase of the development cycle and are far from being commercialized. One challenge for thermoelectric devices based on nanowires lies in finding a suitable host material that will not reduce ZT too much due to the unwanted heat conduction through the host material. Therefore, the host material should have a low thermal conductivity and occupy a volume fraction in the composite material that is as low as possible, while still providing the quantum confinement and the support for the nanowires. In addition, the technology to scale up the sequential bonding of individual nanowires by dopant type (p-type to n-type to p-type, and so on) still needs to be developed.

Meshes of nanowires in the percolation limit have a combination of mechanical, electrical, and optical properties that make them useful for displays and tactile devices. Silver nanowire meshes, in particular, are being tested as transparent conduction films since their properties were shown to be similar to those of state-of-the-art indium tin oxide (ITO) [9.261]. These films are prepared from Ag nanowire suspensions, allowing for inkjet printing and low-temperature processing.

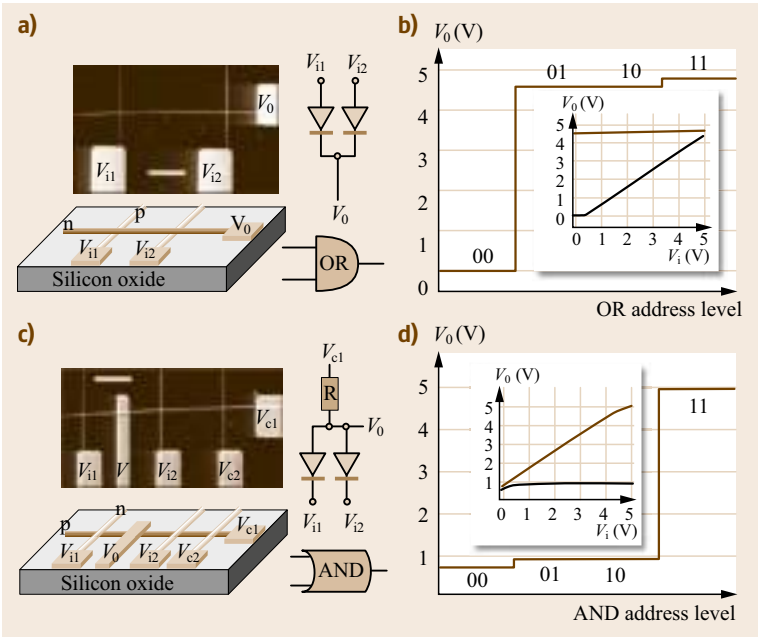


Fig. 9.41 Nanowire logic gates: **(a)** Schematic of logic OR gate constructed from a two (p-Si) by one (n-GaN) crossed nanowire junction. The *inset* shows the SEM image (scale bar: 1 μm) of an assembled OR gate and the symbolic electronic circuit. **(b)** The output voltage of the circuit in **(a)** versus the four possible logic address level inputs: (0,0); (0,1); (1,0); (1,1), where logic 0 input is 0 V and logic 1 is 5 V (same for below). **(c)** Schematic of logic AND gate constructed from a one (p-Si) by three (n-GaN) crossed nanowire junction. The *inset* shows the SEM image (scale bar: 1 μm) of an assembled AND gate and the symbolic electronic circuit. **(d)** The output voltage of the circuit in **(c)** versus the four possible logic address level inputs (after [9.167])

9.3.2 Optical Applications

Nanowires also hold promise for optical applications. One-dimensional systems exhibit a singularity in their joint density of states, allowing quantum effects in nanowires to be optically observable, sometimes even at room temperature. Since the density of states of a nanowire in the quantum limit (small wire diameter) is highly localized in energy, the available states quickly fill up with electrons as the intensity of the incident light is increased. This filling up of the subbands, as well as other effects that are unique to low-dimensional materials, lead to strong optical nonlinearities in quantum wires. Quantum wires may thus yield optical switches with a lower switching energy and increased switching speed compared to currently available optical switches.

Light-emitting diodes (LEDs) have been achieved in junctions between a p-type and an n-type nanowire (Fig. 9.40) [9.242] and in superlattice nanowires with p-type and n-type segments [9.111]. The light emission was localized to the junction area, and was polarized in the superlattice nanowire. An electrically driven laser was fabricated from CdS nanowires. The devices were fabricated by patterning a metal contact onto an n-type CdS nanowire, which resided on a p^+ silicon wafer. The cleaved ends of the wire formed the laser cavity, so that in forward bias, the light characteristic of lasing was observed at the end of the wire [9.262]. LEDs have also been achieved with core-shell structured nanowires made of n-GaN/InGaN/p-GaN [9.263].

Light emission from quantum wire p-n junctions is especially interesting for laser applications, because quantum wires can form lasers with lower excitation thresholds than their bulk counterparts and they also exhibit decreased sensitivity of performance to temperature [9.264]. Furthermore, the emission wavelength can be tuned for a given material composition by simply altering the geometry of the wire.

Lasing action has been reported in ZnO nanowires with wire diameters that are much smaller than the wavelength of the light emitted ($\lambda = 385\,\mathrm{nm}$) [9.110] (Fig. 9.42). Since the edges and lateral surfaces of ZnO nanowires are faceted (Sect. 9.2.1), they form optical cavities that sustain desired cavity modes. Compared to conventional semiconductor lasers, the exciton laser action employed in zinc oxide nanowire lasers exhibits a lower lasing threshold ($\approx 40\,\mathrm{kW/cm^2}$) than their 3-D counterparts ($\approx 300\,\mathrm{kW/cm^2}$). In order to utilize exciton confinement effects in the lasing action, the exciton binding energy ($\approx 60\,\mathrm{meV}$ in ZnO) must be greater than the thermal energy ($\approx 26\,\mathrm{meV}$ at 300 K). Decreasing the wire diameter increases the excitation binding energy and lowers the threshold for lasing. PL NSOM imaging confirmed the waveguiding properties of the anisotropic and the well-faceted structure of ZnO nanowires, limiting the emission to the tips of the ZnO nanowires [9.231]. Time-resolved studies have illuminated the dynamics of the emission process [9.265].

Lasing nanowires were also obtained with other materials systems, such as ZnS nanowires in anodic aluminum oxide templates [9.266], GaN nanowires

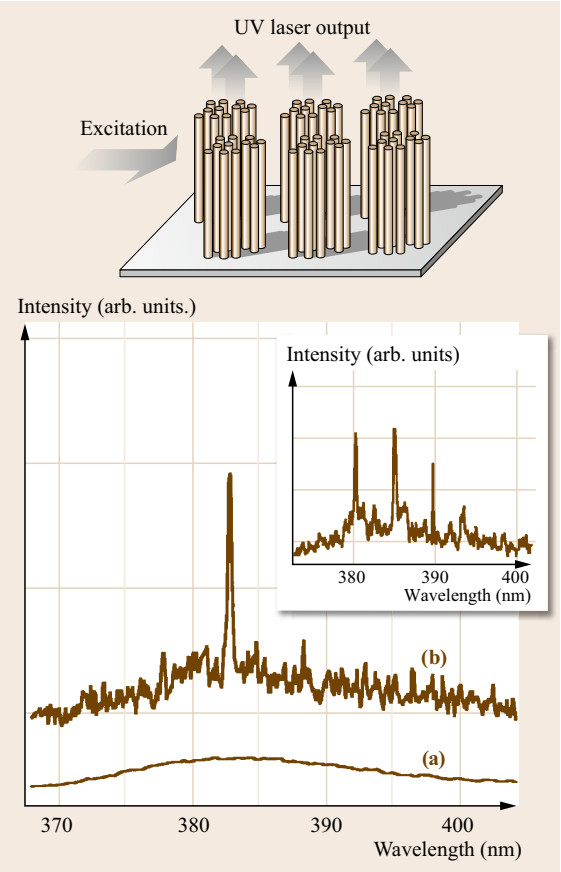


Fig. 9.42 A schematic of lasing in ZnO nanowires and the PL spectra of ZnO nanowires at two excitation intensities. The emission spectrum of trace (a) is taken below the lasing threshold (20 kW/cm^2) and the spectrum of trace (b) above it (100 kW/cm^2) (after [9.110])

[9.267], and a low-lasing-threshold, self-organized GaN(core)/AlGaN(shell) structure [9.268].

Nanowires have also been demonstrated to have good waveguiding properties. Quantitative studies of cadmium sulfide (CdS) nanowire structures show that light propagation takes place with only moderate losses through sharp and even acute angle bends. In addition, active devices made with nanowires have shown that efficient injection into and modulation of light through nanowire waveguides can be achieved [9.269]. By linking ZnO nanowire light sources to SnO_2 waveguides, the possibility of optical integrated circuitry is introduced [9.270].

Nanowire photodetectors were also proposed. ZnO nanowires were found to display a strong photocurrent response to ultraviolet (UV) light irradiation [9.271]. The conductivity of the nanowire increased by four orders of magnitude compared to the dark state. The response of the nanowire was reversible, and selective to photon energies above the bandgap, suggesting that ZnO nanowires could be good candidates for optoelectronic switches.

Nanowires have been also proposed for another type of optical switching. Light with its electric field normal to the wire axis excites a transverse free carrier resonance inside the wire, while light with its electric field parallel to the wire axis excites a longitudinal free carrier resonance inside the wire. Since nanowires are highly anisotropic, these two resonances occur at two different wavelengths and thus result in absorption peaks at two different energies. Gold nanowires dispersed in an aqueous solution align along the electric field when a DC voltage is applied. The energy of the absorption peak can be toggled between the transverse and longitudinal resonance energies by changing the alignment of the nanowires under polarized light illumination using an electric field [9.272, 273]. Thus, electro-optical modulation is achieved.

Nanowires may also be used as barcode tags for optical readout. Nanowires containing gold, silver, nickel, palladium, and platinum were fabricated [9.216] by electrochemical filling of porous anodic alumina, so that each nanowire consisted of segments of various metal constituents. Thus many types of nanowires can be made from a handful of materials, and identified by the order of the metal segments along their main axis, and the length of each segment. Barcode readout is possible by reflectance optical microscopy. The segment length is limited by the Rayleigh diffraction limit, and not by synthesis limitations, and thus can be as small as 145 nm. Figure 9.43a shows an optical image of many Au-Ag-Au-Ag barcoded wires, where the silver segments show higher reflectivity. Figure 9.43b is a backscattering mode field electron SEM (FE-SEM) image of a single nanowire, highlighting the composition and segment length variations along the nanowire.

Nanowires have found applications in solar cells or photovoltaics. Both the large surface area and the high conductivity along the length of nanowires are favorable for their use in photovoltaics. An early example involved n-type nanorods for inorganic-organic solar cells [9.274], which offer promise from a manufacturing and cost-effectiveness standpoint. In a hybrid nanocrystal-organic solar cell, the incident light forms bound electron-hole pairs (excitons) in both the inorganic nanocrystal and in the surrounding organic medium. These excitons diffuse to the inorganic-organic interface and disassociate to form an electron and a hole. Since conjugated polymers usually have poor electron mobilities, the inorganic phase is chosen to have a higher electron affinity than the organic phase so that the organic phase carries the holes and the

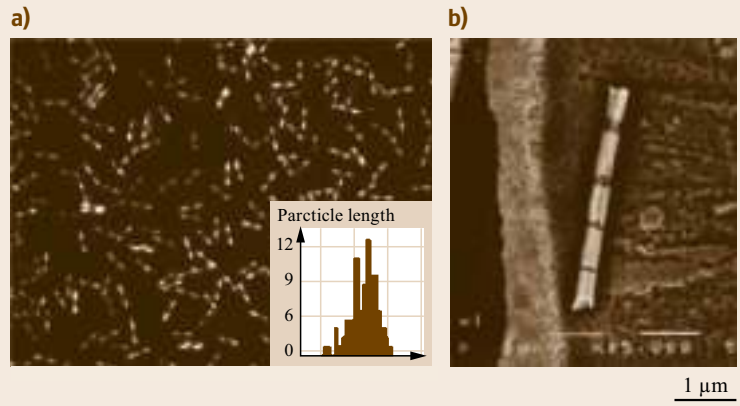


Fig. 9.43 (a) An optical image of many short barcoded Au-Ag-Au-Au wires and (b) an FE-SEM image of an Au/Ag barcoded wire with multiple strips of varying length. The *insert* in (a) shows a histogram of the particle lengths for 106 particles in this image (after [9.216])

semiconductor carries the electrons. The separated electrons and holes drift to the external electrodes through the inorganic and organic materials respectively. However, only those excitons formed within an exciton diffusion length from an interface can dissociate before recombining, and therefore the distance between the dissociation sites limits the efficiency of a solar cell. A solar cell prepared from a composite of CdSe nanorods inside poly(3-ethylthiophene) [9.274] yielded monochromatic power efficiencies of 6.9% and power conversion efficiencies of 1.7% under air mass AM1.5 (AM) illumination. The nanorods provide a large surface area with good chemical bonding to the polymer for efficient charge transfer and exciton dissociation. Furthermore, they provide a good conduction path for the electrons to reach the electrode. Their enhanced absorption coefficient and their tunable bandgap are also characteristics that can be used to enhance the energy conversion efficiency of solar cells.

Inorganic solar cells have the highest efficiencies to date, and nanowire photovoltaics can be an avenue for reduced cost and improved efficiency, by reducing material usage, increasing light absorption, and reducing recombination. Sol Voltaics AB is working to commercialize nanowire solar cells using a low-cost process based on VLS-grown vertical nanowire arrays with built-in axial p-n junctions [9.275, 276]. Devices of InP nanowires and GaAs nanowires achieved 13.8% and 15.3% efficiencies respectively. Silicon nanowires have also been used to make high-efficiency solar cells. Vertically aligned silicon nanowire arrays for solar cells are often made by either MACE or RIE (Sect. 9.1.3). Many studies on silicon nanowires put the p-n junction radially inside the wires [9.277]. These nanowire solar cells may be able to circumvent the Shockley–Queisser limit [9.278]. In radially doped nanowires, the solar-cell junction runs down the nanowire axis, and due to their small diameter, photocarriers are always generated near the junction. Thus in radially doped nanowires, hot carriers can be extracted out of the cell before they relax to the band edge.

Solar cells with silicon nanowire arrays on bulk silicon and the junction inside the bulk silicon are being commercialized by Advanced Silicon Group [9.279]. This cell design combines some of the advantages of nanowire cells with traditional silicon photovoltaics. In particular, these nanowire cells (sometimes called black silicon cells) have very low reflection and efficiently trap light inside the solar cell, with both effects leading to improved efficiency. Since the nanowires are smaller than the wavelength of light, the incident light sees an effective medium (Sect. 9.2.4) [9.280]. By tapering the wires, the index of refraction can be graded leading to a very low reflection of light. In black silicon cells, the junction is placed below the nanowires. It is formed by standard doping in the bulk single-crystal silicon substrate. Such pristine junctions are critical for low carrier recombination, low dark current, and thus high-efficiency cells. Pristine junctions are more challenging to create in the radial nanowire cell design.

9.3.3 Energy Applications

Energy Storage

Nanowires have been researched for energy application. Electrochemical and photoelectrochemical systems benefit from the nanowire array high surface area and one-dimensional conductivity. Individual nanowires can be used as in-situ electrochemical probes to explore intrinsic electrochemical processes and behaviors at the nanoscale.

Silicon has the highest lithium-ion absorption capacity of any known anode material for lithium ion batteries, but it expands and contracts by four times its volume when charged and discharged. This extensive expansion leads to pulverization of the electrode mate-

rial and the capacity of the anode quickly degrades with cycling. Nanowires allow for expansion and contraction during lithiation and delithiation in the radial direction while maintaining the structural integrity of the anode. In addition, compared to silicon nanoparticles, nanowires allow for conduction down the nanowire axis [9.281]. *Mai* et al. reported the characterization of single nanowire anode capacity fading in Li ion–based energy storage devices [9.282], while *Oleshko* et al. incorporated all the battery components in a single heterostructure nanowire for a detailed structural and electrochemical performance study using in-situ multimode analytical STEM imaging [9.283]. Several companies are commercializing silicon nanowire anodes including Nexeon, Advanced Silicon Group, Silexta, and Amprius [9.284–286]. Complete nanowire nanobattery arrays embedded in porous anodic alumina were demonstrated; however, these employed vanadium oxide chemistry [9.287].

Qui and coauthors used TiO_2 passivated InP nanowire arrays to create a photocatalytic system that reduces CO_2 with water to produce methanol [9.288]. More recently, the importance of uniformity of the nanowires in photocatalytic reactions was demonstrated with individual Si nanowire devices [9.289].

Nanowire-based capacitors can be readily produced as arrays in templates, for example by ALD [9.290]. Synthesis of individual nanowire capacitors as coaxial heterostructures by CVD offers extensive material flexibility. Following this approach, *Ajayan* et al. demonstrated the fabrication of single-nanowire capacitors made of a Cu-Cu_2O-C nanowire with capacitance values up to $140\,\mu F/cm^2$, well above the values reported for metal-insulator-metal micro-/nanocapacitors [9.291].

Piezoelectric Generators

Mechanical energy is abundant in the environment and in industrial processes but such energy is usually wasted, be it in the form of mechanical vibration, hydraulic pressure, or air flow, among others. One way to harvest these forms of energy is through the piezoelectric effect, which converts mechanical energies into electricity. This effect is well known at the macroscale in bulk materials such as lead zirconate titanate (PZT) and lead magnesium niobate-lead titanate (PMN–PT). The development of energy nanogenerators, which convert mechanical energy into electricity using one-dimensional piezoelectric materials, now enables the harvesting of small-scale mechanical energy, which can, in turn, be used to charge electrochemical systems or to power low-consumption microdevices [9.292, 293]. *Wang* and *Song* reported the first piezoelectric nanogenerators in 2006. In this seminal work, an output voltage pulse of several millivolts was obtained after bending a ZnO nanowire with an AFM tip [9.294]. With the ultimate objective of improving energy conversion efficiency and output power for practical applications, much research in optimizing materials [9.295–297] and system design [9.298, 299] continues to be performed in the field of nanowire-based technology [9.300, 301].

The major types of piezoelectric nanogenerator structures include the vertically aligned nanowire networks, the laterally aligned nanowire arrays, and the nanowire-based nanocomposites. The fundamental working principle applicable to the various realizations of vertically aligned nanowire arrays, hinges on the electromechanical coupling of the semiconductor behavior and the piezoelectric property of the piezoelectric nanowire. In this case, the application of external stress induces the appearance of a piezoelectric potential along the nanowire, which affects the details of the Schottky barrier at the interface between the nanowire and metallic pads. This, in turn, accounts for the flow of electrons between the metallic electrodes, through the closed circuit, to ensure thermodynamic equilibrium. In addition to the original ZnO-based device, AFM tip-induced lateral bending was also demonstrated for nanowires with other chemical compositions such as GaN, CdS, InN, and ZnS [9.302–306]. To avoid the use of an AFM tip, simplify the fabrication process, and improve the device stability, *Xu* and coworkers proposed a vertical nanowire array integrated nanogenerator (VING), which is created by packaging an array of vertically aligned ZnO nanowires with polymethyl methacrylate (PMMA) and by connecting the nanowire with top and bottom flat electrodes [9.307]. ZnO VINGs are currently among the most popular nanogenerators. Moving beyond ZnO (which has a relatively low piezoelectric constant and therefore yields low-performance nanogenerators), researchers are also exploring more efficient nanowire perovskite piezoelectric materials such as $Pb(Zr, Ti)O_3$ (PZT) [9.308]. This path is promising but suffers from a number of difficulties that are absent in ZnO-based VINGs, namely a more difficult synthesis process and the limited flexibility of the substrate typically used for materials growth.

The first lateral type of piezoelectric nanogenerator (where deformation is always induced by laterally bending the nanowires via the substrate or by applying a pressure in the radial nanowire direction) was demonstrated by *Wang* et al. in 2007 [9.309]. This device is based on the voltage generation of an individual $BaTiO_3$ nanowire under periodic tensile mechanical load. The nanofabrication process of the flexural stage involves complex and expensive steps due to the need to manipulate materials of extremely small sizes. In con-

trast, laterally integrated nanogenerators (LINGs) using a flexible polymer substrate are not only easier to fabricate but also feature improved performance [9.310]. In addition to ZnO, LINGs were made using other nanowire materials, such as potassium sodium niobate, which have the advantage of displaying a much higher piezoelectric constant and a lower free charge carrier density (which reduces stray screening responsible for lower piezoelectric response), and of being biocompatible [9.311]. The field of LINGs is still fairly recent and much effort is being made to simplify the fabrication processes, including the development of a scalable sweeping-printing method for fabricating flexible devices with ZnO nanowire arrays [9.312], or a one-step fabrication of ZnO nanowire networks [9.307]. Notably, the development of electrospinning techniques has enabled the use of ultralong piezoelectric nanofibers and large-scale LING fabrication [9.313, 314].

A novel type of piezoelectric nanogenerator based on nanowire-polymer composites has been explored to simplify the fabrication process compared to that of VINGs and LINGs. In this case, the nanogenerator assembly consists of four functional layers: top and bottom electrodes, flexible substrate, and a nanowire-composite layer. This multilayer structure is reminiscent of that of conventional piezoelectric ceramics. While the composite may feature piezoelectric properties inferior to those of ceramics, the presence of the polymer matrix significantly enhances the flexibility of the device. Compared to VINGs and LINGs, where the device fabrication is largely dictated by the assembly process of the nanowire arrays, the nanowire-composite design can be easily obtained by directly mixing as-synthesized nanowires with the polymer matrix such as polydimethylsiloxane (PDMS). *Hu* et al. realized large output nanogenerators by rational unipolar assembly of conical ZnO nanowires [9.315] and similar structures were also reported for composites of $NaNbO_3$ or GaN nanowires mixed with PDMS [9.296, 316].

Although much effort has already been made in promoting the practical application of piezoelectric nanogenerators, a number of critical issues remain to be fully resolved. This includes issues related to the integration of nanogenerators for high-output power sources, the structural design for increasing the energy harvesting efficiency in different practical conditions, and the need for improved stability and reliability [9.300].

9.3.4 Chemical and Biochemical Sensing Devices

Sensors for chemical and biochemical substances with nanowires as the sensing probe are a very attractive application area for nanowires based on a variety of materials and specific applications. Nanowire sensors will potentially be smaller and more sensitive, demand less power, and react faster than their macroscopic counterparts. Arrays of nanowire sensors could, in principle, achieve nanometer-scale spatial resolution and therefore provide accurate real-time information regarding not only the concentration of a specific analyte but also its spatial distribution. Such arrays could be very useful, for example, for dynamic studies on the effects of chemical gradients on biological cells.

The operation of many sensors made with nanowires, nanotubes, or nano-contacts is based on the reversible change in the conductance of the nanostructure upon absorption of the agent to be detected. The increased sensitivity and faster response time of nanowires are a result of the large surface-to-volume ratio and the small cross-section available for conduction channels, thus increasing the interaction between stray fields and the charge carriers. In the bulk, on the other hand, the abundance of charges can effectively shield external fields, and the volume of material can afford many alternative conduction channels. Therefore, a stronger chemical stimulus and longer response time are necessary to observe changes in the physical properties of a 3-D sensor in comparison to a nanowire.

It is often necessary to modify the surface of the nanowires to achieve a strong interaction with the analytes that need to be detected. Surface modifications utilize the self-assembly, chemisorption, or chemical reactivity of selected organic molecules and polymers towards metal and oxide surfaces. Examples include thiols on gold, isocyanides on platinum, and siloxanes on silica. These surface coatings regulate the binding and chemical reactivity of other molecules towards the nanowire in a predictable manner [9.317].

Cui et al. placed silicon nanowires made by the VLS method (Sect. 9.1.2) between two metal electrodes and modified the silicon oxide coating of the wire through the addition of molecules that are sensitive to the analyte to be detected [9.318]. For example, a pH sensor was made by covalently linking an amine-containing silane to the surface of the nanowire. Variations in the pH of the solution into which the nanowire was immersed caused protonation and deprotonation of the $-NH_2$ and the $-SiOH$ groups on the surface of the nanowire. The variation in surface charge density regulates the conductance of the nanowire; due to the p-type characteristics of a silicon wire, the conductance increases with the addition of negative surface charge. The combined acid and base behavior of the surface groups results in an approximately linear dependence of the conductance on pH in the pH range 2–9, thus leading to a direct-readout pH meter. This same type of approach was used for the detection of

the binding of biomolecules, such as streptavidin using biotin-modified nanowires (Fig. 9.44). This nanowire-based device has high sensitivity and could detect streptavidin binding down to a concentration of 10 pM (10^{-12} mol). Subsequent results demonstrated the capabilities of these functionalized Si nanowire sensors as DNA sensors down to the femtomolar range [9.319].

A similar approach was used by *Favier* et al., who made a nanosensor for the detection of hydrogen from of an array of palladium nanowires between two metal contacts [9.53]. They demonstrated that nanogaps were present in their nanowire structure, and upon absorption of H_2 and formation of Pd hydride, the nanogap structure would close and improve the electrical contact, thereby increasing the conductance of the nanowire array. The response time of these sensors was 75 ms, and they could operate in the range 0.5–5% H_2 before saturation occurred.

A mechanical sensing paradigm typically employs the nanowires as oscillating cantilevers and detects shifts in their mechanical resonance frequency (related to changes in cantilever mass) upon chemical binding events. As above, molecular binding specificity is achieved through functionalization of the nanowire surface. Using a 103 µm-long ZnO nanowire, *Wang* et al. measured a 26 pg weight with ≈ 10% error and 1.2 pg sensitivity [9.320]. These conditions allow for the detection of small nanoparticles (e.g., in aerosols). Single-macromolecule detection was demonstrated through a multimode frequency analysis in a sophisticated doubly clamped nanoelectromechanical resonator [9.321, 322].

Nanowires can act as waveguides to deliver optical signals in the near field and improve signal-to-noise ratio and spatial resolution in spectroscopic measurements [9.323]. *Yang* et al. utilized a tin oxide nanowire to guide the optical signal from a laser passing through the sample. Light scattered by the end of the nanowire was detected in the far field to obtain the PL or Raman spectrum of the analyte. Tin oxide nanowires were further used as a probe to poke through a cell's membrane and collect PL light from the cytoplasm, demonstrating a nanoscale endoscope [9.324]. The device sensed the position of luminescent quantum dots artificially positioned in the cell. Waveguiding in silver nanowires through excitation of surface plasmon polaritons has also been employed to sense molecules through surface-enhanced Raman scattering and other plasmonic effects [9.325].

9.3.5 Magnetic Applications

It has been demonstrated that arrays of single-domain magnetic nanowires can be prepared with controlled nanowire diameter and length, aligned along a common direction and arranged in a close-packed ordered array (Sect. 9.1), and that the magnetic properties (coercivity, remanence, and dipolar magnetic interwire interaction) can be controlled to achieve a variety of magnetic applications [9.49, 88].

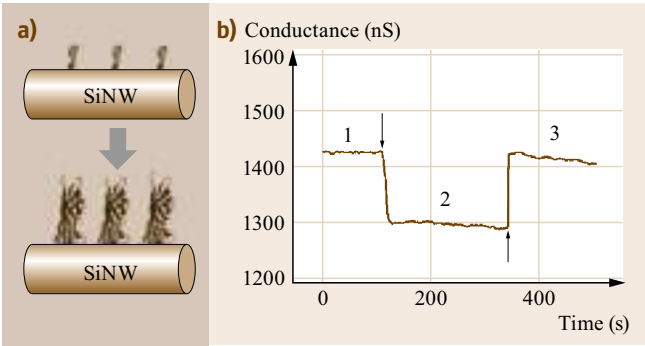


Fig. 9.44 **(a)** Streptavidin molecules bind to a silicon nanowire functionalized with biotin. The binding of streptavidin to biotin causes the nanowire to change its electrical resistance. **(b)** The electrical conductance of a biotin-modified silicon nanowire exposed to streptavidin in a buffer solution (regions 1 and 3) and with the introduction of a solution of antibiotin monoclonal antibody (region 2) (after [9.318])

The most interesting of these applications is for magnetic storage, where the large nanowire aspect ratio (length/diameter) is advantageous for preventing the onset of the *superparamagnetic* limit at which the magnetization direction in the magnetic grains can be reversed by the thermal energy $k_B T$, thereby resulting in loss of recorded data in the magnetic recording medium. The magnetic energy in a grain can be increased by increasing either the volume or the anisotropy of the grain. If the volume is increased, the particle size increases, so the resolution is decreased. For spherical magnetized grains, the superparamagnetic limit at room temperature is reached at 70 Gbit/in^2. In nanowires, the anisotropy is very large and yet the wire diameters are small, so that the magnetostatic switching energy can easily be above the thermal energy while the spatial resolution is large. For magnetic data storage applications, a large aspect ratio is needed for the nanowires in order to maintain a high coercivity, and a sufficient separation between nanowires is needed to suppress interwire magnetic dipolar coupling. Thus nanowires can form stable and highly dense magnetic memory arrays with packing densities in excess of 10^{11} wires/cm^2.

Developments in the controlled movement of domain walls in magnetic nanowires by short pulses of spin-polarized current show the possibility of nonvolatile memory devices with performance and relia-

bility comparable to conventional solid-state memory but at the affordability of conventional magnetic storage [9.326]. For instance, the racetrack memory design is a device made up of an array of horizontally or vertically oriented magnetic nanowires arranged on a silicon chip where individual spintronic read/write processes can operate on a train of tens of domain walls, thereby storing a series of data bits in each nanowire.

Moving beyond information storage, magnetic logic architectures, known as *domain-wall logic*, have the advantage of not requiring the use of transistors and therefore do not show significant heating associated with data switching [9.327]. A domain wall is a mobile interface between regions of magnetization aligned in opposite directions. In particular, a submicrometer planar nanowire made from a soft magnetic material such as Permalloy has been demonstrated as an excellent system featuring domain walls [9.328–330]. The large shape anisotropy of a nanowire provides a preferred magnetization alignment along the axis of the nanowires, leading to a robust binary information encoding. The role of the magnetic domain wall is to provide a transition edge in a changing signal. A logical *1* is defined as the magnetization pointing in the direction of propagation of domain walls and a logical *0* as the magnetization in the opposite direction. It was shown that cusp-shaped planar nanowires can be used to reverse the direction of magnetization, thereby providing the basis for the logical NOT operation [9.331]. Other basic logic functions, especially those with two-input functions, such as AND or OR, are also possible and are necessary elementary operations for computational applications. More complex logic circuits also require functions such as a fan-out structure, which creates two identical copies of an input signal and a cross-over structure, which makes it possible for two signals to pass over each other without interference [9.332].

9.4 Concluding Remarks

In this chapter, we reviewed the synthesis, characterization, and physical properties of nanowires, placing particular emphasis on nanowire properties that differ from those of their bulk counterparts and potential applications that might result from the special structures and properties of nanowires.

We have shown that the field of nanowire research continues to develop very rapidly year after year, driven by the development of a variety of complementary nanowire synthesis methods and effective tools for measuring nanowire structures and properties (Sects. 9.1 and 9.2). At present, much of the progress is at the demonstration-of-concept level, with many gaps in knowledge remaining to be elucidated, theoretical models to be developed, and new nanowire systems to be explored. Having demonstrated that many of the most interesting discoveries to date relate to nanowire properties not present in their bulk material counterparts, we can expect future research emphasis to be increasingly focused on smaller-diameter nanowires, where new unexplored physical phenomena related to quantum confinement effects are more likely to be found. We can also expect the development of applications to follow, some coming sooner and others later. Many promising applications are now at the early demonstration stage (Sect. 9.3), but are moving ahead rapidly because of their promise of new functionality not previously available in the fields of electronics, optoelectronics, biotechnology, magnetics, and energy conversion and generation, among others. Many exciting challenges remain in advancing both the nanoscience and the nanotechnological promise already demonstrated by the nanowire research described in this review.

Acknowledgments. In loving memory of our advisor, mentor, and friend Dr. Millie Dresselhaus who passed away during the production of this handbook.

O.R. acknowledges support from the National Science Foundation (DMR-1151614).

References

9.1 R.P. Feynman: There's plenty of room at the bottom, Eng. Sci. **23**, 22–36 (1960)

9.2 M.S. Dresselhaus, Y.-M. Lin, O. Rabin, A. Jorio, A.G. Souza Filho, M.A. Pimenta, R. Saito, G.G. Samsonidze, G. Dresselhaus: Nanowires and nanotubes, Mater. Sci. Eng. C **23**, 129–140 (2003), also in: Current trends in nanotechnologies: From materials to systems, Eur. Mater. Res. Soc. Symp. Proc., Vol. 140, ed. by W. Jantsch, H. Grimmeiss, G. Marietta (Elsevier, Amsterdam 2002)

9.3 R. Saito, G. Dresselhaus, M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College

Press, London 1998)
9.4 M.S. Dresselhaus, G. Dresselhaus, P. Avouris (Eds.): *Carbon Nanotubes: Synthesis, Structure, Properties and Applications*, Springer Ser. Top. Appl. Phys., Vol. 80 (Springer, Berlin, Heidelberg 2001) pp. 1–447
9.5 R.C. Haddon: Special issue on carbon nanotubes, Acc. Chem. Res. **35**, 997–1113 (2002)
9.6 Y. Mao, S.S. Wong: General, room-temperature method for the synthesis of isolated as well as arrays of single-crystalline ABO_4-type nanorods, J. Am. Chem. Soc. **126**, 15245–15252 (2004)
9.7 E. Braun, Y. Eichen, U. Sivan, G. Ben-Yoseph: DNA-templated assembly and electrode attachment of a conducting silver wire, Nature **391**, 775–778 (1998)
9.8 G. Sauer, G. Brehm, S. Schneider, K. Nielsch, R.B. Wehrspohn, J. Choi, H. Hofmeister, U. Gösele: Highly ordered monocrystalline silver nanowire arrays, J. Appl. Phys. **91**, 3243–3247 (2002)
9.9 K.E. Korte, S.E. Skrabalak, Y. Xia: Rapid synthesis of silver nanowires through a CuCl- or $CuCl_2$-mediated polyol process, J. Mater. Chem. **18**, 437–441 (2008)
9.10 S. Zhu, Y. Gao, B. Hu, J. Li, J. Su, Z. Fan, J. Zhou: Transferable self-welding silver nanowire network as high performance transparent flexible electrode, Nanotechnology **24**, 335202 (2013)
9.11 L.J. Andrés, M.F. Menéndez, D. Gómez, A.L. Martínez, N. Bristow, J.P. Kettle, A. Menéndez, B. Ruiz: Rapid synthesis of ultra-long silver nanowires for tailor-made transparent conductive electrodes: Proof of concept in organic solar cells, Nanotechnology **26**, 265201 (2015)
9.12 G.L. Hornyak, C.J. Patrissi, C.M. Martin: Fabrication, characterization and optical properties of gold nanoparticle/porous alumina composites: The non-scattering Maxwell–Garnett limit, J. Phys. Chem. B **101**, 1548–1555 (1997)
9.13 X.Y. Zhang, L.D. Zhang, Y. Lei, L.X. Zhao, Y.Q. Mao: Fabrication and characterization of highly ordered Au nanowire arrays, J. Mater. Chem. **11**, 1732–1734 (2001)
9.14 F. Kim, K. Sohn, J. Wu, J. Huang: Chemical synthesis of gold nanowires in acidic solutions, J. Am. Chem. Soc. **130**(44), 14442–14443 (2008) doi:10.1021/ja806759v
9.15 Y.-T. Cheng, A.M. Weiner, C.A. Wong, M.P. Balogh, M.J. Lukitsch: Stress-induced growth of bismuth nanowires, Appl. Phys. Lett. **81**, 3248–3250 (2002)
9.16 W. Shim, J. Ham, K.-I. Lee, W.Y. Jeung, M. Johnson, W. Lee: On-film formation of bi nanowires with extraordinary electron mobility, Nano Lett. **9**, 18–22 (2009) doi:10.1021/nl8016829
9.17 J. Heremans, C.M. Thrush, Y.-M. Lin, S. Cronin, Z. Zhang, M.S. Dresselhaus, J.F. Mansfield: Bismuth nanowire arrays: Synthesis, galvanomagnetic properties, Phys. Rev. B **61**, 2921–2930 (2000)
9.18 L. Piraux, S. Dubois, J.L. Duvail, A. Radulescu, S. Demoustier-Champagne, E. Ferain, R. Legras: Fabrication and properties of organic, metal nanocylinders in nanoporous membranes, J. Mater. Res. **14**, 3042–3050 (1999)
9.19 K. Hong, F.Y. Yang, K. Liu, D.H. Reich, P.C. Searson, C.L. Chien, F.F. Balakirev, G.S. Boebinger: Giant positive magnetoresistance of Bi nanowire arrays in high magnetic fields, J. Appl. Phys. **85**, 6184–6186 (1999)
9.20 A.J. Yin, J. Li, W. Jian, A.J. Bennett, J.M. Xu: Fabrication of highly ordered metallic nanowire arrays by electrodeposition, Appl. Phys. Lett. **79**, 1039–1041 (2001)
9.21 Z. Zhang, J.Y. Ying, M.S. Dresselhaus: Bismuth quantum-wire arrays fabricated by a vacuum melting and pressure injection process, J. Mater. Res. **13**, 1745–1748 (1998)
9.22 Z. Zhang, D. Gekhtman, M.S. Dresselhaus, J.Y. Ying: Processing and characterization of single-crystalline ultrafine bismuth nanowires, Chem. Mater. **11**, 1659–1665 (1999)
9.23 T.E. Huber, M.J. Graf, C.A. Foss, P. Constant: Processing and characterization of high-conductance bismuth wire array composites, J. Mater. Res. **15**, 1816–1821 (2000)
9.24 L. Li, G. Li, Y. Zhang, Y. Yang, L. Zhang: Pulsed electrodeposition of large-area, ordered $Bi_{1-x}Sb_x$ nanowire arrays from aqueous solutions, J. Phys. Chem. B **108**, 19380–19383 (2004)
9.25 M.S. Sander, A.L. Prieto, R. Gronsky, T. Sands, A.M. Stacy: Fabrication of high-density, high aspect ratio, large-area bismuth telluride nanowire arrays by electrodeposition into porous anodic alumina templates, Adv. Mater. **14**, 665–667 (2002)
9.26 M. Chen, Y. Xie, J. Lu, Y.J. Xiong, S.Y. Zhang, Y.T. Qian, X.M. Liu: Synthesis of rod-, twinrod-, and tetrapod-shaped CdS nanocrystals using a highly oriented solvothermal recrystallization technique, J. Mater. Chem. **12**, 748–753 (2002)
9.27 D. Xu, Y. Xu, D. Chen, G. Guo, L. Gui, Y. Tang: Preparation of CdS single-crystal nanowires by electrochemically induced deposition, Adv. Mater. **12**, 520–522 (2000)
9.28 D. Routkevitch, T. Bigioni, M. Moskovits, J.M. Xu: Electrochemical fabrication of CdS nanowire arrays in porous anodic aluminum oxide templates, J. Phys. Chem. **100**, 14037–14047 (1996)
9.29 L. Manna, E.C. Scher, A.P. Alivisatos: Synthesis of soluble and processable rod-, arrow-, teardrop-, and tetrapod-shaped CdSe nanocrystals, J. Am. Chem. Soc. **122**, 12700–12706 (2000)
9.30 D. Routkevitch, A.A. Tager, J. Haruyama, D. Al-Mawlawi, M. Moskovits, J.M. Xu: Nonlithographic nano-wire arrays: fabrication, physics, and device applications, IEEE Trans. Electron. Dev. **43**, 1646–1658 (1996)
9.31 D.S. Xu, D.P. Chen, Y.J. Xu, X.S. Shi, G.L. Guo, L.L. Gui, Y.Q. Tang: Preparation of II–VI group semiconductor nanowire arrays by dc electrochemical deposition in porous aluminum oxide templates, Pure Appl. Chem. **72**, 127–135 (2000)
9.32 R. Adelung, F. Ernst, A. Scott, M. Tabib-Azar, L. Kipp, M. Skibowski, S. Hollensteiner, E. Spiecker, W. Jäger, S. Gunst, A. Klein, W. Jägermann, V. Zaporojtchenko, F. Faupel: Self-assembled nanowire networks by deposition of copper onto layered-crystal

surfaces, Adv. Mater. **14**, 1056–1061 (2002)
9.33 T. Gao, G.W. Meng, J. Zhang, Y.W. Wang, C.H. Liang, J.C. Fan, L.D. Zhang: Template synthesis of single-crystal Cu nanowire arrays by electrodeposition, Appl. Phys. A **73**, 251–254 (2001)
9.34 D. Al-Mawlawi, N. Coombs, M. Moskovits: Magnetic-properties of Fe deposited into anodic aluminum-oxide pores as a function of particle-size, J. Appl. Phys. **70**, 4421–4425 (1991)
9.35 F. Li, R.M. Metzger: Activation volume of α-Fe particles in alumite films, J. Appl. Phys. **81**, 3806–3808 (1997)
9.36 A. Sugawara, T. Coyle, G.G. Hembree, M.R. Scheinfein: Self-organized Fe nanowire arrays prepared by shadow deposition on NaCl(110) templates, Appl. Phys. Lett. **70**, 1043–1045 (1997)
9.37 G.S. Cheng, L.D. Zhang, Y. Zhu, G.T. Fei, L. Li, C.M. Mo, Y.Q. Mao: Large-scale synthesis of single crystalline gallium nitride nanowires, Appl. Phys. Lett. **75**, 2455–2457 (1999)
9.38 G.S. Cheng, L.D. Zhang, S.H. Chen, Y. Li, L. Li, X.G. Zhu, Y. Zhu, G.T. Fei, Y.Q. Mao: Ordered nanostructure of single-crystalline GaN nanowires in a honeycomb structure of anodic alumina, J. Mater. Res. **15**, 347–350 (2000)
9.39 Y. Huang, X. Duan, Y. Cui, C.M. Lieber: Gallium nitride nanowire nanodevices, Nano Lett. **2**, 101–104 (2002)
9.40 X. Duan, C.M. Lieber: Laser-assisted catalytic growth of single crystal GaN nanowires, J. Am. Chem. Soc. **122**, 188–189 (2000)
9.41 A.D. Berry, R.J. Tonucci, M. Fatemi: Fabrication of GaAs and InAs wires in nanochannel glass, Appl. Phys. Lett. **69**, 2846–2848 (1996)
9.42 J.R. Heath, F.K. LeGoues: A liquid solution synthesis of single-crystal germanium quantum wires, Chem. Phys. Lett. **208**, 263–268 (1993)
9.43 Y. Wu, P. Yang: Germanium nanowire growth via simple vapor transport, Chem. Mater. **12**, 605–607 (2000)
9.44 Y.F. Zhang, Y.H. Tang, N. Wang, C.S. Lee, I. Bello, S.T. Lee: Germanium nanowires sheathed with an oxide layer, Phys. Rev. B **61**, 4518–4521 (2000)
9.45 S.J. May, J.-G. Zheng, B.W. Wessels, L.J. Lauhon: Dendritic nanowire growth mediated by a self-assembled catalyst, Adv. Mater. **17**, 598–602 (2005)
9.46 S. Han, C. Li, Z. Liu, B. Lei, D. Zhang, W. Jin, X. Liu, T. Tang, C. Zhou: Transition metal oxide core-shell nanowires: Generic synthesis and transport studies, Nano Lett. **4**, 1241–1246 (2004)
9.47 M.P. Zach, K.H. Ng, R.M. Penner: Molybdenum nanowires by electrodeposition, Science **290**, 2120–2123 (2000)
9.48 L. Sun, P.C. Searson, L. Chien: Electrochemical deposition of nickel nanowire arrays in single-crystal mica films, Appl. Phys. Lett. **74**, 2803–2805 (1999)
9.49 K. Nielsch, R. Wehrspohn, S. Fischer, H. Kronmüller, J. Barthel, J. Kirschner, U. Gösele: Magnetic properties of 100 nm nickel nanowire arrays obtained from ordered porous alumina templates, MRS Symp. Proc. **636**, D1.9-1–D1.9-6 (2001)
9.50 Y. Wang, X. Jiang, T. Herricks, Y. Xia: Single crystalline nanowires of lead: Large-scale synthesis, mechanistic studies, and transport measurements, J. Phys. Chem. B **108**, 8631–8640 (2004)
9.51 E. Lifshitz, M. Bashouti, V. Kloper, A. Kigel, M.S. Eisen, S. Berger: Synthesis and characterization of PbSe quantum wires, multipods, quantum rods, cubes, Nano Lett. **3**, 857–862 (2003)
9.52 W. Lu, P. Gao, W.B. Jian, Z.L. Wang, J. Fang: Perfect orientation ordered in-situ one-dimensional self-assembly of Mn-doped PbSe nanocrystals, J. Am. Chem. Soc. **126**, 14816–14821 (2004)
9.53 F. Favier, E.C. Walter, M.P. Zach, T. Benter, R.M. Penner: Hydrogen sensors and switches from electrodeposited palladium mesowire arrays, Science **293**, 2227–2231 (2001)
9.54 X. Huang, N. Zheng: One-pot, high-yield synthesis of 5-fold twinned Pd nanowires and nanorods, J. Am. Chem. Soc. **131**, 4602–4603 (2009) doi:10.1021/ja9009343
9.55 B. Gates, B. Mayers, B. Cattle, Y. Xia: Synthesis, characterization of uniform nanowires of trigonal selenium, Adv. Funct. Mater. **12**, 219–227 (2002)
9.56 C.A. Huber, T.E. Huber, M. Sadoqi, J.A. Lubin, S. Manalis, C.B. Prater: Nanowire array composites, Science **263**, 800–802 (1994)
9.57 Y. Cui, L.J. Lauhon, M.S. Gudiksen, J. Wang, C.M. Lieber: Diameter-controlled synthesis of single crystal silicon nanowires, Appl. Phys. Lett. **78**, 2214–2216 (2001)
9.58 A.M. Morales, C.M. Lieber: A laser ablation method for the synthesis of crystalline semiconductor nanowires, Science **279**, 208–211 (1998)
9.59 N. Wang, Y.F. Zhang, Y.H. Tang, C.S. Lee, S.T. Lee: SiO_2-enhanced synthesis of Si nanowires by laser ablation, Appl. Phys. Lett. **73**, 3902–3904 (1998)
9.60 M.K. Sunkara, S. Sharma, R. Miranda, G. Lian, E.C. Dickey: Bulk synthesis of silicon nanowires using a low-temperature vapor-liquid-solid method, Appl. Phys. Lett. **79**, 1546–1548 (2001)
9.61 K. Peng, M. Zhang, A. Lu, N.-B. Wong, R. Zhang, S.-T. Lee: Ordered silicon nanowire arrays via nanosphere lithography and metal-induced etching, Appl. Phys. Lett. **90**, 163123 (2007)
9.62 S. Vaddiraju, H. Chandrasekaran, M.K. Sunkara: Vapor phase synthesis of tungsten nanowires, J. Am. Chem. Soc. **125**, 10792–10793 (2003)
9.63 J.P. Heremans, C.M. Thrush, D.T. Morelli, M.-C. Wu: Thermoelectric power of bismuth nanocomposites, Phys. Rev. Lett. **88**, 216801-1–216801-4 (2002)
9.64 Y. Li, G.S. Cheng, L.D. Zhang: Fabrication of highly ordered ZnO nanowire arrays in anodic alumina membranes, J. Mater. Res. **15**, 2305–2308 (2000)
9.65 P. Yang, H. Yan, S. Mao, R. Russo, J. Johnson, R. Saykally, N. Morris, J. Pham, R. He, H.-J. Choi: Controlled growth of ZnO nanowires and their optical properties, Adv. Funct. Mater. **12**, 323–331 (2002)
9.66 M.J. Zheng, L.D. Zhang, G.H. Li, W.Z. Shen: Fabrication and optical properties of large-scale uniform zinc oxide nanowire arrays by one-step electrochemical deposition technique, Chem. Phys. Lett. **363**, 123–128 (2002)

9.67 G.A. Ozin: Nanochemistry: Synthesis in diminishing dimensions, Adv. Mater. **4**, 612–649 (1992)

9.68 R.J. Tonucci, B.L. Justus, A.J. Campillo, C.E. Ford: Nanochannel array glass, Science **258**, 783–785 (1992)

9.69 J.Y. Ying: Nanoporous systems and templates, Sci. Spec. **18**, 56–63 (1999)

9.70 J.W. Diggle, T.C. Downie, C.W. Goulding: Anodic oxide films on aluminum, Chem. Rev. **69**, 365–405 (1969)

9.71 J.P. O'Sullivan, G.C. Wood: The morphology and mechanism of formation of porous anodic films on aluminum, Proc. R. Soc. A **317**, 511–543 (1970)

9.72 A.P. Li, F. Müller, A. Birner, K. Nielsch, U. Gösele: Hexagonal pore arrays with a 50–420 nm interpore distance formed by self-organization in anodic alumina, J. Appl. Phys. **84**, 6023–6026 (1998)

9.73 J.P. Sullivan, G.C. Wood: The morphology, mechanism of formation of porous anodic films on aluminum, Proc. R. Soc. A **317**, 511–543 (1970)

9.74 O. Jessensky, F. Müller, U. Gösele: Self-organized formation of hexagonal pore arrays in anodic alumina, Appl. Phys. Lett. **72**, 1173–1175 (1998)

9.75 Y.-M. Lin, X. Sun, S. Cronin, Z. Zhang, J.Y. Ying, M.S. Dresselhaus: Fabrication, transport properties of Te-doped bismuth nanowire arrays. In: *Molecular Electronics: MRS Symposium Proceedings*, Vol. 582, ed. by S.T. Pantelides, M.A. Reed, J. Murday, A. Aviran (Materials Research Society Press, Pittsburgh 2000), pp. 1–6, Chap. H10.3

9.76 F. Li, L. Zhang, R.M. Metzger: On the growth of highly ordered pores in anodized aluminum oxide, Chem. Mater. **10**, 2470–2480 (1998)

9.77 H. Masuda, M. Satoh: Fabrication of gold nanodot array using anodic porous alumina as an evaporation mask, Jpn. J. Appl. Phys. **35**, L126–L129 (1996)

9.78 H. Masuda, H. Yamada, M. Satoh, H. Asoh, M. Nakao, T. Tamamura: Highly ordered nanochannel-array architecture in anodic alumina, Appl. Phys. Lett. **71**, 2770–2772 (1997)

9.79 W. Lee, R. Ji, C.A. Ross, U. Gösele, K. Nielsch: Wafer-scale Ni imprint stamps for porous alumina membranes based on interference lithography, Small **2**, 978–982 (2006)

9.80 C.R. Martin: Nanomaterials: A membrane-based synthetic approach, Science **266**, 1961–1966 (1994)

9.81 P. Apel: Track etching technique in membrane technology, Radiat. Meas. **34**, 559–566 (2001)

9.82 E. Ferain, R. Legras: Track-etched membrane: Dynamics of pore formation, Nucl. Instrum. Methods B **84**, 331–336 (1993)

9.83 A. Blondel, J.P. Meier, B. Doudin, J.-P. Ansermet: Giant magnetoresistance of nanowires of multilayers, Appl. Phys. Lett. **65**, 3019–3021 (1994)

9.84 K. Liu, C.L. Chien, P.C. Searson, Y.Z. Kui: Structural and magneto-transport properties of electrodeposited bismuth nanowires, Appl. Phys. Lett. **73**, 1436–1438 (1998)

9.85 C.A. Huber, T.E. Huber: A novel microstructure: semiconductor-impregnated porous Vycor glass, J. Appl. Phys. **64**, 6588–6590 (1988)

9.86 J.S. Beck, J.C. Vartuli, W.J. Roth, M.E. Leonowicz, C.T. Kresge, K.D. Schmitt, C.T.-W. Chu, D.H. Olson, E.W. Sheppard, S.B. McCullen, J.B. Higgins, J.L. Schlenker: A new family of mesoporous molecular sieves prepared with liquid crystal templates, J. Am. Chem. Soc. **114**, 10834–10843 (1992)

9.87 C.-G. Wu, T. Bein: Conducting polyaniline filaments in a mesoporous channel host, Science **264**, 1757–1759 (1994)

9.88 T. Thurn-Albrecht, J. Schotter, G.A. Kästle, N. Emley, T. Shibauchi, L. Krusin-Elbaum, K. Guarini, C.T. Black, M.T. Tuominen, T.P. Russell: Ultrahigh-density nanowire arrays grown in self-assembled diblock copolymer templates, Science **290**, 2126–2129 (2000)

9.89 P.-Y. Chen, X. Dang, M.T. Klug, N.M. Dorval Courchesne, J. Qi, M.N. Hyder, A.M. Belcher, P.T. Hammond: M13 virus-enabled synthesis of titanium dioxide nanowires for tunable mesoporous semiconducting networks, Chem. Mater. **27**, 1531–1540 (2015) doi:10.1021/cm503803u

9.90 J.C. Zhou, C.M. Soto, M.-S. Chen, M.A. Bruckman, M.H. Moore, E. Barry, B.R. Ratna, P.E. Pehrsson, B.R. Spies, T.S. Confer: Biotemplating rod-like viruses for the synthesis of copper nanorods and nanowires, J. Nanobiotechnol. **10**, 18 (2012) doi:10.1186/1477-3155-10-18

9.91 M. Knez, M. Sumser, A.M. Bittner, C. Wege, H. Jeske, T.P. Martin, K. Kern: Spatially selective nucleation of metal clusters on the tobacco mosaic virus, Adv. Funct. Mater. **14**, 116–124 (2004) doi:10.1002/adfm.200304376

9.92 X.Z. Fan, E. Pomerantseva, M. Gnerlich, A. Brown, K. Gerasopoulos, M. McCarthy, J. Culver, R. Ghodssi: Tobacco mosaic virus: A biological building block for micro/nano systems, J. Vac. Sci. Technol. A **31**, 050815-1–05081524 (2013)

9.93 Y.-M. Lin, S.B. Cronin, J.Y. Ying, M.S. Dresselhaus, J.P. Heremans: Transport properties of Bi nanowire arrays, Appl. Phys. Lett. **76**, 3944–3946 (2000)

9.94 A.W. Adamson: *Physical Chemistry of Surfaces* (Wiley, New York 1982) p. 338

9.95 R. Ferré, K. Ounadjela, J.M. George, L. Piraux, S. Dubois: Magnetization processes in nickel and cobalt electrodeposited nanowires, Phys. Rev. B **56**, 14066–14075 (1997)

9.96 H. Zeng, M. Zheng, R. Skomski, D.J. Sellmyer, Y. Liu, L. Menon, S. Bandyopadhyay: Magnetic properties of self-assembled Co nanowires of varying length and diameter, J. Appl. Phys. **87**, 4718–4720 (2000)

9.97 Y. Peng, H.L. Zhang, S.-L. Pan, H.-L. Li: Magnetic properties and magnetization reversal of α-Fe nanowires deposited in alumina film, J. Appl. Phys. **87**, 7405–7408 (2000)

9.98 L. Piraux, J.M. George, J.F. Despres, C. Leroy, E. Ferain, R. Legras, K. Ounadjela, A. Fert: Giant magnetoresistance in magnetic multilayered nanowires, Appl. Phys. Lett. **65**, 2484–2486 (1994)

9.99 S. Bhattacharrya, S.K. Saha, D. Chakravorty: Nanowire formation in a polymeric film, Appl. Phys. Lett. **76**, 3896–3898 (2000)

9.100 G. Yi, W. Schwarzacher: Single crystal superconductor nanowires by electrodeposition, Appl. Phys. Lett. **74**, 1746–1748 (1999)

9.101 D. Al-Mawlawi, C.Z. Liu, M. Moskovits: Nanowires formed in anodic oxide nanotemplates, J. Mater. Res. **9**, 1014–1018 (1994)

9.102 Z. Yao, C. Wang, Y. Li, N.-Y. Kim: AAO-assisted synthesis of highly ordered, large-scale TiO_2 nanowire arrays via sputtering and atomic layer deposition, Nanoscale Res. Lett. **10**, 166 (2015) doi:10.1186/s11671-015-0872-9

9.103 J.W. Elam, D. Routkevitch, P.P. Mardilovich, S.M. George: Conformal coating on ultrahigh-aspect-ratio nanopores of anodic alumina by atomic layer deposition, Chem. Mater. **15**, 3507–3517 (2003) doi:10.1021/cm0303080

9.104 M. Daub, M. Knez, U. Goesele, K. Nielsch: Ferromagnetic nanotubes by atomic layer deposition in anodic alumina membranes, J. Appl. Phys. **101**, 09J111 (2007) doi:10.1063/1.2712057

9.105 R.S. Wagner, W.C. Ellis: Vapor-liquid-solid mechanism of single crystal growth, Appl. Phys. Lett. **4**, 89–90 (1964)

9.106 Y. Wu, P. Yang: Direct observation of vapor-liquid-solid nanowire growth, J. Am. Chem. Soc. **123**, 3165–3166 (2001)

9.107 Y. Wu, R. Fan, P. Yang: Block-by-block growth of single-crystalline Si/SiGe superlattice nanowires, Nano Lett. **2**, 83–86 (2002)

9.108 S. Sharma, M.K. Sunkara, R. Miranda, G. Lian, E.C. Dickey: A novel low temperature synthesis method for semiconductor nanowires. In: *Synthesis, Functional Properties and Applications of Nanostructures, Mater. Res. Soc. Symp. Proc., San Francisco*, Vol. 676, ed. by H.W. Hahn, D.L. Feldheim, C.P. Kubiak, R. Tannenbaum, R.W. Siegel (Materials Research Society, Pittsburgh 2001) p. Y1.6

9.109 J. Johansson, B.A. Wacaser, K.A. Dick, W. Seifert: Growth related aspects of epitaxial nanowires, Nanotechnology **17**, S355–S361 (2006) doi:10.1088/0957-4484/17/11/S21

9.110 M.H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, P. Yang: Room-temperature ultraviolet nanowire nanolasers, Science **292**, 1897–1899 (2001)

9.111 M.S. Gudiksen, L.J. Lauhon, J. Wang, D.C. Smith, C.M. Lieber: Growth of nanowire superlattice structures for nanoscale photonics and electronics, Nature **415**, 617–620 (2002)

9.112 Y. Wu, J. Xiang, C. Yang, W. Lu, C.M. Lieber: Single-crystal metallic nanowires and metal/semiconductor nanowire heterostructures, Nature **430**, 61–65 (2004)

9.113 M.T. Björk, B.J. Ohlsson, T. Sass, A.I. Persson, C. Thelander, M.H. Magnusson, K. Deppert, L.R. Wallenberg, L. Samuelson: One-dimensional steeplechase for electrons realized, Nano Lett. **2**, 87–89 (2002)

9.114 N. Wang, Y.H. Tang, Y.F. Zhang, C.S. Lee, S.T. Lee: Nucleation and growth of Si nanowires from silicon oxide, Phys. Rev. B **58**, R16024–R16026 (1998)

9.115 Y.F. Zhang, Y.H. Tang, N. Wang, C.S. Lee, I. Bello, S.T. Lee: One-dimensional growth mechanism of crystalline silicon nanowires, J. Cryst. Growth **197**, 136–140 (1999)

9.116 S.T. Lee, Y.F. Zhang, N. Wang, Y.H. Tang, I. Bello, C.S. Lee, Y.W. Chung: Semiconductor nanowires from oxides, J. Mater. Res. **14**, 4503–4507 (1999)

9.117 D.D.D. Ma, C.S. Lee, Y. Lifshitz, S.T. Lee: Periodic array of intramolecular junctions of silicon nanowires, Appl. Phys. Lett. **81**, 3233–3235 (2002)

9.118 S.V. Thombare, A.F. Marshall, P.C. McIntyre: Kinetics of germanium nanowire growth by the vapor-solid-solid mechanism with a Ni-based catalyst, APL Mater. **1**, 061101 (2013) doi:10.1063/1.4833935

9.119 J. Lensch-Falk, E. Hemesath, D. Perea, L. Lauhon: Alternative catalysts for VSS growth of silicon and germanium nanowires, J. Mater. Chem. **19**, 849–857 (2009) doi:10.1039/b817391e

9.120 F. Wang, A. Dong, J. Sun, R. Tang, H. Yu, W.E. Buhro: Solution-liquid-solid growth of semiconductor nanowires, Inorganic Chem. **45**, 7511–7521 (2006)

9.121 H. Jansen, M. de Boer, R. Legtenberg, M. Elwenspoek: The black silicon method: A universal method for determining the parameter setting of a fluorine-based reactive ion etcher in deep silicon trench etching with profile control, J. Micromech. Microeng. **5**, 115–120 (1995)

9.122 Y.Q. Fu, A. Colli, A. Fasoli, J.K. Luo, A.J. Flewitt, A.C. Ferrari, W.I. Milne: Deep reactive ion etching as a tool for nanostructure fabrication, J. Vac. Sci. Technol. B **27**, 1520–1526 (2009)

9.123 X. Liu, P.R. Coxon, M. Peters, B. Hoex, J.M. Cole, D.J. Fray: Black silicon: Fabrication methods, properties and solar energy applications, Energy Environ. Sci. **7**, 3223–3263 (2014)

9.124 F. Toor, J.B. Miller, L.M. Davidson, L. Nichols, W. Duan, M.P. Jura, J. Yim, J. Forziati, M.R. Black: Nanostructured silicon via metal assisted catalyzed etch (MACE): Chemistry fundamentals and pattern engineering, Nanotechnology **27**(41), 412003 (2016)

9.125 Z. Huang, N. Geyer, P. Werner, J. de Boor, U. Gösele: Metal-assisted chemical etching of silicon: A review, Adv. Mater. **23**, 285–308 (2011)

9.126 B. Gates, Y. Yin, Y. Xia: A solution-phase approach to the synthesis of uniform nanowires of crystalline selenium with lateral dimensions in the range of 10–30 nm, J. Am. Chem. Soc. **122**, 12582–12583 (2000)

9.127 B. Mayers, B. Gates, Y. Yin, Y. Xia: Large-scale synthesis of monodisperse nanorods of Se/Te alloys through a homogeneous nucleation and solution growth process, Adv. Mater. **13**, 1380–1384 (2001)

9.128 B. Gates, Y. Wu, Y. Yin, P. Yang, Y. Xia: Single-crystalline nanowires of Ag_2Se can be synthesized by templating against nanowires of trigonal Se, J. Am. Chem. Soc. **123**, 11500–11501 (2001)

9.129 B. Gates, B. Mayers, Y. Wu, Y. Sun, B. Cattle, P. Yang, Y. Xia: Synthesis and characterization of crystalline Ag_2Se nanowires through a template-engaged reaction at room temperature, Adv. Funct. Mater. **12**, 679–686 (2002)

9.130 H. Yu, P.C. Gibbons, W.E. Buhro: Bismuth, tellurium and bismuth telluride nanowires, J. Mater. Chem. **14**, 595–602 (2004)

9.131 X. Peng, J. Wickham, A.P. Alivisatos: Kinetics of II–VI, III–V colloidal semiconductor nanocrystal growth: 'Focusing' of size distributions, J. Am. Chem. Soc. **120**, 5343–5344 (1998)

9.132 M.P. Zach, K. Inazu, K.H. Ng, J.C. Hemminger, R.M. Penner: Synthesis of molybdenum nanowires with millimeter-scale lengths using electrochemical step edge decoration, Chem. Mater. **14**, 3206–3216 (2002)

9.133 Dedi, P.-C. Lee, C.-H. Chien, G.-P. Dong, W.-C. Huang, C.-L. Chen, C.-M. Tseng, S.R. Harutyunyan, C.-H. Lee, Y.-Y. Chen: Stress-induced growth of single-crystalline lead telluride nanowires and their thermoelectric transport properties, Appl. Phys. Lett. **103**, 023115 (2013) doi:10.1063/1.4813606

9.134 J. Ham, W. Shim, D.H. Kim, S. Lee, J. Roh, S.W. Sohn, K.H. Oh, P.W. Voorhees, W. Lee: Direct growth of compound semiconductor nanowires by on-film formation of nanowires: Bismuth telluride, Nano Lett. **9**, 2867–2872 (2009) doi:10.1021/nl9010518

9.135 N.A. Melosh, A. Boukai, F. Diana, B. Gerardot, A. Badolto, P.M. Petroff, J.R. Heath: Ultrahigh-density nanowire lattices and circuits, Science **300**, 112–115 (2003)

9.136 P. Yang, F. Kim: Langmuir–Blodgett assembly of one-dimensional nanostructures, ChemPhysChem **3**, 503–506 (2002)

9.137 B. Messer, J.H. Song, P. Yang: Microchannel networks for nanowire patterning, J. Am. Chem. Soc. **122**, 10232–10233 (2000)

9.138 P.A. Smith, C.D. Nordquist, T.N. Jackson, T.S. Mayer, B.R. Martin, J. Mbindyo, T.E. Mallouk: Electric-field assisted assembly and alignment of metallic nanowires, Appl. Phys. Lett. **77**, 1399–1401 (2000)

9.139 D. Whang, S. Jin, C.M. Lieber: Large-scale hierarchical organization of nanowires for functional nanosystems, Jpn. J. Appl. Phys. **43**, 4465–4470 (2004)

9.140 S. Jin, D.M. Whang, M.C. McAlpine, R.S. Friedman, Y. Wu, C.M. Lieber: Scalable interconnection and integration of nanowire devices without registration, Nano Lett. **4**, 915–919 (2004)

9.141 M. Li, R.B. Bhiladvala, T.J. Morrow, J.A. Sioss, K.-K. Lew, J.M. Redwing, C.D. Keating, T.S. Mayer: Bottom-up assembly of large-area nanowire resonator arrays, Nat. Nanotechnol. **3**, 88–92 (2008)

9.142 S. Evoy, N. DiLello, V. Deshpande, A. Narayanan, H. Liu, M. Riegelman, B.R. Martin, B. Hailer, J.-C. Bradley, W. Weiss, T.S. Mayer, Y. Gogotsi, H.H. Bau, T.E. Mallouk, S. Raman: Dielectrophoretic assembly and integration of nanowire devices with functional CMOS operating circuitry, Microelectron. Eng. **75**, 31–42 (2004)

9.143 T. Kuykendall, P.J. Pauzauskie, Y.F. Zhang, J. Goldberger, D. Sirbuly, J. Denlinger, P.D. Yang: Crystallographic alignment of high-density gallium nitride nanowire arrays, Nat. Mater. **3**, 524–528 (2004)

9.144 O. Rabin, P.R. Herz, S.B. Cronin, Y.-M. Lin, A.I. Akinwande, M.S. Dresselhaus: Nanofabrication using self-assembled alumina templates. In: *Nonlithographic and Lithographic Methods for Nanofabrication, Mater. Res. Soc. Symp. Proc., Boston*, Vol. 636, ed. by J.A. Rogers, A. Karim, L. Merhari, D. Norris, Y. Xia (Materials Research Society, Pittsburgh 2001) pp. D4.7-1–D4.7-6

9.145 O. Rabin, P.R. Herz, Y.-M. Lin, A.I. Akinwande, S.B. Cronin, M.S. Dresselhaus: Formation of thick porous anodic alumina films and nanowire arrays on silicon wafers and glass, Adv. Funct. Mater. **13**, 631–638 (2003)

9.146 O. Rabin, P.R. Herz, Y.-M. Lin, S.B. Cronin, A.I. Akinwande, M.S. Dresselhaus: Arrays of nanowires on silicon wafers. In: *21st Int. Conf. Thermoelectr. Proc. ICT '02 Long Beach* (IEEE, Piscataway 2002) pp. 276–279

9.147 Y.H. Tang, Y.F. Zhang, N. Wang, C.S. Lee, X.D. Han, I. Bello, S.T. Lee: Morphology of Si nanowires synthesized by high-temperature laser ablation, J. Appl. Phys. **85**, 7981–7983 (1999)

9.148 Y. Ding, Z.L. Wang: Structure analysis of nanowires and nanobelts by transmission electron microscopy, J. Phys. Chem. B **108**, 12280–12291 (2004)

9.149 S.B. Cronin, Y.-M. Lin, O. Rabin, M.R. Black, G. Dresselhaus, M.S. Dresselhaus, P.L. Gai: Bismuth nanowires for potential applications in nanoscale electronics technology, Microsc. Microanal. **8**, 58–63 (2002)

9.150 M.S. Sander, R. Gronsky, Y.-M. Lin, M.S. Dresselhaus: Plasmon excitation modes in nanowire arrays, J. Appl. Phys. **89**, 2733–2736 (2001)

9.151 L.J. Lauhon, M.S. Gudiksen, D. Wang, C.M. Lieber: Epitaxial core-shell and core-multishell nanowire heterostructures, Nature **420**, 57–61 (2002)

9.152 L. Venkataraman, C.M. Lieber: Molybdenum selenide molecular wires as one-dimensional conductors, Phys. Rev. Lett. **83**, 5334–5337 (1999)

9.153 A. Majumdar: Scanning thermal microscopy, Annu. Rev. Mater. Sci. **29**, 505–585 (1999)

9.154 K.M. Unruh, T.E. Huber, C.A. Huber: Melting and freezing behavior of indium metal in porous glasses, Phys. Rev. B **48**, 9021–9027 (1993)

9.155 Y.Y. Wu, P.D. Yang: Melting and welding semiconductor nanowires in nanotubes, Adv. Mater. **13**, 520–523 (2001)

9.156 P.M. Ajayan, S. Iijima: Capillarity-induced filling of carbon nanotubes, Nature **361**, 333–334 (1993)

9.157 Y. Gao, Y. Bando: Carbon nanothermometer containing gallium, Nature **415**, 599 (2002)

9.158 M.E. Toimil-Molares, A.G. Balogh, T.W. Cornelius, R. Neumann, C. Trautmann: Fragmentation of nanowires driven by Rayleigh instability, Appl. Phys. Lett. **84**, 5337–5339 (2004)

9.159 J.L. Costa-Krämer, N. Garcia, H. Olin: Conductance quantization histograms of gold nanowires at 4 K, Phys. Rev. B **55**, 12910–12913 (1997)

9.160 D.A. Wharam, T.J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J.E.F. Frost, D.G. Hasko, D.C. Peacock, D.A. Ritchie, G.A.C. Jones: One-dimensional transport and the quantization of the ballistic resistance, J. Phys. C **21**, L209–L214 (1988)

9.161 B.J. van Wees, H. van Houten, C.W.J. Beenakker, J.G. Williamson, L.P. Kouvenhoven, D. van der Marel, C.T. Foxon: Quantized conductance of point contacts in a two-dimensional electron gas, Phys.

Rev. Lett. **60**, 848–850 (1988)
9.162 C.J. Muller, J.M. van Ruitenbeek, L.J. de Jongh: Conductance and supercurrent discontinuities in atomic-scale metallic constrictions of variable width, Phys. Rev. Lett. **69**, 140–143 (1992)
9.163 C.J. Muller, J.M. Krans, T.N. Todorov, M.A. Reed: Quantization effects in the conductance of metallic contacts at room temperature, Phys. Rev. B **53**, 1022–1025 (1996)
9.164 J.L. Costa-Krämer, N. Garcia, H. Olin: Conductance quantization in bismuth nanowires at 4 K, Phys. Rev. Lett. **78**, 4990–4993 (1997)
9.165 C.Z. Li, H.X. He, A. Bogozi, J.S. Bunch, N.J. Tao: Molecular detection based on conductance quantization of nanowires, Appl. Phys. Lett. **76**, 1333–1335 (2000)
9.166 J.L. Costa-Krämer, N. Garcia, P. Garcia-Mochales, P.A. Serena, M.I. Marques, A. Correia: Conductance quantization in nanowires formed between micro and macroscopic metallic electrodes, Phys. Rev. B **55**, 5416–5424 (1997)
9.167 Y. Huang, X. Duan, Y. Cui, L.J. Lauhon, K.-H. Kim, C.M. Lieber: Logic gates and computation from assembled nanowire building blocks, Science **294**, 1313–1317 (2001)
9.168 J.-R. Kim, H. Oh, H.M. So, J.-J. Kim, J. Kim, C.J. Lee, S.C. Lyu: Schottky diodes based on a single GaN nanowire, Nanotechnology **13**, 701–704 (2002)
9.169 X. Duan, Y. Huang, C.M. Lieber: Nonvolatile memory and programmable logic from molecule-gated nanowires, Nano Lett. **2**, 487–490 (2002)
9.170 E.C. Walter, R.M. Penner, H. Liu, K.H. Ng, M.P. Zach, F. Favier: Sensors from electrodeposited metal nanowires, Surf. Interface Anal. **34**, 409–412 (2002)
9.171 E.C. Walter, K.H. Ng, M.P. Zach, R.M. Penner, F. Favier: Electronic devices from electrodeposited metal nanowires, Microelectron. Eng. **61/62**, 555–561 (2002)
9.172 Y.-M. Lin, X. Sun, M.S. Dresselhaus: Theoretical investigation of thermoelectric transport properties of cylindrical Bi nanowires, Phys. Rev. B **62**, 4610–4623 (2000)
9.173 K. Liu, C.L. Chien, P.C. Searson: Finite-size effects in bismuth nanowires, Phys. Rev. B **58**, R14681–R14684 (1998)
9.174 Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, J. Heremans: Magnetotransport investigations of ultrafine single-crystalline bismuth nanowire arrays, Appl. Phys. Lett. **73**, 1589–1591 (1998)
9.175 J. Heremans, C.M. Thrush, Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, D.T. Morelli: Magnetoresistance of bismuth nanowire arrays: A possible transition from one-dimensional to three-dimensional localization, Phys. Rev. B **58**, R10091–R10095 (1998)
9.176 L. Sun, P.C. Searson, C.L. Chien: Finite-size effects in nickel nanowire arrays, Phys. Rev. B **61**, R6463–R6466 (2000)
9.177 Y.-M. Lin, S.B. Cronin, O. Rabin, J.Y. Ying, M.S. Dresselhaus: Transport properties and observation of semimetal-semiconductor transition in Bi-based nanowires. In: *Quantum Confined Semiconductor Nanostructures: MRS Symposium Proceedings, Boston*, Vol. 737-C, ed. by J.M. Buriak, D.D.M. Wayner, F. Priolo, B. White, V. Klimov, L. Tsybeskov (Materials Research Society, Pittsburgh 2003) p. F3.14
9.178 Y.-M. Lin, M.S. Dresselhaus: Transport properties of superlattice nanowires and their potential for thermoelectric applications. In: *Quantum Confined Semiconductor Nanostructures: MRS Symposium Proceedings, Boston*, Vol. 737-C, ed. by J.M. Buriak, D.D.M. Wayner, F. Priolo, B. White, V. Klimov, L. Tsybeskov (Materials Research Society Press, Pittsburgh 2003) p. F8.18
9.179 Y.-M. Lin, O. Rabin, S.B. Cronin, J.Y. Ying, M.S. Dresselhaus: Semimetal-semiconductor transition in $Bi_{1-x}Sb_x$ alloy nanowires and their thermoelectric properties, Appl. Phys. Lett. **81**, 2403–2405 (2002)
9.180 J. Heremans, C.M. Thrush, Y.-M. Lin, S.B. Cronin, M.S. Dresselhaus: Transport properties of antimony nanowires, Phys. Rev. B **63**, 085406-1–085406-8 (2001)
9.181 Y.-M. Lin, S.B. Cronin, O. Rabin, J.Y. Ying, M.S. Dresselhaus: Transport properties of $Bi_{1-x}Sb_x$ alloy nanowires synthesized by pressure injection, Appl. Phys. Lett. **79**, 677–679 (2001)
9.182 D.E. Beutler, N. Giordano: Localization and electron-electron interaction effects in thin Bi wires and films, Phys. Rev. B **38**, 8–19 (1988)
9.183 Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, J. Heremans: Electronic transport properties of single crystal bismuth nanowire arrays, Phys. Rev. B **61**, 4850–4861 (2000)
9.184 J. Heremans, C.M. Thrush: Thermoelectric power of bismuth nanowires, Phys. Rev. B **59**, 12579–12583 (1999)
9.185 D. Li, S.T. Huxtable, A.R. Abramson, A. Majumdar: Thermal transport in nanostructured solid-state cooling devices, J. Heat Transfer **127**, 108–114 (2005) doi:10.1115/1.1839588
9.186 D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, A. Majumdar: Thermal conductivity of individual silicon nanowires, Appl. Phys. Lett. **83**, 2934–2936 (2003)
9.187 A.I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J.-K. Yu, W.A. Goddard III, J.R. Heath: Silicon nanowires as efficient thermoelectric materials, Nature **451**, 168–171 (2008) doi:10.1038/nature06458
9.188 A.I. Hochbaum, R. Chen, R.D. Delgado, W. Liang, E.C. Garnett, M. Najarian, A. Majumdar, P. Yang: Enhanced thermoelectric performance of rough silicon nanowires, Nature **451**, 163–167 (2008) doi:10.1038/nature06381
9.189 D. Li, Y. Wu, R. Fan, P. Yang, A. Majumdar: Thermal conductivity of Si/SiGe superlattice nanowires, Appl. Phys. Lett. **83**, 3186–3188 (2003)
9.190 S.C. Andrews, M.A. Fardy, M.C. Moore, S. Aloni, M. Zhang, V. Radmilovic, P. Yang: Atomic-level control of the thermoelectric properties in polytypoid nanowires, Chem. Sci. **2**, 706–714 (2011) doi:10.1039/C0SC00537A
9.191 F. Zhou, A.L. Moore, M.T. Pettes, Y. Lee, J.H. Seol, Q.L. Ye, L. Rabenberg, L. Shi: Effect of growth base pressure on the thermoelectric properties of

indium antimonide nanowires, J. Phys. D Appl. Phys. **43**, 025406 (2010) doi:10.1088/0022-3727/43/2/025406
9.192 T.S. Tighe, J.M. Worlock, M.L. Roukes: Direct thermal conductance measurements on suspended monocrystalline nanostructures, Appl. Phys. Lett. **70**, 2687–2689 (1997)
9.193 S.T. Huxtable, A.R. Abramson, C.-L. Tien, A. Majumdar, C. LaBounty, X. Fan, G. Zeng, J.E. Bowers, A. Shakouri, E.T. Croke: Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices, Appl. Phys. Lett. **80**, 1737–1739 (2002)
9.194 L. Shi, C. Yu, J. Zhou: Thermal characterization and sensor applications of one-dimensional nanostructures employing microelectromechanical systems, J. Phys. Chem. B **109**, 22102–22111 (2005)
9.195 R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn: Thin-film thermoelectric devices with high room-temperature figures of merit, Nature **413**, 597–602 (2001)
9.196 C. Dames, G. Chen: Theoretical phonon thermal conductivity of Si-Ge superlattice nanowires, J. Appl. Phys. **95**, 682–693 (2004)
9.197 C. Dames, G. Chen: Modeling the thermal conductivity of a SiGe segmented nanowire. In: *21st Int. Conf. Thermoelectr. Proc. ICT '02, Long Beach* (IEEE, Piscataway 2002) pp. 317–320
9.198 N. Mingo, L. Yang, D. Li, A. Majumdar: Predicting the thermal conductivity of Si and Ge nanowires, Nano Lett. **3**, 1713–1716 (2003) doi:10.1021/nl034721i
9.199 G. Chen, M.S. Dresselhaus, G. Dresselhaus, J.-P. Fleurial, T. Caillat: Recent developments in thermoelectric materials, Int. Mater. Rev. **48**, 45–66 (2003)
9.200 M. Kazan, G. Guisbiers, S. Pereira, M.R. Correia, P. Masri, A. Bruyant, S. Volz, P. Royer: Thermal conductivity of silicon bulk and nanowires: Effects of isotopic composition, phonon confinement, and surface roughness, J. Appl. Phys. **107**, 083503 (2010) doi:10.1063/1.3340973
9.201 K. Schwab, J.L. Arlett, J.M. Worlock, M.L. Roukes: Thermal conductance through discrete quantum channels, Physica E **9**, 60–68 (2001)
9.202 K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes: Measurement of the quantum of thermal conductance, Nature **404**, 974–977 (2000)
9.203 L.D. Hicks, M.S. Dresselhaus: Thermoelectric figure of merit of a one-dimensional conductor, Phys. Rev. B **47**, 16631–16634 (1993)
9.204 O. Rabin, Y.-M. Lin, M.S. Dresselhaus: Anomalously high thermoelectric figure of merit in $Bi_{1-x}Sb_x$ nanowires by carrier pocket alignment, Appl. Phys. Lett. **79**, 81–83 (2001)
9.205 Y.M. Zuev, J.S. Lee, C. Galloy, H. Park, P. Kim: Diameter dependence of the transport properties of antimony telluride nanowires, Nano Lett. **10**, 3037–3040 (2010)
9.206 W. Liang, A.I. Hochbaum, M. Fardy, O. Rabin, M. Zhang, P. Yang: Field-effect modulation of Seebeck coefficient in single PbSe nanowires, Nano Lett. **9**, 1689–1693 (2009)
9.207 Y. Tian, M.R. Sakr, J.M. Kinder, D. Liang, M.J. MacDonald, R.L.J. Qiu, H.-J. Gao, X.P.A. Gao: One-dimensional quantum confinement effect modulated thermoelectric properties in InAs nanowires, Nano Lett. **12**, 6492–6497 (2012)
9.208 P.M. Wu, J. Gooth, X. Zianni, S.F. Svensson, J.G. Gluschke, K.A. Dick, C. Thelander, K. Nielsch, H. Linke: Large thermoelectric power factor enhancement observed in InAs nanowires, Nano Lett. **13**, 4080–4086 (2013)
9.209 J.E. Cornett, O. Rabin: Thermoelectric figure of merit calculations for semiconducting nanowires, Appl. Phys. Lett. **98**, 182104 (2011)
9.210 J.E. Cornett, O. Rabin: Universal scaling relations for the thermoelectric power factor of semiconducting nanostructures, Phys. Rev. B **84**, 205410 (2011)
9.211 Y. Pei, X. Shi, A. LaLonde, H. Wang, L. Chen, G.J. Snyder: Convergence of electronic bands for high performance bulk thermoelectrics, Nature **473**, 66–69 (2011)
9.212 J. Tang, H. Wang, D.H. Lee, M. Fardy, Z. Huo, T.P. Russell, P. Yang: Holey silicon as an efficient thermoelectric material, Nano Lett. **10**, 4279–4283 (2010)
9.213 L.D. Hicks, M.S. Dresselhaus: The effect of quantum well structures on the thermoelectric figure of merit, Phys. Rev. B **47**, 12727–12731 (1993)
9.214 T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. LaForge: Quantum dot superlattice thermoelectric materials and devices, Science **297**, 2229–2232 (2002)
9.215 Y.-M. Lin, M.S. Dresselhaus: Thermoelectric properties of superlattice nanowires, Phys. Rev. B **68**, 075304 (2003)
9.216 S.R. Nicewarner-Peña, R.G. Freeman, B.D. Reiss, L. He, D.J. Peña, I.D. Walton, R. Cromer, C.D. Keating, M.J. Natan: Submicrometer metallic barcodes, Science **294**, 137–141 (2001)
9.217 M.T. Björk, B.J. Ohlsson, C. Thelander, A.I. Persson, K. Deppert, L.R. Wallenberg, L. Samuelson: Nanowire resonant tunneling diodes, Appl. Phys. Lett. **81**, 4458–4460 (2002)
9.218 M. Cardona: *Light Scattering in Solids* (Springer, Berlin, Heidelberg 1982)
9.219 P.Y. Yu, M. Cardona: *Fundamentals of Semiconductors* (Springer, Berlin, Heidelberg 1995), Chap. 7
9.220 J.C.M. Garnett: Colours in metal glasses, in metallic films, and in metallic solutions, Philos. Trans. R. Soc. A **205**, 237–288 (1906)
9.221 D.E. Aspnes: Optical properties of thin films, Thin Solid Films **89**, 249–262 (1982)
9.222 U. Kreibig, L. Genzel: Optical absorption of small metallic particles, Surf. Sci. **156**, 678–700 (1985)
9.223 M.R. Black, Y.-M. Lin, S.B. Cronin, O. Rabin, M.S. Dresselhaus: Infrared absorption in bismuth nanowires resulting from quantum confinement, Phys. Rev. B **65**, 195417–1–195417–9 (2002)
9.224 Veselago 1967 Usp. Fiz. Nauk 92 517
9.225 A. Poddubny, I. Iorsh, P. Belov, Y. Kivshar: Hyperbolic metamaterials, Nat. Photonics **7**, 948–957 (2013)
9.226 J. Yao, Z. Liu, Y. Liu, Y. Wang, C. Sun, G. Bartal, A.M. Stacy, X. Zhang: Optical negative refraction

in bulk metamaterials of nanowires, Science **321**, 930–930 (2008)
9.227 M.W. Lee, H.Z. Twu, C.-C. Chen, C.H. Chen: Optical characterization of wurtzite gallium nitride nanowires, Appl. Phys. Lett. **79**, 3693–3695 (2001)
9.228 M.S. Gudiksen, J. Wang, C.M. Lieber: Size-depent photoluminescence from single indium phosphide nanowires, J. Phys. Chem. B **106**, 4036–4039 (2002)
9.229 D.M. Lyons, K.M. Ryan, M.A. Morris, J.D. Holmes: Tailoring the optical properties of silicon nanowire arrays through strain, Nano Lett. **2**, 811–816 (2002)
9.230 S. Bhattacharya, D. Banerjee, K.W. Adu, S. Samui, S. Bhattacharyya: Confinement in silicon nanowires: Optical properties, Appl. Phys. Lett. **85**, 2008–2010 (2004) doi:10.1063/1.1787164
9.231 J.C. Johnson, H. Yan, R.D. Schaller, L.H. Haber, R.J. Saykally, P. Yang: Single nanowire lasers, J. Phys. Chem. B **105**, 11387–11390 (2001)
9.232 S. Blom, L.Y. Gorelik, M. Jonson, R.I. Shekhter, A.G. Scherbakov, E.N. Bogachek, U. Landman: Magneto-optics of electronic transport in nanowires, Phys. Rev. B **58**, 16305–16314 (1998)
9.233 J.P. Pierce, E.W. Plummer, J. Shen: Ferromagnetism in cobalt-iron alloy nanowire arrays on w(110), Appl. Phys. Lett. **81**, 1890–1892 (2002)
9.234 S. Melle, J.L. Menendez, G. Armelles, D. Navas, M. Vazquez, K. Nielsch, R.B. Wehrsphon, U. Gösele: Magneto-optical properties of nickel nanowire arrays, Appl. Phys. Lett. **83**, 4547–4549 (2003)
9.235 J.C. Johnson, H. Yan, R.D. Schaller, P.B. Petersen, P. Yang, R.J. Saykally: Near-field imaging of nonlinear optical mixing in single zinc oxide nanowires, Nano Lett. **2**, 279–283 (2002)
9.236 M.R. Black, P.L. Hagelstein, S.B. Cronin, Y.-M. Lin, M.S. Dresselhaus: Optical absorption from an indirect transition in bismuth nanowires, Phys. Rev. B **68**, 235417 (2003)
9.237 M.R. Black, Y.-M. Lin, S.B. Cronin, M.S. Dresselhaus: Using optical measurements to improve electronic models of bismuth nanowires. In: *21st Int. Conf. Thermoelectr. Proc. ICT '02, Long Beach*, ed. by T. Caillat, J. Snyder (IEEE, Piscataway 2002) pp. 253–256
9.238 H. Richter, Z.P. Wang, L. Ley: The one phonon Raman-spectrum in microcrystalline silicon, Solid State Commun. **39**, 625–629 (1981)
9.239 I.H. Campbell, P.M. Fauchet: The effects of microcrystal size and shape on the one phonon Raman-spectra of crystalline semiconductors, Solid State Commun. **58**, 739–741 (1986)
9.240 H.-L. Liu, C.-C. Chen, C.-T. Chia, C.-C. Yeh, C.-H. Chen, M.-Y. Yu, S. Keller, S.P. DenBaars: Infrared and Raman-scattering studies in single-crystalline GaN nanowires, Chem. Phys. Lett. **345**, 245–251 (2001)
9.241 R. Gupta, Q. Xiong, C.K. Adu, U.J. Kim, P.C. Eklund: Laser-induced Fano resonance scattering in silicon nanowires, Nano Lett. **3**, 627–631 (2003)
9.242 X. Duan, Y. Huang, Y. Cui, J. Wang, C.M. Lieber: Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices, Nature **409**, 66–69 (2001)
9.243 Y. Cui, C.M. Lieber: Functional nanoscale electronic devices assembled using silicon nanowire building blocks, Science **291**, 851–853 (2001)
9.244 Y. Cui, X. Duan, J. Hu, C.M. Lieber: Doping and electrical transport in silicon nanowires, J. Phys. Chem. B **104**, 101–104 (2000)
9.245 G.F. Zheng, W. Lu, S. Jin, C.M. Lieber: Synthesis and fabrication of high-performance n-type silicon nanowire transistors, Adv. Mater. **16**, 1890–1891 (2004)
9.246 J. Goldberger, D.J. Sirbuly, M. Law, P. Yang: ZnO nanowire transistors, J. Phys. Chem. B **109**, 9–14 (2005)
9.247 D.H. Kang, J.H. Ko, E. Bae, J. Hyun, W.J. Park, B.K. Kim, J.J. Kim, C.J. Lee: Ambient air effects on electrical characteristics of gap nanowire transistors, J. Appl. Phys. **96**, 7574–7577 (2004)
9.248 S.-W. Chung, J.-Y. Yu, J.R. Heath: Silicon nanowire devices, Appl. Phys. Lett. **76**, 2068–2070 (2000)
9.249 C. Li, W. Fan, B. Lei, D. Zhang, S. Han, T. Tang, X. Liu, Z. Liu, S. Asano, M. Meyyappan, J. Han, C. Zhou: Multilevel memory based on molecular devices, Appl. Phys. Lett. **84**, 1949–1951 (2004)
9.250 B. Lei, C. Li, D.Q. Zhang, Q.F. Zhou, K. Shung, C.W. Zhou: Nanowire transistors with ferroelectric gate dielectrics: Enhanced performance and memory effects, Appl. Phys. Lett. **84**, 4553–4555 (2004)
9.251 H.T. Ng, J. Han, T. Yamada, P. Nguyen, Y.P. Chen, M. Meyyappan: Single crystal nanowire vertical surround-gate field-effect transistor, Nano Lett. **4**, 1247–1252 (2004)
9.252 J.-P. Colinge, C.-W. Lee, A. Afzalian, N.D. Akhavan, R. Yan, I. Ferain, P. Razavi, B. O'Neill, A. Blake, M. White, A.-M. Kelleher, B. McCarthy, R. Murphy: Nanowire transistors without junctions, Nat. Nanotechnol. **5**, 225–229 (2010) doi:10.1038/nnano.2010.15
9.253 J.P. Colinge, A. Kranti, R. Yan, C.W. Lee, I. Ferain, R. Yu, A.N. Dehdashti, P. Razavi: Junctionless nanowire transistor (JNT): Properties and design guidelines, Solid-State Electron. **65/66**, 33–37 (2011) doi:10.1016/j.sse.2011.06.004
9.254 W. Wu, X. Wen, Z.L. Wang: Taxel-addressable matrix of vertical-nanowire piezotronic transistors for active and adaptive tactile imaging, Science **340**, 952–957 (2013)
9.255 M. Ding, H. Kim, A.I. Akinwande: Observation of valence band electron emission from n-type silicon field emitter arrays, Appl. Phys. Lett. **75**, 823–825 (1999)
9.256 F.-H. Chu, C.-W. Huang, C.-L. Hsin, C.-W. Wang, S.-Y. Yu, P.-H. Yeh, W.-W. Wu: Well-aligned ZnO nanowires with excellent field emission and photocatalytic properties, Nanoscale **4**, 1471–1475 (2012)
9.257 F.C.K. Au, K.W. Wong, Y.H. Tang, Y.F. Zhang, I. Bello, S.T. Lee: Electron field emission from silicon nanowires, Appl. Phys. Lett. **75**, 1700–1702 (1999)
9.258 P.M. Ajayan, O.Z. Zhou: Applications of carbon nanotubes. In: *Carbon Nanotubes: Synthesis, Structure, Properties and Applications*, Springer Ser. Top. Appl. Phys., Vol. 80, ed. by M.S. Dresselhaus, G. Dresselhaus, P. Avouris (Springer, Berlin, Heidel-

berg 2001) pp. 391–425
9.259 M. Lu, M.K. Li, Z.J. Zhang, H.L. Li: Synthesis of carbon nanotubes/si nanowires core-sheath structure arrays and their field emission properties, Appl. Surf. Sci. **218**, 196–202 (2003)
9.260 L. Vila, P. Vincent, L. Dauginet-DePra, G. Pirio, E. Minoux, L. Gangloff, S. Demoustier-Champagne, N. Sarazin, E. Ferain, R. Legras, L. Piraux, P. Legagneux: Growth and field-emission properties of vertically aligned cobalt nanowire arrays, Nano Lett. **4**, 521–524 (2004)
9.261 J.Y. Lee, S.T. Connor, Y. Cui, P. Peumans: Solution-processed metal nanowire mesh transparent electrodes, Nano Lett. **8**, 689–692 (2008) doi:10.1021/nl073296g
9.262 X. Duan, Y. Huang, R. Agarwal, C.M. Lieber: Single-nanowire electrically driven lasers, Nature **421**, 241 (2003)
9.263 F. Qian, Y. Li, S. Gradecak, D.L. Wang, C.J. Barrelet, C.M. Lieber: Gallium nitride-based nanowire radial heterostructures for nanophotonics, Nano Lett. **4**, 1975–1979 (2004)
9.264 V. Dneprovskii, E. Zhukov, V. Karavanskii, V. Poborchii, I. Salamatini: Nonlinear optical properties of semiconductor quantum wires, Superlattice. Microst. **23**(6), 1217–1221 (1998)
9.265 J.C. Johnson, K.P. Knutsen, H. Yan, M. Law, Y. Zhang, P. Yang, R.J. Saykally: Ultrafast carrier dynamics in single ZnO nanowire and nanoribbon lasers, Nano Lett. **4**, 197–204 (2004)
9.266 J.X. Ding, J.A. Zapien, W.W. Chen, Y. Lifshitz, S.T. Lee, X.M. Meng: Lasing in ZnS nanowires grown on anodic aluminum oxide templates, Appl. Phys. Lett. **85**, 2361 (2004)
9.267 J.C. Johnson, H.-J. Choi, K.P. Knutsen, R.D. Schaller, P. Yang, R.J. Saykally: Single gallium nitride nanowire lasers, Nat. Mater. **1**, 106–110 (2002)
9.268 H.J. Choi, J.C. Johnson, R. He, S.K. Lee, F. Kim, P. Pauzauskie, J. Goldberger, R.J. Saykally, P. Yang: Self-organized GaN quantum wire UV lasers, J. Phys. Chem. B **107**, 8721–8725 (2003)
9.269 C.J. Barrelet, A.B. Greytak, C.M. Lieber: Nanowire photonic circuit elements, Nano Lett. **4**, 1981–1985 (2004)
9.270 M. Law, D.J. Sirbuly, J.C. Johnson, J. Goldberger, R.J. Saykally, P. Yang: Ultralong nanoribbon waveguides for sub-wavelength photonics integration, Science **305**, 1269–1273 (2004)
9.271 H. Kind, H. Yan, B. Messer, M. Law, P. Yang: Nanowire ultraviolet photodetectors and optical switches, Adv. Mater. **14**, 158–160 (2002)
9.272 B.M.I. van der Zande, M.R. Böhmer, L.G.J. Fokkink, C. Schöneberger: Colloidal dispersions of gold rods: Synthesis and optical properties, Langmuir **16**, 451–458 (2000)
9.273 B.M.I. van der Zande, G.J.M. Koper, H.N.W. Lekkerkerker: Alignment of rod-shaped gold particles by electric fields, J. Phys. Chem. B **103**, 5754–5760 (1999)
9.274 W.U. Huynh, J.J. Dittmer, A.P. Alivisatos: Hybrid nanorod-polymer solar cells, Science **295**, 2425–2427 (2002)
9.275 I. Åberg, G. Vescovi, D. Asoli: A GaAs nanowire array solar cell with 15.3% efficiency at 1 Sun, IEEE J. Photovolt. **6**, 185–190 (2016)
9.276 J. Wallentin, N. Anttu, D. Asoli, M. Huffman, I. Åberg, M.H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzigmann, H.Q. Xu, L. Samuelson, K. Deppert, M.T. Borgström: InP nanowire array solar cells achieving 13.8% efficiency by exceeding the ray optics limit, Science **339**, 1057–1060 (2013)
9.277 B.X. Tian, T.J. Zheng: Kempa, Y. Fang, N. Yu, G. Yu, J. Huang, C. M. Lieber: Coaxial silicon nanowires as solar cells and nanoelectronic power sources, Nature **449**, 885–889 (2007)
9.278 P. Krogstrup, H.I. Jørgensen, M. Heiss, O. Demichel, J.V. Holm, M. Aagesen, J. Nygard: A. Fontcuberta i Morral: Single-nanowire solar cells beyond the Shockley-Queisser limit, Nat. Photon. **7**, 306–310 (2013)
9.279 B. A. Buchine, F. Modawar, M. R. Black: Nanostructured Devices. US Patent No. 20,100,122,725. (2010)
9.280 L. Hu, G. Chen: Analysis of optical absorption in silicon nanowire arrays for photovoltaic applications, Nano Lett. **7**, 3249–3252 (2007)
9.281 C.K. Chan, H. Peng, G. Liu, K. McIlwrath, X.F. Zhang, R.A. Huggins, Y. Cui: High-performance lithium battery anodes using silicon nanowires, Nat. Nanotechnol. **3**, 31–35 (2008)
9.282 L. Mai, Y. Dong, L. Xu, C. Han: Single nanowire electrochemical devices, Nano Lett. **10**, 4273–4278 (2010) doi:10.1021/nl102845r
9.283 V.P. Oleshko, T. Lam, D. Ruzmetov, P. Haney, H.J. Lezec, A.V. Davydov, S. Krylyuk, J. Cumings, A.A. Talin: Miniature all-solid-state heterostructure nanowire Li-ion batteries as a tool for engineering and structural diagnostics of nanoscale electrochemical processes, Nanoscale **6**(20), 11756 (2014)
9.284 M. Green: U.S. Patent No. 7402829 (U.S. Patent and Trademark Office, Washington DC 2008)
9.285 B. A. Buchine, F. Modawar, M. R. Black: Nanostructured silicon for battery anodes, U.S. Patent No. 8791449. (2014)
9.286 G. E. Loveness, C.I. Stefan, S. Han: Multidimensional electrochemically active structures for battery electrodes, U.S. Patent No. 9172088. (2015)
9.287 C. Liu, E.I. Gillette, X. Chen, A.J. Pearse, A.C. Kozen, M.A. Schroeder, K.E. Gregorczyk, S.L. Bok, G.W. Rubloff: An all-in-one nanopore battery array, Nat. Nanotechnol. **9**, 1031–1039 (2014) doi:10.1038/nnano.2014.247
9.288 J. Qiu, G. Zeng, M.-A. Ha, M. Ge, Y. Lin, M. Hettick, B. Hou, A.N. Alexandrova, A. Javey, S.B. Cronin: Artificial photosynthesis on TiO_2-passivated InP nanopillars, Nano Lett. **15**, 6177–6181 (2015)
9.289 Y. Su, C. Liu, S. Brittman, J. Tang, A. Fu, N. Kornienko, Q. Kong, P. Yang: Single-nanowire photoelectrochemistry, Nat. Nanotechnol. **11**, 609–612 (2016) doi:10.1038/nnano.2016.30
9.290 P. Banerjee, I. Perez, L. Henn-Lecordier, S.B. Lee, G.W. Rubloff: Nanotubular metal–insulator–metal capacitor arrays for energy storage, Nat. Nanotech-

nol. **4**(5), 292–296 (2009)
9.291 Z. Liu, Y. Zhan, G. Shi, S. Moldovan, M. Gharbi, L. Song, L. Ma, W. Gao, J. Huang, R. Vajtai, F. Banhart, P. Sharma, J. Lou, P.M. Ajayan: Anomalous high capacitance in a coaxial single nanowire capacitor, Nat. Commun. (2012) doi:10.1038/ncomms1833
9.292 P.M. Rorvik, T. Grande, M.A. Einarsrud: One-dimensional nanostructures of ferroelectric perovskites, Adv. Mater. **23**, 4007–4034 (2011) doi:10.1002/adma.201004676
9.293 Z.L. Wang: From nanogenerators to piezotronics–A decade-long study of ZnO nanostructures, MRS Bull. **37**, 814–827 (2012) doi:10.1557/mrs.2012.186
9.294 Z.L. Wang, J.H. Song: Piezoelectric nanogenerators based on zinc oxide nanowire arrays, Science **312**, 242–246 (2006) doi:10.1126/science.1124005
9.295 L. Gu, N. Cui, L. Cheng, Q. Xu, S. Bai, M. Yuan, W. Wu, J. Liu, Y. Zhao, F. Ma, Y. Qin, Z. Lin Wang: Flexible fiber nanogenerator with 209 V output voltage directly powers a light-emitting diode, Nano Lett. **13**, 91–94 (2013) doi:10.1021/nl303539c
9.296 J.H. Jung, M. Lee, J.-I. Hong, Y. Ding, C.-Y. Chen, L.-J. Chou, Z.L. Wang: Lead-free $NaNbO_3$ nanowires for a high output piezoelectric nanogenerator, Acs Nano **5**, 10041–10046 (2011) doi:10.1021/nn2039033
9.297 J. Kwon, W. Seung, B.K. Sharma, S.W. Kim, J.H. Ahn: A high performance PZT ribbon-based nanogenerator using graphene transparent electrodes, Energy Environ. Sci. **5**, 8970–8975 (2012) doi:10.1039/c2ee22251e
9.298 C. Falconi, G. Mantinia, A. D'Amico, Z.L. Wang: Studying piezoelectric nanowires and nanowalls for energy harvesting, Sens. Actuat. B-Chem. **139**, 511–519 (2009) doi:10.1016/j.snb.2009.02.071
9.299 G. Mantini, Y.F. Gao, A. D'Amico, C. Falconi, Z.L. Wang: Equilibrium piezoelectric potential distribution in a deformed ZnO nanowire, Nano Res **2**, 624–629 (2009) doi:10.1007/s12274-009-9063-2
9.300 Z. Wang, X. Pan, Y. He, Y. Hu, H. Gu, Y. Wang: Piezoelectric nanowires in energy harvesting applications, Adv. Mater. Sci. Eng. (2015) doi:10.1155/2015/165631
9.301 H.D. Espinosa, R.A. Bernal, M. Minary-Jolandan: A review of mechanical and electromechanical properties of piezoelectric nanowires, Adv. Mater. **24**, 4656–4675 (2012) doi:10.1002/adma.201104810
9.302 C.-T. Huang, J. Song, W.-F. Lee, Y. Ding, Z. Gao, Y. Hao, L.-J. Chen, Z.L. Wang: GaN nanowire arrays for high-output nanogenerators, J. Am. Chem. Soc. **132**, 4766–4771 (2010) doi:10.1021/ja909863a
9.303 C.T. Huang, J. Song, C.-M. Tsai, W.-F. Lee, D.-H. Lien, Z. Gao, Y. Hao, L.-J. Chen, Z.L. Wang: Single-InN-nanowire nanogenerator with upto 1 V output voltage, Adv. Mater. **22**, 4008–4013 (2010) doi:10.1002/adma.201000981
9.304 Y.F. Lin, J. Song, Y. Ding, S.Y. Lu, Z.L. Wang: Alternating the output of a CdS nanowire nanogenerator by a white-light-stimulated optoelectronic effect, Adv. Mater. **20**, 3127–3130 (2008) doi:10.1002/adma.200703236
9.305 M.-Y. Lu, J. Song, M.-P. Lu, C.-Y. Lee, L.-J. Chen, Z.L. Wang: ZnO-ZnS heterojunction and ZnS nanowire arrays for electricity generation, Acs Nano **3**, 357–362 (2009) doi:10.1021/nn800804r
9.306 Y.F. Lin, J. Song, Y. Ding, S.Y. Lu, Z.L. Wang: Piezoelectric nanogenerator using CdS nanowires, Appl. Phys. Lett. **92**, 022105 (2008) doi:10.1063/1.2831901
9.307 S. Xu, Y. Qin, C. Xu, Y. Wei, R. Yang, Z.L. Wang: Self-powered nanowire devices, Nat. Nanotechnol. **5**, 366–373 (2010) doi:10.1038/nnano.2010.46
9.308 S. Xu, B.J. Hansen, Z.L. Wang: Piezoelectric-nanowire-enabled power source for driving wireless microelectronics, Nat. Commun. (2010) doi:10.1038/ncomms1098
9.309 Z.Y. Wang, J. Hu, A.P. Suryavanshi, K. Yum, M.F. Yu: Voltage generation from individual $BaTiO_3$ nanowires under periodic tensile mechanical load, Nano Lett. **7**, 2966–2969 (2007) doi:10.1021/nl070814e
9.310 R.S. Yang, Y. Qin, L.M. Dai, Z.L. Wang: Power generation with laterally packaged piezoelectric fine wires, Nat. Nanotechnol. **4**, 34–39 (2009) doi:10.1038/nnano.2008.314
9.311 Z. Wang, H. Gu, Y. Hu, K. Yang, M. Hu, D. Zhoua, J. Guan: Synthesis, growth mechanism and optical properties of $(K,Na)NbO_3$ nanostructures, Crystengcomm **12**, 3157–3162 (2010) doi:10.1039/c000169d
9.312 G.A. Zhu, R.S. Yang, S.H. Wang, Z.L. Wang: Flexible high-output nanogenerator based on lateral ZnO nanowire array, Nano Lett. **10**, 3151–3155 (2010) doi:10.1021/nl101973h
9.313 J.Y. Chang, M. Domnner, C. Chang, L.W. Lin: Piezoelectric nanofibers for energy scavenging applications, Nano Energy **1**, 356–371 (2012) doi:10.1016/j.nanoen.2012.02.003
9.314 X. Chen, S.Y. Xu, N. Yao, Y. Shi: 1.6 V nanogenerator for mechanical energy harvesting using PZT nanofibers, Nano Lett. **10**, 2133–2137 (2010) doi:10.1021/nl100812k
9.315 Y.F. Hu, Y. Zhang, C. Xu, G.A. Zhu, Z.L. Wang: High-output nanogenerator by rational unipolar assembly of conical nanowires and its application for driving a small liquid crystal display, Nano Lett. **10**, 5025–5031 (2010) doi:10.1021/nl103203u
9.316 L. Lin, C. Lai, Y. Hu, Y. Zhang, X. Wang, C. Xu, R.L. Snyder, L.J. Chen, Z.L. Wang: High output nanogenerator based on assembly of GaN nanowires, Nanotechnology (2011) doi:10.1088/0957-4484/22/47/475401
9.317 L.A. Bauer, N.S. Birenbaum, G.J. Meyer: Biological applications of high aspect ratio nanoparticles, J. Mater. Chem. **14**, 517–526 (2004)
9.318 Y. Cui, Q. Wei, H. Park, C. Lieber: Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species, Science **293**, 1289–1292 (2001)
9.319 J. Hahm, C. Lieber: Direct ultra-sensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors, Nano Lett. **4**, 51–54 (2004)
9.320 J. Zhoua, C.S. Lao, P. Gao, W. Mai, W.L. Hughes, S. Zhi Deng, N.S. Xu, Z.L. Wang: Nanowire as picogram balance at workplace atmosphere, Solid State

Commun. **139**, 222–226 (2006)
9.321 M.S. Hanay, S. Kelber, A.K. Naik, D. Chi, S. Hentz, E.C. Bullard, E. Colinet, L. Duraffourg, M.L. Roukes: Single-protein nanomechanical mass spectrometry in real time, Nat. Nanotechnol. **7**, 602–608 (2012) doi:10.1038/nnano.2012.119.
9.322 M. L. Roukes, M. S. Hanay, S. Kelber, A. Naik: US Patent Application No. 13/890,087 (2013)
9.323 D.J. Sirbuly, M. Law, H. Yan, P. Yang: Semiconductor nanowires for subwavelength photonics integration, J. Phys. Chem. B **109**, 15190–15213 (2005)
9.324 R. Yan, J. Park, Y. Choi, C. Heo, S. Yang, L.P. Lee, P. Yang: Nanowire-based single cell endoscopy, Nat. Nanotechnol. **7**, 191–196 (2012)
9.325 Y. Huang, Y. Fang, Z. Zhang, L. Zhu, M. Sun: Nanowire-supported plasmonic waveguide for remote excitation of surface-enhanced Raman scattering, Light Sci. Appl. **3**, e199 (2014) doi:10.1038/lsa.2014.80
9.326 S.S.P. Parkin, M. Hayashi, L. Thomas: Magnetic domain-wall racetrack memory, Science **320**, 190–194 (2008) doi:10.1126/science.1145799
9.327 D.A. Allwood, G. Xiong, C.C. Faulkner, D. Atkinson, D. Petit, R.P. Cowburn: Magnetic domain-wall logic, Science **309**, 1688–1692 (2005) doi:10.1126/science.1108813
9.328 D. Atkinson, D.A. Allwood, G. Xiong, M.D. Cooke, C.C. Faulkner, R.P. Cowburn: Magnetic domain-wall dynamics in a submicrometre ferromagnetic structure, Nat. Mater. **2**, 85–87 (2003) doi:10.1038/nmat803
9.329 R.D. McMichael, M.J. Donahue: Head to head domain wall structures in thin magnetic strips, IEEE Trans. Magnet. **33**, 4167–4169 (1997) doi:10.1109/20.619698
9.330 T. Ono, H. Miyajima, K. Shigeto, K. Mibu, N. Hosoito, T. Shinjo: Propagation of a magnetic domain wall in a submicrometer magnetic wire, Science **284**, 468–470 (1999) doi:10.1126/science.284.5413.468
9.331 D.A. Allwood, G. Xiong, M.D. Cooke, C.C. Faulkner, D. Atkinson, N. Vernier, R.P. Cowburn: Submicrometer ferromagnetic NOT gate and shift register, Science **296**, 2003–2006 (2002) doi:10.1126/science.1070595
9.332 C.C. Faulkner, D.A. Allwood, M.D. Cooke, G. Xiong, D. Atkinson, R.P. Cowburn: Controlled switching of ferromagnetic wire junctions by domain wall injection, IEEE Trans. Magnet. **39**, 2860–2862 (2003) doi:10.1109/tmag.2003.816247

10. Nanoribbons

Toshiaki Enoki, Shintaro Sato

Graphene nanoribbons have intriguing electronic structures, which are large edge geometry dependent. Armchair-edged graphene nanoribbons, which are energetically stable, have a ribbon-width-dependent intrinsic energy gap, while zigzag-edged ones have spin-polarized nonbonding edge states in the vicinity of the edge region. The edge state is the origin of electronic, magnetic and chemical activities. These features of the electronic structures can be characterized using microprobe techniques such as scanning tunneling microscopy/spectroscopy (STM/STS), atomic force microscopy (AFM), transmission electron microscopy (TEM), Raman spectroscopy, x-ray absorption, angle-resolved photoemission spectroscopy, electron transport, and magnetic measurements. Graphene nanostructures are synthesized using top-down and bottom-up methods, in the latter of which graphene nanostructures with atomically precise edges can be created. The presence of bandgap, which varies depending on the ribbon width and the edge geometry, makes graphene an important candidate for electronics device applications. The spin-polarized edge states localized in the vicinity of edges in zigzag-edged nanoribbons are expected to be utilized for spintronics applications.

10.1 **Graphene Nanoribbons** 303
10.2 **Electronic and Magnetic Properties** 305
10.3 **Characterizations** 310
10.4 **Syntheses: Top-Down Methods and Bottom-Up Methods** 317
10.4.1 Graphene Nanoribbon Formation by Top-Down Approaches 317
10.4.2 Graphene Nanoribbon Formation by Bottom-Up Approaches 319
10.4.3 Graphene Nanoribbon Formation by Other Approaches 321
10.5 **Potential Applications** 324
10.5.1 Application to Transistors 324
10.5.2 Interconnect Application 326
10.5.3 Other Potential Applications 328
10.6 **Conclusions** 330
References 331

10.1 Graphene Nanoribbons

A one-atom-thick two-dimensional carbon sheet called graphene was first isolated from graphite in 2004 [10.1]. It consists of a two-dimensional honeycomb lattice of carbon atoms and is regarded to be an infinite-size polycyclic aromatic hydrocarbon molecule derived by fusing benzene molecules in a two-dimensional manner. Graphene is just a precursor sheet from which carbon nanotubes [10.2, 3] or fullerenes [10.2] are created by cutting it into appropriate shapes as shown in Fig. 10.1. Here a carbon nanotube is formed by rolling up a rectangular-shaped graphene sheet into a cylinder, while a fullerene such as C_{60} is obtained by folding an appropriately cut graphene sheet into a ball with all the edges being closed. By cutting a graphene sheet, we can create a variety of nanographenes (graphene nanostructures) having different sizes and shapes. Among these, ribbon-shaped ones are called graphene nanoribbons (GNRs) (Fig. 10.2). As naturally deduced in chemistry, nanographenes are considered as nanosized polycyclic aromatic hydrocarbon molecules. What distinguishes nanographenes from fullerenes and carbon nanotubes is the presence of open edges in the former.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_10

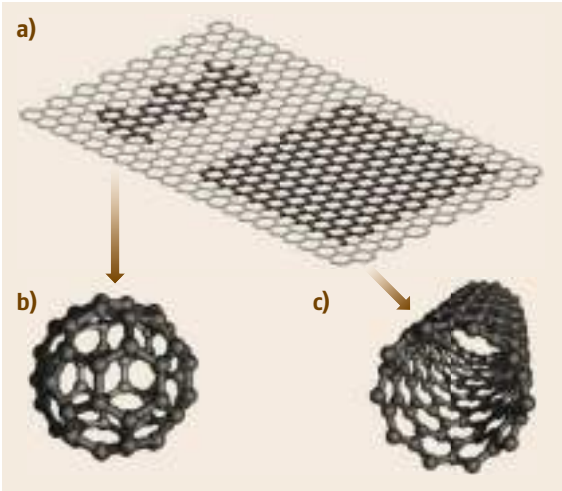


Fig. 10.1a–c Nanographene (**a**), fullerene (**b**) and carbon nanotube (**c**)

The presence of open edges having different geometries crucially contributes to a variety of properties in nanographenes.

The valence electrons, which play an important role in the physical and chemical properties, consist of π-electrons of constituent carbon atoms forming an sp^2-bond network. The π-electrons in an infinite-size graphene is described effectively in terms of massless Dirac fermion in relativistic quantum mechanics running in a two-dimensional honeycomb lattice in physics [10.4–6]. In chemistry we can discuss the behavior of the π-electrons in graphene in terms of aromaticity similarly to those in polycyclic aromatic hydrocarbon molecules [10.5, 6]. Here *Clar*'s aromatic sextet rule can phenomenologically explain their properties comprehensively [10.7, 8]. In graphene nanoarchitects such as carbon nanotubes, fullerenes and nanographenes, the electronic structure of π-electrons, which is modified from graphene, governs their physical and chemical properties. In particular, the electronic structure of nanographene is derived simply by cutting infinite graphene into nanofragments or increasing the number of constituent benzene rings to the nanodimension in polycyclic aromatic hydrocarbon molecules. The electronic structure of nanographene can be obtained by introducing a boundary condition, which is given from the open edges, in the massless Dirac fermion in physics, and also by extending the electronic structure of polycyclic aromatic molecules to the nanodimension in chemistry.

When we cut a graphene sheet, there are two fundamental directions, that is, armchair and zigzag directions, as shown in Fig. 10.2a. Rotating the cutting direction by 30° changes the armchair-zigzag direction to zigzag-armchair direction. Cutting a graphene sheet along these two fundamental directions and subsequently terminating the edge carbon atoms using foreign chemical species such as hydrogen gives two types of edges; that is, armchair edge and zigzag edge [10.5, 6], as exhibited in Fig. 10.2a. By cutting a graphene sheet into nanofragments in general directions, various edge geometries are created, in which armchair edge and zigzag edge are extreme cases. Accordingly, the edge of an arbitrary-shaped nanographene can be described in terms of a combination of armchair and zigzag edges. In the case of graphene nanoribbons, we have two types of ribbons as typical cases; armchair nanoribbon (AGNR) and zigzag nanoribbon (ZGNR) as given in Fig. 10.2b,c respectively. By cutting along an arbitrary direction between armchair and zigzag directions, nanoribbons having chiral edges are created. We can create wiggled nanoribbons depending on how we cut a graphene sheet. The electronic structure, which determines the physical and chemical properties, is essentially different between armchair and zigzag nanoribbons. Armchair nanoribbons are energetically stable with an intrinsic energy gap being opened due to aromatic stability, whereas zigzag nanoribbons are

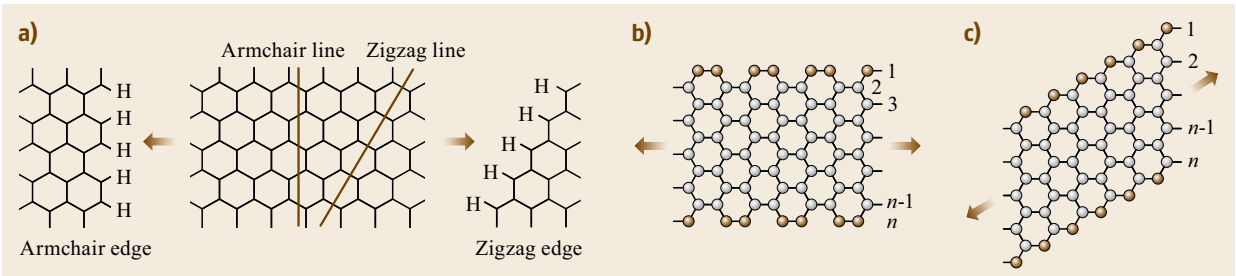


Fig. 10.2 (**a**) Cutting a graphene sheet along the armchair and zigzag directions gives an armchair edge and a zigzag edge respectively. (**b**) Armchair-edged graphene nanoribbon. (**c**) Zigzag-edged graphene nanoribbon. The ribbon width is defined in terms of the number (n) of C−C dimer chains and the number (n) of zigzag chains for armchair and zigzag nanoribbons, respectively; the ribbons are named n-AGNR and n-ZGNR

unstable as they are less aromatic. Interestingly, spin-polarized nonbonding π-electron states localized in the vicinity of zigzag edges are created, which are the origin of electronic, magnetic and chemical activities [10.5, 9]. Therefore, the presence of armchair and zigzag nanoribbons together with chiral-edged nanoribbons brings about a variety of physical and chemical properties.

Apart from edge geometry, how we terminate edge carbon atoms gives rise to another variety in the electronic structures, as functional groups we use in terminating edge carbon atoms affect the electronic structures [10.10]. For example, terminating them with hydrogen atoms gives an electronic structure different from terminating with carbonyl groups. Substituting constituent carbon atoms with heteroatoms such as boron or nitrogen atoms in nanoribbons also brings about a large change in the electronic structure.

There are two fundamental methods in preparing graphene nanoribbons; that is, the top-down method and the bottom-up method [10.5]. In the top-down method, we employ electron beam lithography or microprobe techniques such as scanning tunneling microscopy and atomic force microscopy for cutting graphene sheets into nanoribbons [10.11, 12]. Cutting a carbon nanotube can also create graphene nanoribbons, where zigzag, armchair and chiral tubes of different diameters can be utilized to produce nanoribbons having different edge geometries and widths [10.13]. The drawback in the top-down method is edge roughness, which is unavoidable. In the bottom-up method, precursor monomer molecules are polymerized typically on a metal substrate, which works as catalyst, to prepare nanoribbons. The bottom-up method can create atomically precise edges [10.14]. In addition to this, depending on which precursor monomer we employ, nanoribbons having a variety of edge geometries and chemistries can be obtained. Accordingly the bottom-up method is expected to be available in preparing nanoribbon-based molecular electronics devices.

The variety of electronic properties is a source of various functionalities in nanographenes such as graphene nanoribbons. Therefore, nanographenes will be an important base material for future molecular electronics/spintronics devices. Incorporating atomically precise nanoribbons into electronics devices brings about carbon-based molecular devices in the next generation.

10.2 Electronic and Magnetic Properties

Nanographene has open edges, which importantly contribute to modifying the electronic structure from infinite graphene. Here we have two fundamental edges; armchair and zigzag edges as shown in Fig. 10.2. The modification depends crucially on the edge geometry. We can discuss the electronic structures of nanographene and their edge geometry dependence in terms of both physics and chemistry languages, as they are in between infinite graphene and polycyclic aromatic hydrocarbon molecules, with which physicists and chemists are relatively familiar.

In chemistry, we have *Clar*'s aromatic sextet rule, which is an important phenomenological rule for discussing the electronic and chemical properties of polycyclic aromatic hydrocarbon molecules [10.7, 8]. Using this rule, we can discuss the energetic stability of polycyclic aromatic molecules by tiling aromatic sextets (Fig. 10.3a), which are benzene in the resonance state with all six carbon atoms singly bonded to the surrounding, on a molecule concerned. According to Clar's rule, the most stable structure is that having a maximal number of sextets tiled. We can employ this rule also for nanographenes as they are classified as polycyclic aromatic hydrocarbon molecules, whose sizes are in the nanodimension. Figure 10.3 shows Clar's representations of infinite graphene and several examples of armchair-edged and zigzag-edged polycyclic aromatic molecules. In infinite graphene having a two-dimensional honeycomb hexagonal lattice, aromatic sextets are tiled in every three hexagon ring and the tiled sextets form a $\sqrt{3} \times \sqrt{3}$ superlattice as shown in Fig. 10.3b. It should be noted that there are three Clar's representations degenerate in infinite graphene. So the sextets are migrating between the three states. This is the reason why infinite graphene is electrically conductive from the chemistry aspect.

When a graphene is cut, an edge is created, resulting in a modification of the arrangement of sextets. Figure 10.3c exhibits examples of armchair-edged aromatic molecules, that is, armchair-edged graphene nanofragments. We have large numbers of sextets in these fragments, and consequently their energetic stability (aromatic stability) is confirmed on the basis of Clar's aromatic rule. Note that there is no degeneracy in the Clar's representations, and, instead, only a unique Clar's representation exists for armchair-edged fragments, resulting in the localization of sextets. The localization of sextets suggests that armchair-edged nanographenes are electrically less conductive in contrast to the highly conductive properties of infinite graphene.

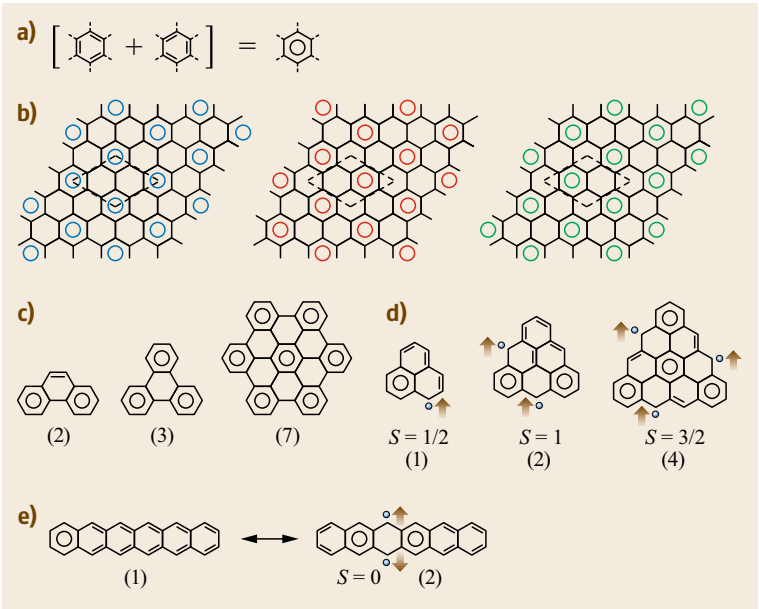


Fig. 10.3 **(a)** Clar's sextet. Clar's representations of **(b)** infinite graphene, **(c)** armchair-edged molecules, **(d)** zigzag-edged triangular molecule and **(e)** zigzag-edged linear molecule. In an infinite graphene, three Clar's representations are degenerate with sextets being arranged in a $\sqrt{3} \times \sqrt{3}$ superlattice (*dotted rhombus*) **(b)**. The *dot with an arrow* denotes an unpaired electron having a spin $S = 1/2$ in **(d,e)**. The number in the parenthesis and S indicate the number of sextets and the total spin respectively in **(c–e)**

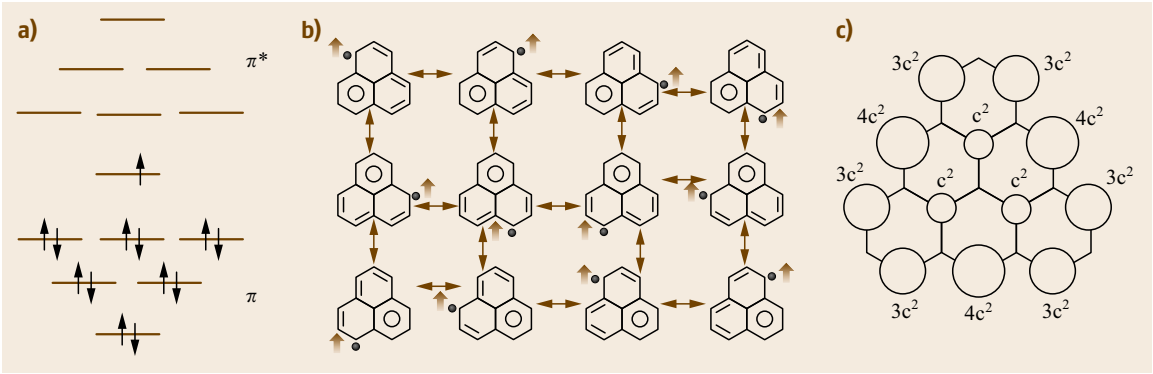


Fig. 10.4 **(a)** The electronic structure of a triangle molecule with three benzene rings fused (phenalene) (*left* molecule in Fig. 10.3d), **(b)** 12 degeneracy of Clar's representations in phenalene and **(c)** the spatial distribution of the local density of states of the nonbonding π-electron state (edge state) in triangulene consisting of six benzene rings. The nonbonding π-state is half-filled with an unpaired electron having $S = 1/2$ in **(a)**. The *arrow with a dot* denotes an unpaired electron in **(b)**

The situation is completely different in zigzag-edged fragments compared to armchair-edged ones, as shown in Fig. 10.3d,e, with examples of triangle and linear fragments respectively. Zigzag-edged nanographenes have small numbers of sextets, suggesting less energetic stability. In addition, there are unpaired electrons remaining in the edge region as shown in Fig. 10.3d,e. An unpaired electron occupies a nonbonding π-electron state in single occupancy between the π-valence states and the π^*-conduction states, as shown in Fig. 10.4a. In the case of triangle nanographenes, the spins $s = 1/2$ of unpaired electrons are arranged in parallel owing to the Hund rule, forming a ferromagnetic state (Fig. 10.3d). In contrast, the spins of two unpaired electrons in linear nanographene are arranged in antiparallel, forming a closed shell singlet state, or antiferromagnetic state in other words (Fig. 10.3e). Accordingly zigzag-edged nanographenes are magnetic due to the presence of nonbonding states half-filled with unpaired electrons. It should be noted that there exist many Clar's representations degeneratein zigzag-edged nanographenes, as can be understood from the presence of 12 representations in a fragment with three hexagons fused in triangle shape in Fig. 10.4b. Sextets migrate between 12 states, resulting in the creation of electrical conductive properties. The unpaired electron also migrates around the

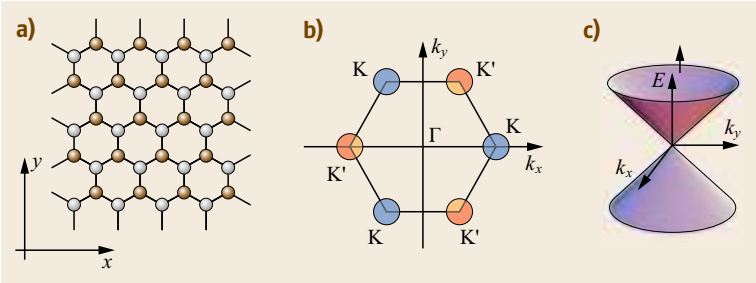


Fig. 10.5 **(a)** The two-dimensional hexagonal bipartite lattice of graphene, which has two independent sublattices; A and B sublattices denoted by *filled* and *open circles* respectively. **(b)** The hexagonal Brillouin zone of graphene with two independent reciprocal lattice points K and K′. **(c)** The electronic structure of graphene expressed with the Dirac cones of valence π- and conduction π^*-band, which correspond to the lower cone and upper inverted cone, respectively. The *arrow* shown on the Dirac cone indicates the pseudospin. Two independent sets of Dirac cones are placed at K and K′ points

zigzag-shaped peripheral region. Note that calculation proves the unpaired electron mostly populates in the vicinity of zigzag edges, as can be seen in Fig. 10.4c. Owing to this reason, the nonbonding π-state is called the edge state [10.9].

Next let us discuss the electronic structures of nanographenes in the physics aspect [10.4, 15]. The two-dimensional honeycomb hexagonal lattice of graphene is a bipartite lattice, which consists of two independent sublattices (A sublattice and B sublattice), as given in Fig. 10.5a. Here, the electron running in the hexagonal bipartite lattice can be described effectively in terms of massless Dirac fermion in relativistic quantum mechanics, as expressed in the Weyl equation with linear dependence on momentum $\boldsymbol{p}$

$$\boldsymbol{H} = v_{\mathrm{F}}\boldsymbol{\sigma}\boldsymbol{p}\,, \tag{10.1}$$

where v_{F} is the Fermi velocity given by

$$v_{\mathrm{F}} = \frac{3\gamma_0 a_{\mathrm{CC}}}{2\hbar} \approx \frac{1}{300}c$$

with a_{CC}, the C–C bond length (0.142 nm) and γ_0, the transfer integral (3.16 eV), representing the strength of the C–C interaction of π-electron, and $\boldsymbol{\sigma}$ is the Pauli matrix of a pseudospin. The Fermi velocity is estimated as $v_{\mathrm{F}} \approx (1/300)c$ (with speed of light c), which is 1000 km/s. Consequently the electrons on the graphene sheet run very fast. This is the reason why graphene is interesting for future fast electronics devices. The Pauli matrix $\boldsymbol{\sigma}$ represents a pseudospin, which is not a real spin, but is effectively converted from the structure degree of two in the bipartite lattice. In the Weyl equation, the application of magnetic field $\boldsymbol{B}$ modifies momentum as $\boldsymbol{p} \to \boldsymbol{p} - e\boldsymbol{A}$, where the gauge field $\boldsymbol{A}$ is related to the magnetic field applied; $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ [10.15]. Consequently the pseudospin becomes susceptible to the magnetic field applied. The electronic structure of graphene expressed in the Weyl equation is exhibited in Fig. 10.5c. The valence π- and conduction π^*-bands having linear $\boldsymbol{p}$ $(= \hbar\boldsymbol{k})$ dependence are called Dirac cones. The bottom of the conduction band and the top of the valence band touch each other at a point called the Dirac point. The Dirac point is located at the Fermi level when graphene is neutral in charge. The absence of an energy gap between the valence and conduction bands gives zero-gap semiconductor features to infinite graphene. The structural degree of freedom two is conserved in the momentum space. So there are two independent reciprocal lattice points K and K′ in the hexagonal Brillouin zone in graphene as shown in Fig. 10.5b. Accordingly two independent sets of Dirac cones, which are called K and K′ valleys, exist at K and K′ points.

When an edge is created by cutting a graphene sheet, the edge works as a boundary condition for the electron in graphene, modifying the electronic structure depending on the edge geometry. Let us examine the scattering process of the electron at the edges [10.15]. Figure 10.6a,b exhibits the scattering processes of the electron at armchair and zigzag edges respectively. In the case of the scattering process at an armchair edge shown in Fig. 10.6a, the momentum $\hbar k_x$ is changed to $-\hbar k_x$ (see also Fig. 10.5b for the coordinates). This is a transition from the K valley to the K′ one, or the K′ valley to K one, that is, the intervalley transition. Consequently the scattered wave can be described in terms of a linear combination of the wave functions of the K valley and K′ valley, as given in the following

$$\psi_{\mathrm{B}}(R_{\mathrm{B}}) \propto \left(-\exp\left(-\mathrm{i}KR_{\mathrm{B}}\right) + \exp\left(\mathrm{i}K'R_{\mathrm{B}}\right)\right), \tag{10.2}$$

for a site R_{B} belonging to sublattice B. This suggests that the two wave functions are subjected to an interference. The electron wave interference brings about the formation of a standing wave, that can be seen in the absence of degeneracy in the Clar's representations in

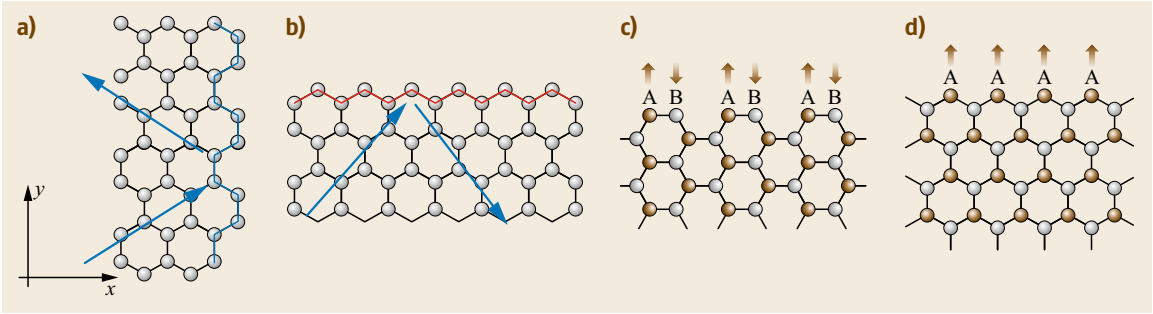


Fig. 10.6a–d The scattering processes of the electron at **(a)** armchair and **(b)** zigzag edges. Sublattice symmetry in armchair **(c)** and zigzag **(d)** edges

armchair-edged nanographenes. In other words, a single unique Clar's representation is the consequence of the formation of a standing wave. The standing wave formation makes armchair-edged nanographenes energetically stabilized together with the formation of an energy gap between the valence and conduction bands.

In zigzag-edged nanographene, the situation is in sharp contrast to that in armchair-edged nanographene. The momentum change is given as $\hbar k_y \rightarrow -\hbar k_y$ in the scattering process at a zigzag edge as shown in Fig. 10.6b. This is the transition from K valley to K one or K′ valley to K′ one, that is, only one valley participates in the transition process (intravalley transition). Accordingly there is no electron wave interference, a standing wave being absent in zigzag-edged nanographenes. What happens in zigzag-edge nanographene can be more clearly understood by comparing sublattice symmetry between armchair- and zigzag-edged nanographenes. Figure 10.6c,d shows the arrangements of A and B sublattices in the vicinity of armchair and zigzag edges. In an armchair edge, A and B sublattices are always paired the same to the interior region, the sublattice symmetry being conserved. In contrast, a zigzag edge consists of only one of the sublattices, A or B, that is, A-B symmetry is broken in the zigzag edge. In general, broken symmetry works to create a gauge field, which is lattice distortion-induced gauge field $\boldsymbol{A}_{\mathrm{q}}$. In the case of armchair edge, in which A-B symmetry is conserved, $\boldsymbol{A}_{\mathrm{q}} = 0$, whereas it becomes finite, ensuring that the pseudospin works to give a magnetic degree of freedom to zigzag-edged nanographene [10.15]. In other words the pseudospin behaves as real spin, that is equivalent to the spin of an unpaired electron, which occupies the nonbonding π-electron state in the chemistry aspect. We can count the number of spins N_{S} created in zigzag edges, in which A-B symmetry is broken, using *Lieb*'s theorem [10.16] given as follows

$$N_{\mathrm{S}} = |N_{\mathrm{A}} - N_{\mathrm{B}}| \,, \tag{10.3}$$

using which the total spin S is given in the following equation

$$S = \frac{1}{2}|N_{\mathrm{A}} - N_{\mathrm{B}}| \,. \tag{10.4}$$

Here N_{A} and N_{B} denote the numbers of sites in A and B sublattices, respectively.

As we discussed above, the formation of a standing wave in armchair-edged nanographenes and the presence of a nonbonding π-electron state in zigzag-edged ones are confirmed from both chemistry and physics aspects. Now we discuss the electron structures of these two types of nanographenes using local density approximation (LDA) calculations. Figure 10.7 shows the band structures of armchair- and zigzag-edged graphene nanoribbons (AGNR and ZGNR respectively) whose edge carbon atoms are monohydrogenated, together with the calculated spatial distribution of the local density of states and the corresponding Clar's representations. The electronic structure of an armchair-edged graphene nanoribbon (Fig. 10.7a) has an energy gap (E_{g}) between the valence and conduction bands in contrast to the absence of an energy gap in infinite graphene. The presence of an energy gap is responsible for the energetic stability of armchair-edged nanoribbons. The spatial distribution of the local density of states shown in Fig. 10.7b indicates the formation of a standing wave with a $\sqrt{3} \times \sqrt{3}$ superlattice as a signature of sextet localization (Fig. 10.7c). Therefore armchair-edged nanoribbons are subjected to aromatic stability, similar to Kekulé molecules in polycyclic aromatic hydrocarbons. The magnitude of an energy gap has a decreasing trend accompanied by an oscillation upon the increase in the ribbon width as shown in Fig. 10.8, and it is extrapolated to zero as the width becomes infinite. The aromatic stability declines upon the increase in the ribbon width. The oscillating behavior can be expressed as follows

$$E_{\mathrm{g}}^{3p+1} > E_{\mathrm{g}}^{3p} > E_{\mathrm{g}}^{3p+2} \,,$$

for widths of $n = 3p+1$, $n = 3p$ and $n = 3p+2$, where

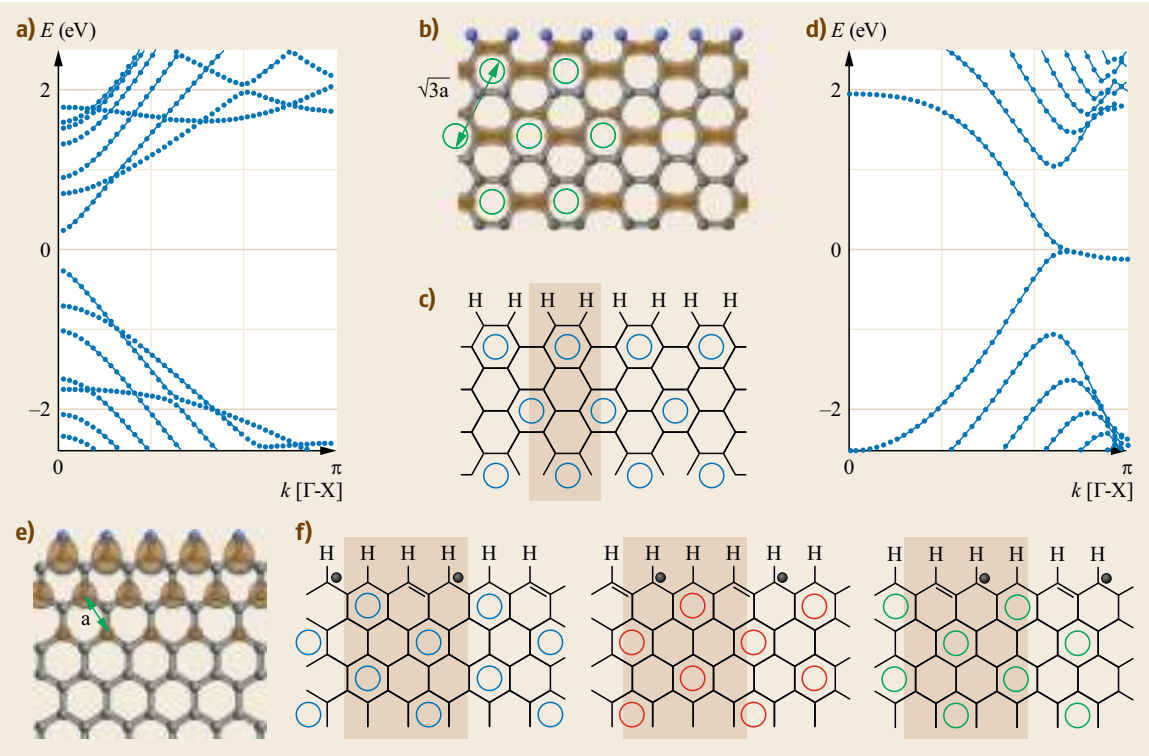


Fig. 10.7a–f The electronic structures of armchair-edged **(a–c)** and zigzag-edged graphene nanoribbons **(d–f)**, whose edge carbon atoms are monohydrogenated. **(a,d)** Calculated band structures, **(b,e)** calculated spatial distributions of the local density of states with LDA in the vicinity of the Fermi level (0–0.1 V), **(c,f)** Clar's representations. The Fermi level E_F is located at 0 eV in **(a)** and **(d)**. The ribbon widths are $n = 15$ and 10 respectively for armchair-edged and zigzag-edged nanoribbons (15-AGNR and 10-ZGNR)

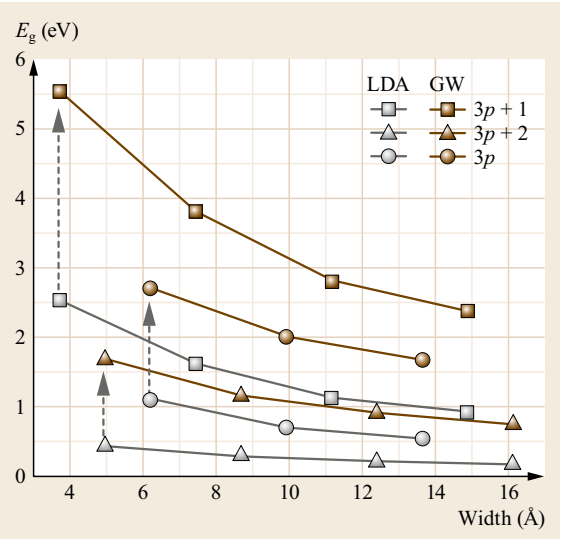


Fig. 10.8 Variation in energy gap E_g with the width of armchair-edged graphene nanoribbons. Three families of armchair graphene nanoribbons are represented by different symbols. Values for members of the same family of armchair nanographenes are connected by *solid lines* as guides to the eye. *Open symbols* represent bandgaps calculated by LDA; *solid symbols* represent the corresponding quasiparticle bandgaps obtained by the GW method. *Dashed arrows* indicate the self-energy correction for the narrowest ribbon in each of the three families (after [10.17]) ◄

the ribbon width is spanned by n C–C dimer chains, and p is an integer [10.17, 18]. The oscillating behavior is just the consequence of the change in aromaticity related to how aromatic sextets are tiled on nanoribbons. The electron correlation works to enhance the bandgap as can be seen in the calculation with the GW method in Fig. 10.8.

A zigzag-edged graphene nanoribbon has no gap, and instead it features the presence of a nonbonding π-electron state (edge state) as shown in Fig. 10.7d. A flat band having no energy dispersion between the valence and conduction bands is assigned to the edge

state in Fig. 10.7d. The spatial distribution of the edge states is found to be localized well in the vicinity of the zigzag edge region as can be seen in Fig. 10.7e. The Clar's representations are triply degenerate as confirmed in Fig. 10.7f. The sextet migration is further evidence for the absence of energy gap. In the edge, every three carbon atom has an unpaired electron with spin $s = 1/2$, which migrates between the three edge carbon atoms. Accordingly Clar's rule suggests the presence of a magnetic moment $1/3\,\mu_B$/C atom. The creation of a nonbonding edge state together with the absence of an energy gap indicates that zigzag-edged graphene nanoribbons are energetically unstable or less aromatic in other words. Note that the electron correlation effect works to create an energy gap even in zigzag-edged graphene nanoribbons [10.17] despite simple tight binding theory predicting the absence of a gap [10.9].

10.3 Characterizations

In Sect. 10.2, we discussed the electronic structures of graphene nanoribbons (GNR) having different edge geometries. Here structural and electronic characterizations of these nanoribbons are presented on the basis of what we learned in Sect. 10.2, in terms of several characterization techniques.

Microprobe techniques such as scanning tunneling microscopy (STM), scanning tunneling spectroscopy (STS) and atomic force microscopy (AFM) are powerful in characterizing both detailed atomic arrangements and electronic structures in graphene nanostructures. We can get information on the formation of standing waves in armchair edges and the presence of nonbonding edge states in zigzag edges. Figure 10.9 shows the results of STM observations in the vicinities of armchair and zigzag edges whose edge carbon atoms

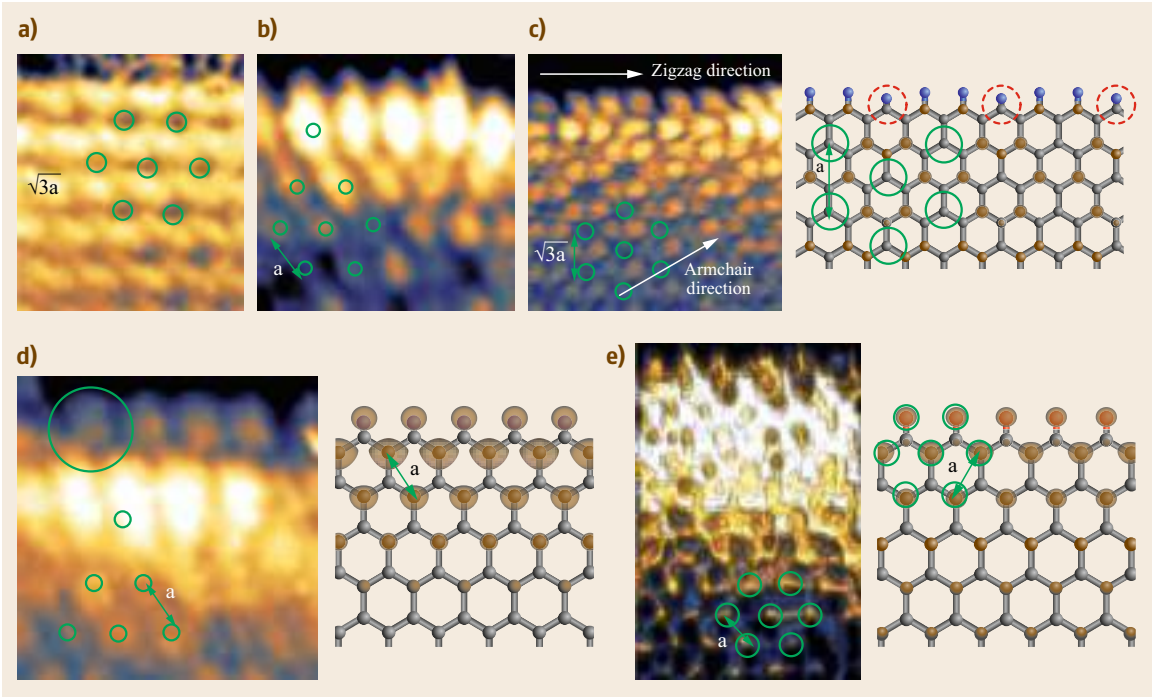


Fig. 10.9a–e STM images of armchair and zigzag edges. **(a)** Monohydrogenated armchair edge, **(b)** monohydrogenated zigzag edge, **(c)** zigzag edge in which every three edge carbon atom is dihydrogenated with others remaining monohydrogenated, **(d)** dihydrogenated zigzag edge and **(e)** zigzag edge terminated with carbonyl group (imaging conditions: $V_s = 0.1$ V, $I = 0.9$ nA for **(a,c,d)**, $V_s = 0.1$ V, $I = 0.7$ nA for **(b)**, $V_s = 0.2$ V for **(e)**). For **(c–e)**, the spatial distribution mapping of the local density of states calculated with LDA in the vicinity of the Fermi level (0–0.1 V) are also shown with *brown circles* on the right side of the STM images for comparison. The *green circles* with a and $\sqrt{3}a$ denote the periodicities of the graphene unit cell and the $\sqrt{3}\times\sqrt{3}$ superlattice respectively. The *large green circle* around the zigzag edge denotes the two hydrogen atoms terminating the edge carbon atom in **(d)**. The sites encircled with *red dotted lines* indicate the two hydrogen atoms terminating the edge carbon atoms in **(c)**

are terminated with different chemical species [10.10]. In the vicinity of a monohydrogenated armchair edge shown in Fig. 10.9a, a $\sqrt{3}\times\sqrt{3}$ superlattice is observed, which is the signature of a standing wave formed in an armchair edge. In contrast to the armchair edge, the bright spots assigned to edge states are observed in the vicinity of monohydrogenated zigzag edges in Fig. 10.9b. The local density of states of the edge state decays sharply as we go from the carbon atoms of a zigzag edge to the interior of the graphene sheet, the edge state being well localized at the zigzag edge. These findings in armchair and zigzag edges confirm the theoretical results shown in Fig. 10.7 of Sect. 10.2.

STM observations interestingly prove that the electronic structure varies depending on how the edge carbon atoms are terminated with foreign chemical species. Figure 10.9c gives the result of a zigzag edge in which every three edge carbon atom is dihydrogenated with others remaining monohydrogenated. Bright spots assigned to edge states disappear at the edge, instead a $\sqrt{3}\times\sqrt{3}$ superlattice is observed, completely similar to what happens in an armchair edge. This change can be understood using LDA calculations and also Clar's aromatic sextet rule as we can see in Fig. 10.9c and Fig. 10.10a respectively. The spatial distribution mapping of the local density of states in the vicinity of the Fermi level in Fig. 10.9c confirms the absence of edge states around the zigzag edge, instead the formation of a $\sqrt{3}\times\sqrt{3}$ superlattice is in good agreement with the experimental result. The formation of the superlattice together with the absence of edge states can be understood from the Clar's representation as we can see in Fig. 10.10a. A combination of two Clar's representations suggests sextet localization, which results in the formation of a $\sqrt{3}\times\sqrt{3}$ superlattice. When we add more hydrogen to the zigzag edge to form a completely dihydrogenated zigzag edge, the creation of an edge state and the disappearance of the superlattice are observed again in the STM image as shown in Fig. 10.9d, similar to those in monohydrogenated zigzag edges. This is confirmed with the spatial distribution mapping of the local density of states. The revival of the

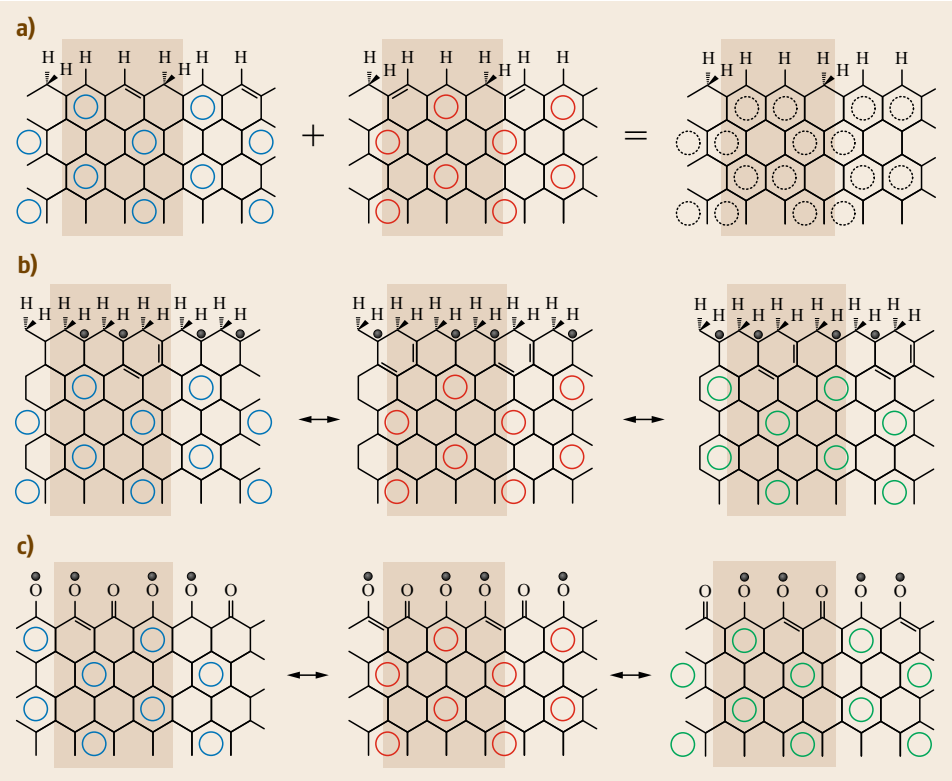


Fig. 10.10a–c Clar's representations of **(a)** zigzag edge, in which every three edge carbon atom is dihydrogenated with others remaining monohydrogenated, **(b)** dihydrogenated zigzag edge and **(c)** carbonylated zigzag edge. The *dots* denote an unpaired electron occupying the nonbonding edge state. In **(a)**, the combination of two Clar's representations (*left* and *center*) gives rise to the formation of a standing wave of $\sqrt{3}\times\sqrt{3}$ superlattice (the representation on the *right*) as a result of sextet localization, whereas the degeneracy of three Clar's representations results in sextet migration for **(b,c)**

electronic feature when we go from the partially dihydrogenated zigzag edge to fully dihydrogenated zigzag edge can be understood in terms of the migration of sextets in the triply degenerate Clar's representations in the fully dihydrogenated zigzag edge in Fig. 10.10b. The presence of unpaired electrons (edge state) in the zigzag edge is confirmed in the Clar's representations. Here it should be noted that the local density of states is distributed also on the two hydrogen sites terminating the edge carbon atoms as can be seen in the STM image and the local density of states distribution. This is the consequence of hyperconjugation; the direction connecting the two hydrogen atoms in the dihydrogenated site is in parallel to the direction of the carbon $2p_z$ orbital, which is perpendicular to the graphene sheet, resulting in the hybridization of hydrogen 1s and carbon $2p_z$ orbitals. Accordingly, the edge state can be extended to the two hydrogen sites terminating the edge carbon atom. It is also interesting that the spatial distribution of the edge state is extended to the interior of the nanoribbon. In the triply degenerate Clar's representations, two edge carbon atoms have an unpaired electron, which migrates between three representations. The magnetic moment of the edge state is estimated as $2/3\,\mu_B$/C atom, which is twice that in the monohydrogenated zigzag edge.

Oxygen-containing functional groups, which terminate edge carbon atoms, give electronic features different from hydrogen atoms. Edges terminated by oxygen-containing functional groups are more realistic for electronics device applications as most devices are handled in ambient atmosphere. Therefore investigation of nanoribbons whose edge carbon atoms are terminated by oxygen-containing functional groups is important. Figure 10.9e shows the STM image of a carbonylated zigzag edge together with the calculated spatial distribution of the local density of states around the Fermi level. Here the presence of an edge state is confirmed as a signature of a zigzag edge. Note that we can see the edge state around the edge and the absence of the superlattice in the STM image and also in the spatial distribution mapping of the density of states in the LDA calculation. The migration of sextets and unpaired electrons in the triply degenerate Clar's representations supports the STM observation and LDA calculation result. The magnetic moment is estimated as $2/3\,\mu_B$/C atom, the same as that in a fully dihydrogenated zigzag edge. What is interesting here is the extension of the edge states into the oxygen atom site in the carbonyl group as can be seen in Fig. 10.9e. In addition the local density of states in the edge state penetrates deeply into the interior of the graphene with a long decay length. Indeed, the decay length of the edge state is estimated as 2 nm from the experiment. The reason the feature different from a conventional monohydrogenated zigzag edge originates from the participation of oxygen $2p_z$ orbitals in the edge state. The oxygen $2p_z$ orbital is in parallel to the carbon $2p_z$ orbital, resulting in the hybridization of these two orbitals. The charge transfer from the edge carbon atom to the electronegative oxygen atom causes the population of the edge state in the oxygen atom and also the extended feature of the edge state.

STM/STS can provide information on the formation of an energy gap in armchair-edged nanoribbons [10.19, 20]. Figure 10.11 shows the STM image and STS spectrum of an armchair-edged graphene nanoribbon (ribbon width: $n = 13$) prepared using the bottom-up method. The armchair-edged graphene nanoribbon obtained through the bottom-up method is confirmed to have atomically precise edges as can be seen in Fig. 10.11a. In the STS spectrum ($\mathrm{d}I/\mathrm{d}V$ versus V plot), the presence of an energy gap is observed between the valence and conduction bands as shown in Fig. 10.11b, in which the energy gap is estimated as $E_g^{n=13} = 1.4 \pm 0.1$ eV for a nanoribbon having a width of $n = 13$. STS observations indicate that narrower nanoribbons have larger energy gap, consistent with theoretical prediction [10.17]; for example, $E_g^{n=7} = 2.5$ eV for $n = 7$. Nanoribbons with widths of $n = 7$ and 13 belong to the same subfamily of $n = 3p + 1$ with $p = 2$ and 4 respectively (Sect. 10.2). The gaps experimentally observed are systematically narrower than those predicted theoretically (3.8 and 2.4 eV for $n = 7$ and 13 respectively) with GW approximation, in which electron correlation is taken into consideration. But they are wider than those predicted by LDA calculation (1.6 eV for $n = 7$). This indicates that the electron correlation effect plays an important role in the gap formation. The discrepancy in the observed gap from the LDA value likely arises from image charge screening due to the gold substrate, on which the nanoribbons are placed. It should be noted that states of the valence and conduction band edges have higher intensity near the edges of the graphene nanoribbon, while the nanoribbon appears relatively featureless for energies within the gap as evidenced in Fig. 10.11b (B-D).

Raman spectroscopy is a powerful tool in characterizing the electronic and lattice vibrational structures of graphene nanoribbons and their edge geometry effect [10.21], where resonance Raman spectra specific to the electronic feature of zero-gap semiconductors in graphene is sensitive enough to get information on individual nanoribbons even though they are in the nanodimension. In principle there are important Raman bands called G-band, D-band and G′ (or 2D) at 1580, 1350 and 2700 cm^{-1} respectively, which can be utilized for characterizations of nanoribbons. The former one originates from the first-order Raman process asso-

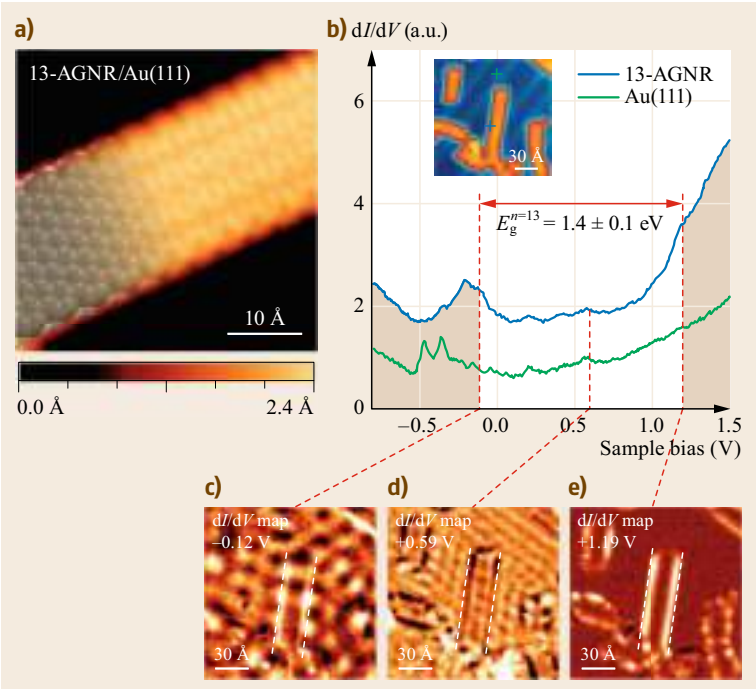
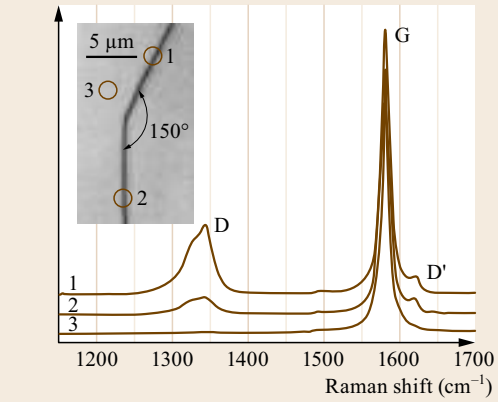
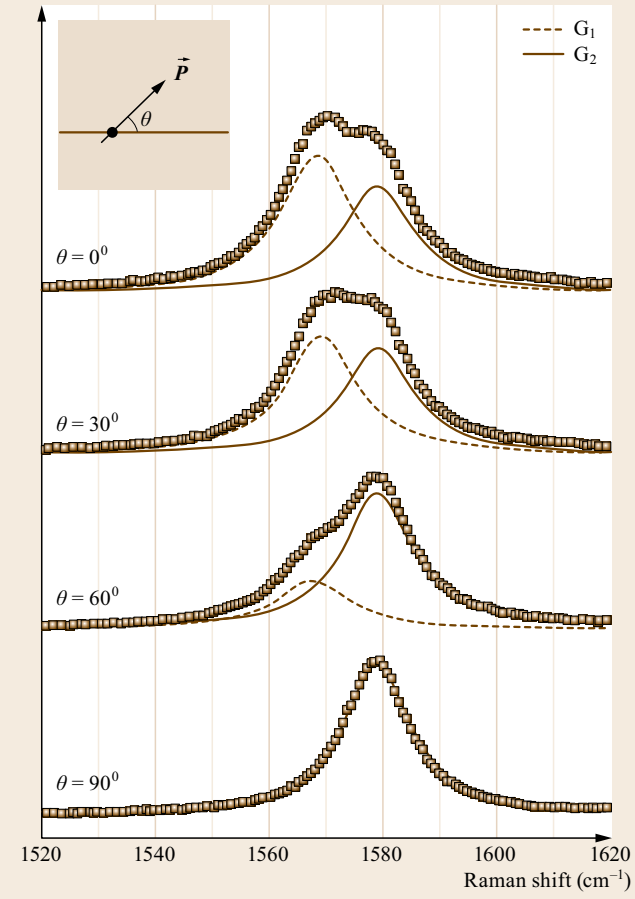
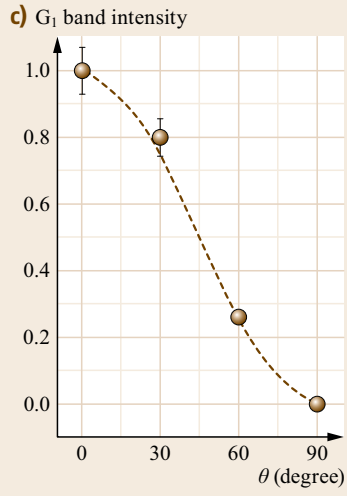


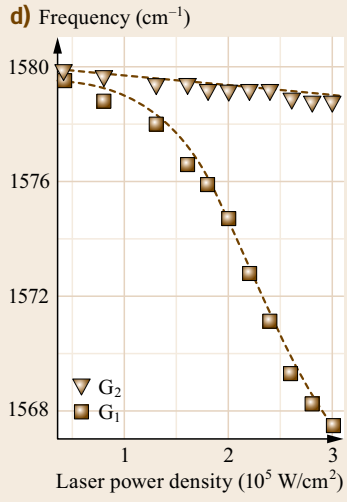
Fig. 10.11 **(a)** STM image of armchair-edged graphene nanoribbon (width $n = 13$) (13-ACNR) prepared on Au(111) substrate ($V_s = -0.70$ V, $I = 7.02$ nA). A structural model is overlaid onto the STM image. **(b)** STM dI/dV spectroscopic measurement of energy gap in 13-ACNR. dI/dV spectra recorded on a nanoribbon (*blue line*) and on bare Au(111) (*green line*). The nanoribbon spectrum is offset vertically by 1 au for clarity (open-feedback parameters: $V_s = 1.00$ V, $I = 35$ pA; modulation voltage $V_{rms} = 10$ mV). *Crosses in inset* topographic image indicate the positions where spectra were recorded (inset $V_s = 1.00$ V, $I = 35$ pA). **(c–e)** dI/dV maps of nanoribbon ($n = 13$) taken at **(c)** valence band edge ($V_s = -0.12$ V, $I = 35$ pA), **(d)** in the gap ($V_s = 0.59$ V, $I = 35$ pA) and **(e)** at the conduction band edge ($V_s = 1.19$ V, $I = 35$ pA). *Dashed lines* indicate the outer edges of the nanoribbon. (Reprinted with permission from [10.19], Copyright 2013 American Chemical Society)

ciated with intravalley scattering with the participation of a zone center phonon, while the latter two are owing to the double resonant Raman process related to the intervalley transition. In particular the intervalley scattering process in the D-band takes place together with the participation of elastic defect scattering. The G- and D-bands are sensitive to the edge geometry in graphene nanoribbons. We have learned how the electron scattering process takes place in armchair and zigzag edges in Sect. 10.2; that is, the K-K′ intervalley transition is responsible for the elastic scattering at the armchair edge, giving rise to the electron wave interference. Therefore, the D-band, which is owing to the elastic edge scattering, appears only in armchair-edge nanoribbons, whereas it does not in zigzag-edged ones. This can be seen in observations of Raman spectra as shown in Fig. 10.12a [10.22]. In the observed Raman spectra, the D-band appears in the armchair edge despite being weak in the zigzag edge and not observed in the interior of the graphene sheet, in good agreement with theoretical suggestion. The G-band can be observed both in armchair and zigzag edges as confirmed in Fig. 10.12a. Figure 10.12b shows the details of a G-band observed in armchair-edged graphene nanoribbon exhibited in the AFM image [10.23]. The experiment is carried out using a polarized laser beam. The G-band has large polarization angle dependence, whose intensity is expressed as $\cos^2\theta$ with polarization angle θ. This angular dependence is reproduced theoretically in terms of the K-K′ intervalley scattering coupled with longitudinal optical phonon mode as given in Fig. 10.12f [10.24]. The G-band shows angular dependence different in zigzag-edged nanoribbons compared to armchair-edge ones, as can be seen in theoretical predictions in Fig. 10.12f. It is given as $\cos^2\theta$ in armchair-edged nanoribbons. These findings in the experimental and theoretical works allow us to utilize Raman G- and D-bands as a powerful tool for characterizing the electronic and structural features in graphene nanoribbons.


a) Intensity (a.u.)
5 μm
1
2
3
150°
G
D
D'
1200
1300
1400
1500
1600
1700
Raman shift (cm−1)


b) Intensity (a.u.)
G1
G2
P
θ
θ = 0°
θ = 30°
θ = 60°
θ = 90°
1520
1540
1560
1580
1600
1620
Raman shift (cm−1)


c) G1 band intensity
1.0
0.8
0.6
0.4
0.2
0.0
0
30
60
90
θ (degree)


d) Frequency (cm−1)
1580
1576
1572
1568
G2
G1
1
2
3
Laser power density (10^5 W/cm^2)


e)
80 nm

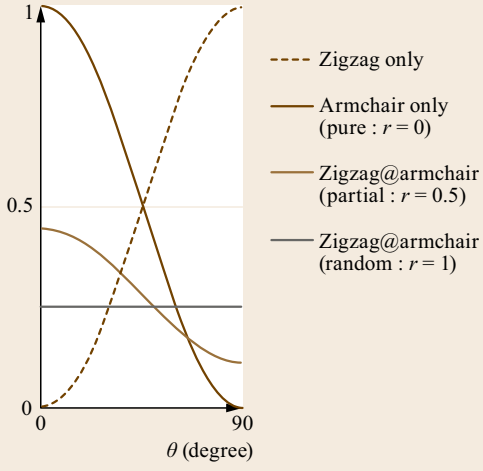

f) |M^opt(A)|^2 (a.u.)
1
0.5
0
0
90
θ (degree)
Zigzag only
Armchair only
(pure : r = 0)
Zigzag@armchair
(partial : r = 0.5)
Zigzag@armchair
(random : r = 1)

Fig. 10.12 **(a)** Raman spectra with D-, G-, and D′-bands obtained in three different regions of the highly oriented pyrolytic graphite (HOPG) sample. *Inset* shows an optical image of the step and the regions where spectra are taken: 1, armchair edge; 2, zigzag edge; and 3, interior of the graphene sheet. (Reprinted with permission from [10.22]) **(b)** Raman G-band spectra obtained for incident light having different polarization angles (θ). *Inset* shows a schematic of the graphene nanoribbon sample in **(e)** (*horizontal gray line*) showing the angle θ between the ribbon's axis and the light polarization vector ($\boldsymbol{P}$). G_1 and G_2 are assigned to the G-bands of the graphene nanoribbon and the HOPG substrate respectively. **(c)** Intensity of the G_1 peak versus θ. *Dotted line* is a $\cos^2\theta$ theoretical curve in **(f)**. **(d)** Raman frequencies of the G_1 (*squares*) and G_2 (*triangles*) peaks as a function of laser power density. Excitation laser energy is $E_{laser} = 2.41$ eV. **(e)** AFM image of single-layer graphene nanoribbons (width: 8 nm, length: $> 1\,\mu$m) (after [10.23], Copyright 2004 American Physical Society). **(f)** Theoretical polarization angle dependence of the G-band intensity for zigzag, armchair and zigzag-armchair mixed edges (after [10.24]) ◄

Near-edge x-ray absorption fine structure (NEXAFS) in x-ray absorption spectroscopy [10.25, 26] and electron energy loss near edge structure (ELNES) in electron microscopy [10.27] also give important information about graphene nanoribbons. These are element-specific for investigating the electronic structure of states above the Fermi level E_F, near which the edge state is located for zigzag-edged nanoribbons. Therefore, we can extract information on the electronic structure of the edge state that is populated around the edge carbon atoms. In particular, ELNES together with microscope images can offer details on the electronic structure in atomic resolution in graphene nanostructures. Figure 10.13 shows the experimentally observed NEXAFS for nanographene, compared with that for HOPG [10.25]. The NEXAFS spectra of nanographene

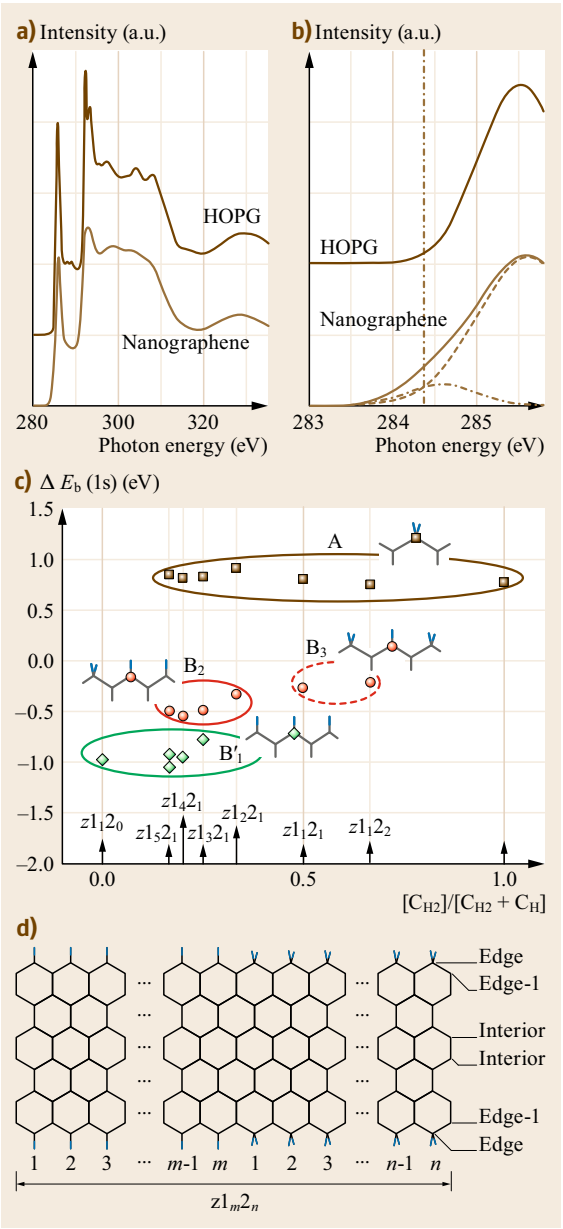


Fig. 10.13 **(a)** Carbon K-edge NEXAFS spectra of nanographene and HOPG. **(b)** Close-up of the pre-edge region of the NEXAFS spectra (*solid line*) and convoluted curve fit. HOPG spectrum is fitted with a single Gaussian; in nanographene, the deconvolution consists of two Gaussian peaks corresponding to the edge state (*dot-dash curve*) and the π^* conduction band (*dashed curve*). *Vertical dot-dash line* indicates the position of the Fermi level of HOPG (284.36 eV) (after [10.25]). **(c)** Chemical shifts (ΔE_b (1 s)) of 1 s core-level binding energy of carbon at zigzag edges in zigzag-edged graphene nanoribbons having monohydrogenated edges, dihydrogenated edges and edges consisting of monohydrogenated and dihydrogenated regions. *Insets* show the edge structures of the graphene nanoribbons. There are four groups (A, B'_1, B_2, B_3) having different chemical shifts in the edge structures. **(d)** Schematic illustration of geometric structures of graphene nanoribbons with different degrees of hydrogenation at zigzag edge. z1 and z2 denote the monohydrogenated (H1) and dihydrogenated (H2) zigzag edges and m and n represent the numbers of H1 and H2 terminations at edges, respectively. *Arrows* indicate the periodic boundary of the unit cell (after [10.26]) ►

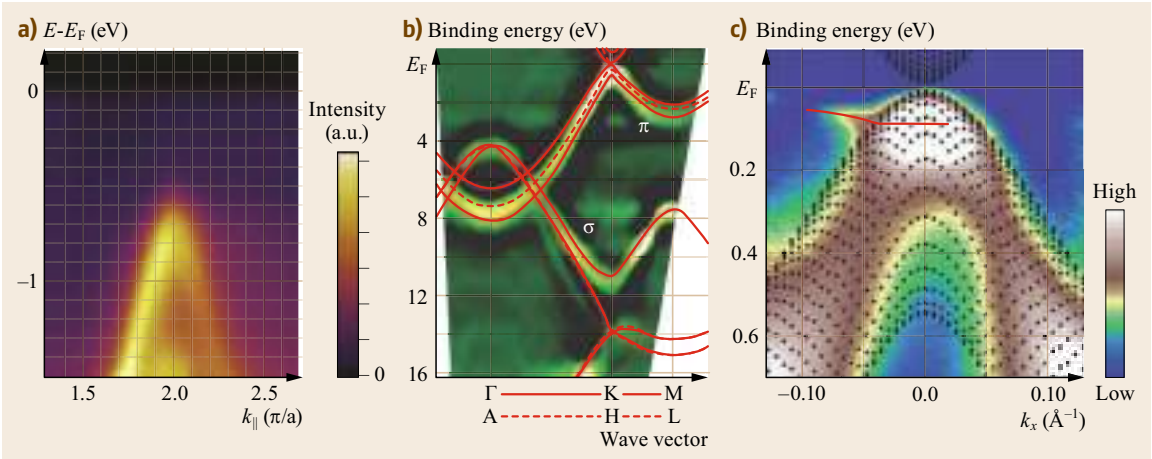


Fig. 10.14 **(a)** Angle-resolved PES intensity plot ($(E-E_F)$, $k_\parallel$) recorded along the ribbon axis for atomically precise 7-AGNR. $h\nu = 37\,\text{eV}$, $T = 300\,\text{K}$. K point is located at $k_\parallel = 2\pi/a$ ($a = 3a_{CC} = 0.43\,\text{nm}$). (Reprinted with permission from [10.20], copyright 2012 American Chemical Society). **(b)** Experimental band structure (*bright area*) of graphite obtained by plotting the second derivative of the ARPES intensity as a function of the wave vector and binding energy, compared with the modified first-principles band structure calculation (*solid* and *dashed lines* represent $\Gamma \leftrightarrow K \leftrightarrow M$ and $A \leftrightarrow H \leftrightarrow L$ respectively). **(c)** Experimental band structure near E_F around the K(H) point obtained by the same method as in **(b)**. The experimental result is compared with a projection of the calculated π band onto the (001) plane (*small dots*). The *line* from $k_x \approx 0.0$ to $-0.10\,\text{Å}^{-1}$ just below E_F shows the peak position of a weakly dispersed edge state near E_F. (Reprinted with permission from [10.28], copyright 2006 American Physical Society)

and graphite (HOPG) have typically two peaks, at 285.5 and 291.9 eV, which correspond to the C1s to π^* and σ^* transitions respectively, as shown in Fig. 10.13a. In the nanographene spectrum, an extra small peak around E_F (284.36 eV) is observed at 284.5 eV, which is assigned to the edge state.

The position of the edge state in energy is sensitive to the detailed structure of the graphene edge [10.26] arising from the detailed shape of the edges and any chemical modifications. The chemical shift $\Delta E_b(1s)$, which is defined as the peak position with respect to the 1s core-level binding energy $E_b(1s)$ of interior C, in the NEXAFS spectra of graphene nanoribbons calculated by density functional theory (DFT) theory, show a systematic trend depending on the edge shape and chemical modifications, as shown in Fig. 10.13c,d [10.26], where monohydrogenated or dihydrogenated carbons appear at the zigzag edges. Monohydrogenated edges, dihydrogenated edges, and zigzag edges with a mixture of monohydrogenated and dihydrogenated edge carbon atoms are investigated. Here, $z1_m2_n$ denotes a zigzag edge in which the unit cell consists of a monohydrogenated region with m edge carbon atoms and a dihydrogenated region with n edge carbon atoms. In the edge carbon atom of a pure monohydrogenated zigzag edge or the monohydrogenated edge carbon atom, which is neighbored by monohydrogenated sites on both side, in a zigzag edge having a mixture of monohydrogenated and dehydrogenated sites, the chemical shift shows a large negative value of $\approx -1.0\,\text{eV}$, suggesting the presence of an edge state well localized at the edge carbon atom. Note the region encircled with B_1' in Fig. 10.13c. The edge carbon atom at a site surrounded by dihydrogenated edge carbon site(s) has a less negative chemical shift of ≈ -0.2 to $-0.5\,\text{eV}$ (regions B_2, B_3), which is in good agreement with the STM observations of a dihydrogenated zigzag edge discussed above [10.5]. The edge carbon atom having a larger number of dihydrogenated neighboring sites has less a negative chemical shift. In contrast, the dihydrogenated edge carbon atom (region A) has a positive chemical shift of ≈ 0.6–$0.8\,\text{eV}$, and the edge state is absent at the carbon site.

Photoemission spectroscopy (PES) is available for characterizations of nanoribbons, in particular their electronic structures of the valence states below the Fermi level [10.20, 28]. The formation of a gap is observed in angle-resolved photoemission spectroscopy (ARPES) spectra for armchair-edged graphene nanoribbon ($n = 7$) 7-AGNR grown on narrow terraces of (111) structure (3.83 nm in width) periodically created on a stepped Au(788) surface. The grown nanoribbons are unidirectionally aligned along the terrace. Figure 10.14a shows the ARPES spectrum of the nanoribbons. The spectrum proves that the valence band

top, which is parabolic shaped, is located 0.7 eV below the Fermi level, as a consequence of the gap formation. The valence band top position is in quantitative agreement with the STS result, suggesting that the bandgap (2.3 eV) is not positioned symmetrically with respect to the substrate Fermi level. This is explained on the basis of the difference in the work function between the graphene and Au substrate. The effective mass is estimated as $m^* = 0.21 m_e$.

The formation of edge state located around the Fermi level is confirmed in zigzag edges also using ARPES. Figure 10.14b,c shows the results of ARPES experiments. The graphite π valence bands, whose top touches the Fermi energy at the K point, are clearly visible, in addition to the σ bands well below E_F, in excellent agreement with a first-principles band structure calculation. In Fig. 10.14c, an anomalous feature of the edge state, denoted by a solid line, appears just at the top of the π valence band below E_F. This state is almost flat at wave numbers from $k_x \approx 0.0$ to -0.05 Å^{-1}, and it shows a slight upward dispersion toward E_F at $k_x \approx 0.05$ to -0.10 Å^{-1}.

10.4 Syntheses: Top-Down Methods and Bottom-Up Methods

10.4.1 Graphene Nanoribbon Formation by Top-Down Approaches

GNRs have been fabricated by top-down approaches and bottom-up approaches. A typical top-down method makes use of electron beam lithography. In fact, *Han* et al. fabricated GNRs with widths down to 14 nm and demonstrated that the energy gap, E_g, for which quantum-size effect and aromatic stability are responsible (Sects. 10.2 and 10.3), is expressed as a function of the ribbon width as

$$E_g = \frac{\alpha}{W - W^*}, \qquad (10.5)$$

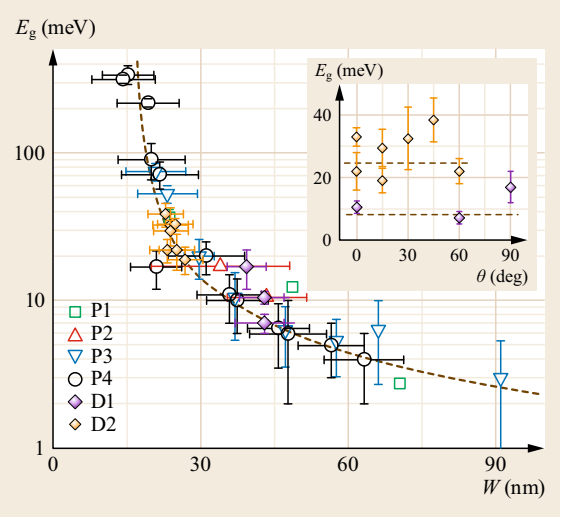


Fig. 10.15 Dependence of energy gap (E_g) on ribbon width (W) for six nanoribbon samples (P1, P2, P3, P4, D1, D2). The *insert* shows that the dependence of E_g on the crystallographic direction of GNRs. The angle (θ) of GNR was taken from an arbitrarily determined direction. No angle dependence was observed (after [10.29])

where $\alpha = 2$ eV and $W^* = 16$ nm (Fig. 10.15) [10.29]. W^* is considered to be the inactive edge width and is interpreted as the difference between the apparent ribbon width and the actual ribbon width or the decrease in the effective ribbon width as a result of edge defects. Electron beam lithography systems keep improving, and now it can pattern GNRs with widths below 10 nm [10.30]. Studies on the electron transport in GNRs fabricated by electron beam lithography have been performed further, and the energy gap above is now interpreted as a transport gap, which is generally larger than the theoretically calculated bandgap [10.31, 32]. The transport gap has been interpreted using various models including Anderson localization, variable-range hopping, and the Coulomb blockade model. In all the models, the existence of defects at the edges is considered to cause the transport gap.

Graphene nanoribbons were also made by using other top-down approaches. In fact, an array of GNRs with widths down to 8 nm have been fabricated by using directed self-assembly (DSA) of a cylindrical poly(styrene-b-dimethylsiloxane) (PS-b-PDMS) block copolymer as shown in Fig. 10.16 [10.33]. The DSA

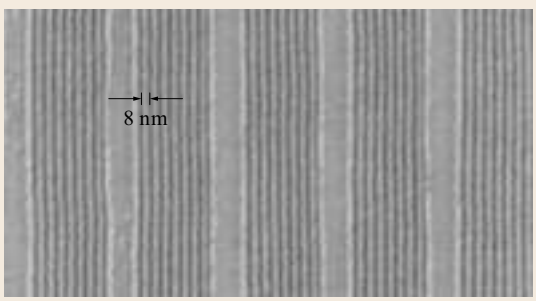


Fig. 10.16 SEM image of 8 nm-wide graphene nanoribbons capped with PDMS. (Reprinted with permission from [10.33], copyright 2012 American Chemical Society)

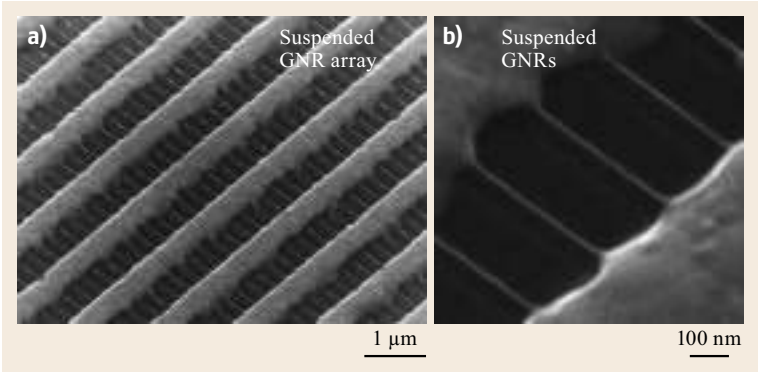


Fig. 10.17a,b SEM images of suspended GNRs (scale bar: **(a)** 1 μm, **(b)** 100 nm) [10.34]

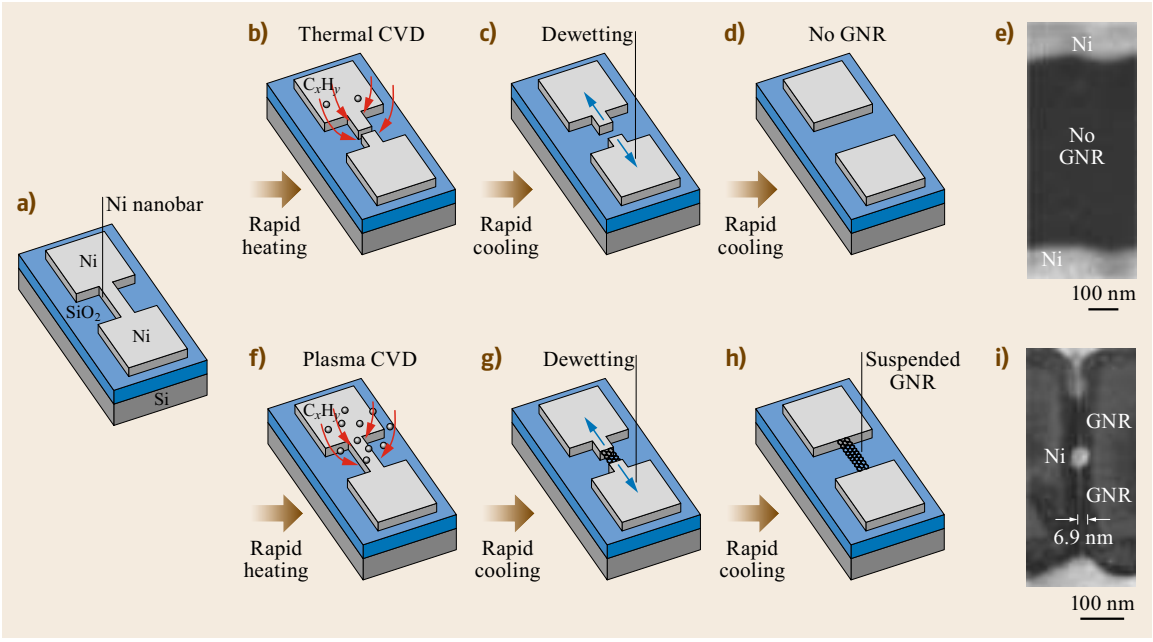


Fig. 10.18a–i Schematic diagram of growth method for GNR by **(a–d)** rapid heating thermal CVD and **(a)**, **(f–h)** rapid heating plasma. *Arrows* in **(b,f)** and **(c,g)** indicates the direction of hydrocarbon supply and dewetting of Ni nanobar respectively. **(e,i)** SEM images of the **(e)** broken Ni nanobar and **(i)** suspended GNR with sub-10 nm width. Scale bar: 100 nm **(e,i)** (after [10.34])

technique is one of the promising techniques for patterning sub-10 nm features in the aggressively scaled semiconductor devices. The variations in width of fabricated GNRs were still process dependent. The variations in width in each single GNR, and those in width of multiple GNRs were both at least around 3 nm. These variations actually caused poor transistor characteristics, as described in the next section.

Direct patterning of GNRs by focused helium ion beams in a helium ion microscope has also been performed [10.35, 36], fabricating a 5 nm-wide GNR array [10.36]. However, the G-band and 2D band (at 2700 cm^{-1}) in Raman spectra for such GNRs broaden as the GNR width decreases from 15 to 5 nm. In fact, the 2D band is almost indistinguishable for GNRs narrower than 8 nm [10.36]. These Raman spectra can be explained by softening of the phonons due to the dominance of defects in the GNR lattice especially around the edges. The defects may have been due to imperfect etching by helium ions. It is also possible that helium ions backscattered from the substrate damaged the GNRs.

Scanning tunneling microscopy (STM) lithography, where high bias potential is applied to the STM tip in order to etch graphene, was demonstrated to form GNRs [10.12]. The constant-current STM im-

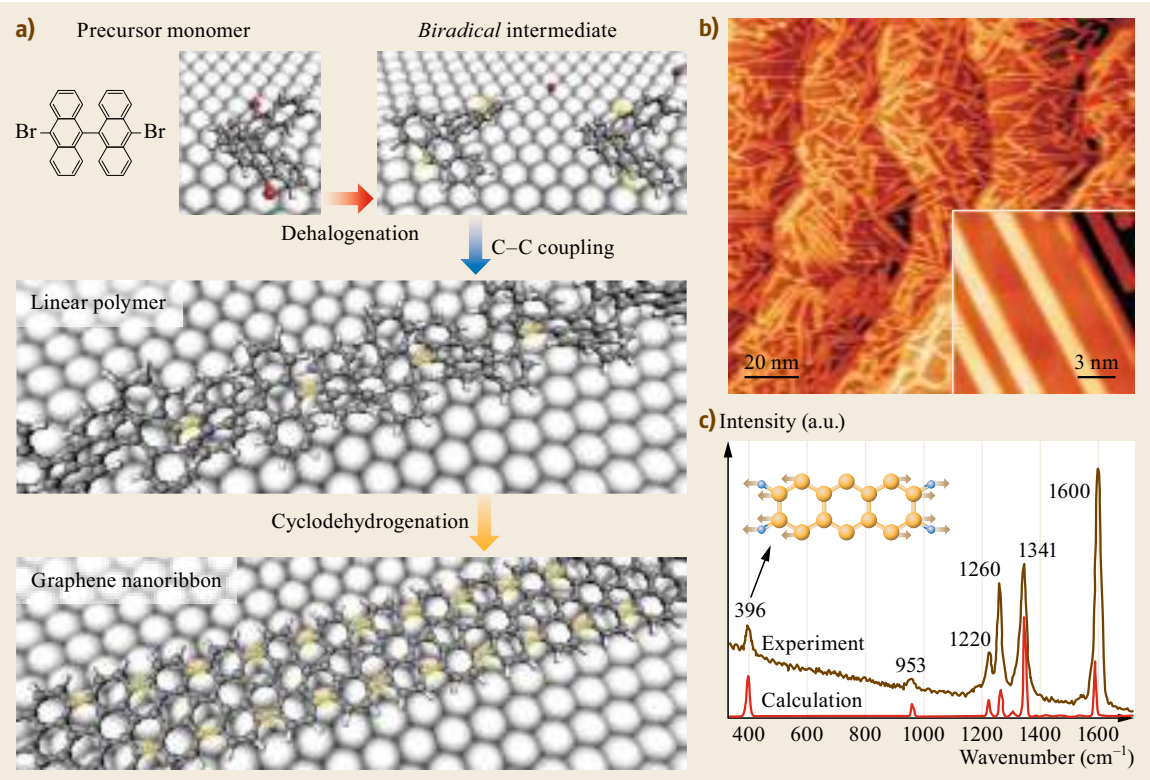


Fig. 10.19 **(a)** Fabrication process of atomically precise bottom-up 7-AGNRs. **(b)** Typical STM images of 7-AGNRs. **(c)** Typical Raman spectrum of the 7-AGNR. The peak at 396 cm^{-1} is characteristic for the 0.74 nm width of 7-AGNR, assigned to the radial-breathing like mode with the atomic displacements (see the *inset*) (after [10.37])

age of a 10 nm-wide armchair GNR (AGNR) exhibited confinement-induced standing wave patterns. An AGNR with a width down to 2.5 nm was formed by this technique. The energy gap obtained by scanning tunneling spectroscopy (STS) for the 2.5 nm-wide AGNR was about 0.5 eV, agreeing with first-principles calculations using local density approximations (LDA). Another interesting approach to fabricating arrays of suspended GNRs were recently reported (Fig. 10.17), where GNRs, which mainly consisted of multilayer graphene, were grown by plasma-enhanced chemical vapor deposition (PECVD) on arrays of nickel nanobars with widths from 20 to 100 nm [10.34]. The fabrication process is illustrated in Fig. 10.18. The GNRs with widths down to 6.9 nm were obtained using this approach. From polarized Raman spectra, it was concluded that the GNRs had near-zigzag edges, although the mechanism for having preferential edge orientation is not yet clear.

As described above, several top-down approaches to form GNRs have been reported. It is now even possible to fabricate arrays of GNRs. However, their Raman spectra and electrical properties, which are explained later, are not as good as expected. This is, as mentioned above, due to rough edges of GNRs caused by limitations in the current top-down methods. However, these kinds of difficulties may be solved in the near future, considering the history of semiconductor technologies.

10.4.2 Graphene Nanoribbon Formation by Bottom-Up Approaches

As a bottom-up approach, a method of fabricating atomically precise armchair-edged GNRs from 10,10-dibromo-9,9-bianthryl (DBBA) as a precursor was reported in 2010 [10.37]. Figure 10.19 illustrates the fabrication method. First, DBBA molecules are evaporated onto Au(111) substrate under ultrahigh vacuum (UHV). When the substrate temperature is increased to around 200 °C, halogen atoms (Br) are removed from the precursors and the radical intermediates diffuse and meet on the surface. The intermediates are then bonded with each other via aryl-aryl coupling, forming linear chains, which is basically the Ullmann reaction on a surface. When the temperature is further increased to 400 °C, cyclodehydrogenation occurs

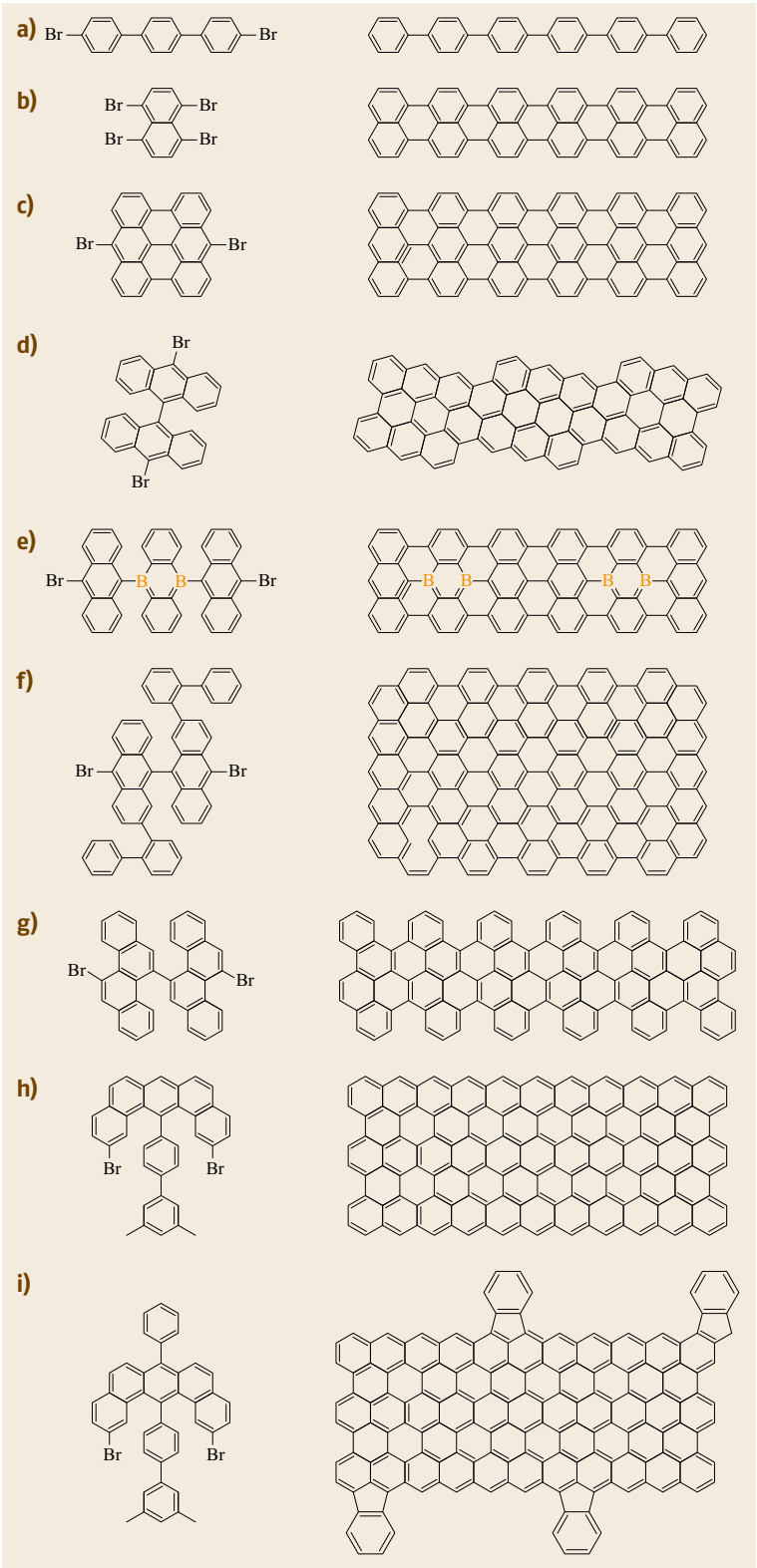


Fig. 10.20a–i Precursors (*left*) and GNRs (*right*) formed from the corresponding precursors (after [10.19, 38, 39])

to transform polymers into atomically precise armchair GNRs. As shown in Fig. 10.19, the ribbon width is well controlled and a Raman vibrational mode that depends on the ribbon width is observed. The fabricated GNRs are those with armchair-shaped edges with seven dimer lines (7-AGNR) and the bandgap is theoretically predicted to be about 3.7 eV [10.17]. *Ruffieux* et al. estimated the bandgap by STS to be approximately 2.3 eV [10.20], in reasonable agreement with the theoretical bandgap [10.17] considering the mirror charge effects of the Au substrate. The electrical conductance of such GNRs was first measured by picking up a 7-AGNR with an STM tip and measuring the conductance between the tip and the substrate. The GNR exhibited nearly insulating behavior due to the large bandgap [10.40].

The formation mechanism of GNR above has been studied further theoretically and experimentally [10.41–44]. Since GNRs are formed by the on-surface Ullmann reaction, the choice of substrate greatly affects the reaction. The reaction may be divided into three main steps: namely, dehalogenation, polymerization, and cyclodehydrogenation. *Simonov* et al. experimentally showed that the onset temperature of each step depends on the substrate type [10.43]. For instance, at room temperature, the DBBA precursor molecules on Au(111) remain intact, while on Cu(111) a complete surface-assisted dehalogenation occurs. On Cu(111), tilted polymer chains are formed even at room temperature or slightly elevated temperatures. Intramolecular cyclodehydrogenation begins at the annealing temperature above 100 °C on Cu(111) and flat AGNRs are formed at 200 °C, in contrast to the case of Au(111) described above. Theoretical simulations have also revealed that energy barriers for each reaction step are greatly reduced by the existence of metal surfaces, and that the degree of barrier lowering depends on the choice of metals [10.41, 42]. The choice of halogen atoms introduced to precursors can also play a role in GNR formation [10.42, 44]. For example, iodine atoms desorb more easily from precursors than bromine atoms. The desorbed halogens seem to remain on the surface until the polymerization process is completed and may poison the substrate, hindering long-range ordering of the self-assembled structures [10.42]. However, the role of halogens on the surface still remains unclear, and further studies are needed.

The bandgap of 3.7 eV in 7-AGNR may be too large for the application of GNRs to electronic devices. However, different types of atomically precise bottom-up graphene nanoribbons have been synthesized since 2010 (Fig. 10.20) [10.19, 38, 39, 45–49]. Such GNRs have been realized by carefully designing appropriate precursor molecules, as also illustrated in Fig. 10.20. As for AGNRs, it is now claimed that 3-, 5-, 7-, and 13-AGNRs have successfully been formed. The bandgaps observed by STS agree reasonably well with theoretical predictions [10.17], by taking into account the influences of metal surface. The typical length of 7-AGNRs was a few tens of nanometers. On the other hand, newer AGNRs, say, 5- and 13-AGNRs appeared to be shorter, typically less than 10 nm. This is not good considering real-world applications. This might be related to the lifetime of radicals for each precursor and further studies are needed. 13-AGNRs above were actually formed in attempt to obtain AGNRs with a smaller bandgap. In order to lower the bandgap further, wider AGNRs would be required in general. It is imagined that formation of wider AGNRs requires relatively large precursor molecules, which can be difficult to evaporate. Therefore, a different strategy may be needed to obtain wider AGNRs.

A zigzag-edged GNR (ZGNR) has recently been formed on Au(111) surface by employing precursors shown in Fig. 10.20h [10.38]. A noncontact AFM image of the 6-ZGNR obtained with a CO-functionalized tip, as well as STM images, are shown in Fig. 10.21. Anticipated edge states of 6-ZGNR were not observed for 6-ZGNR on Au(111) due to the surface state of the underlying Au(111) substrate; however, the edge states were obtained for 6-ZGNRs on insulating NaCl islands, where they are electronically decoupled from the metal substrate. The edge states were observed to have a bandgap, in good agreement with theoretical calculations using the GW approximation, in which the electron correlation effect is involved [10.17]. The successful formation of ZGNRs may open up possibilities for ZGNR-based spintronics devices, as discussed later.

Boron-doped AGNRs have also been formed as shown in Fig. 10.22 [10.46, 47]. However, the ratio of boron atoms to carbon atoms is too high to refer this case to doping, completely changing the original band structure. Therefore, replacing carbon atoms in the GNR body may not be a good idea for doping GNRs. However, this kind of technique can produce GNR families with similar dimensions, but having various electronic states, and can provide another set of building blocks for carbon-based nanoelectronics.

10.4.3 Graphene Nanoribbon Formation by Other Approaches

Other methods have also been used in attempts to form GNRs. For example, *Li* et al. formed GNRs by heating exfoliated expandable graphite to 1000 °C in a forming gas, dispersing the resulting graphite in a 1,2-dichloroethane (DCE) solution of poly[(*m*-phenylenevinylene-*co*)-2,5-(dioctoxy-p-phenylenevin-

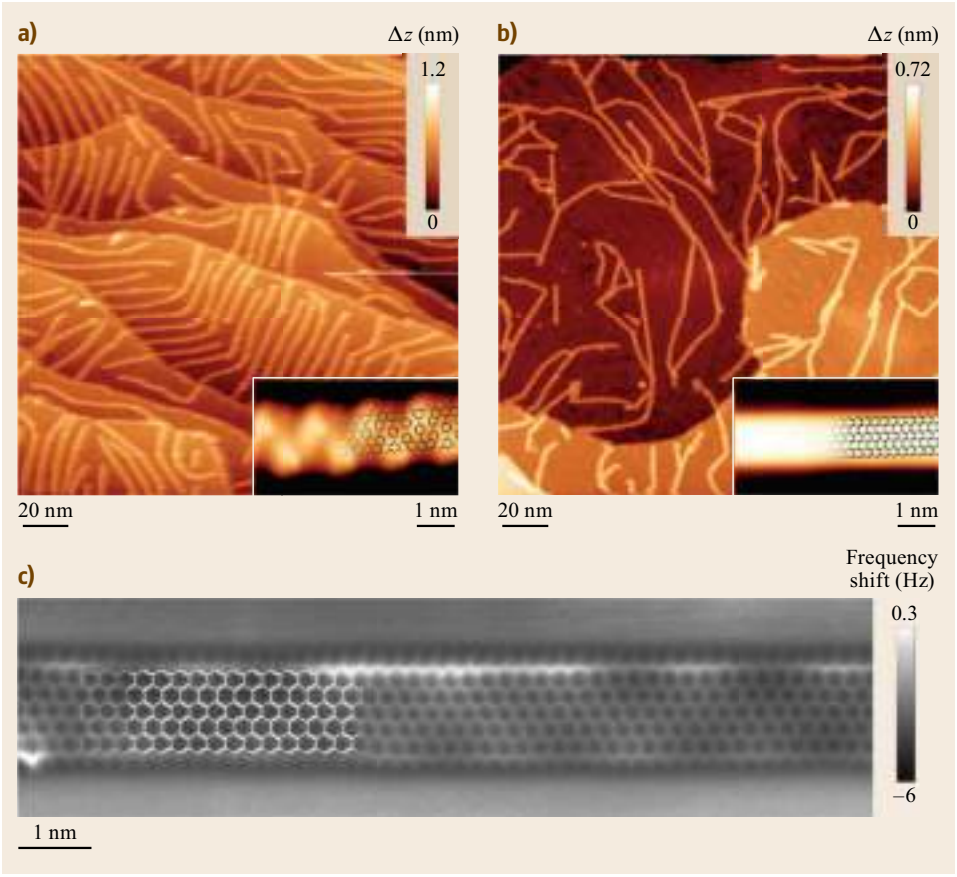


Fig. 10.21a–c Bottom-up synthesis of 6-ZGNR. **(a)** Large-scale STM image of the Au(111) surface after deposition of monomer (Fig. 10.20h) on the surface held at 475 K. Formation of polymers is observed. Scale bar: 20 nm. *Inset*: high-resolution STM image of the polymer. Scale bar: 1 nm. **(b)** STM image of the Au(111) surface after annealing at 625 K. 6-ZGNRs are observed. Scale bar: 20 nm. *Inset*: high-resolution STM of a 6-ZGNR. Scale bar: 1 nm. **(c)** Constant height noncontact AFM frequency shift image taken with a CO-functionalized tip. Scale bar: 1 nm (after [10.38])

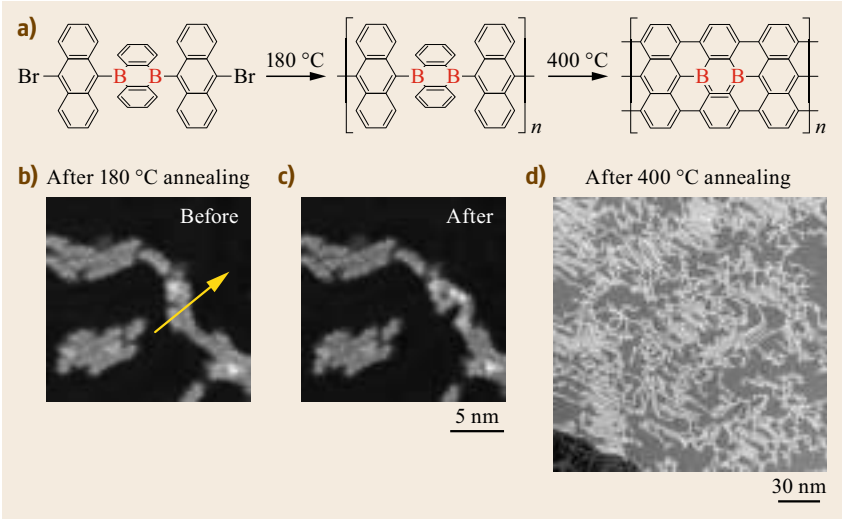


Fig. 10.22a–d Synthesis of boron-doped graphene nanoribbons. **(a)** Schematic diagram of the chemical reaction. **(b)** STM image of molecules, polymerized by annealing at 180 °C and **(c)** after the tip-induced manipulation. Scale bar: 5 nm. **(d)** Boron-doped GNR, synthesized by annealing at 400 °C. Scale bar: 30 nm (after [10.47])

ylene)] (PmPV) and ultrasonicating it [10.13]. GNRs with widths narrower than 10 nm were obtained. TEM revealed that these GNRs had ultrasmooth edges. As explained later, electrical properties of these GNRs are probably better than those of GNRs obtained by any other approaches so far. The bandgap was also in reasonable agreement with that obtained by first-principles calculations (without consideration of the

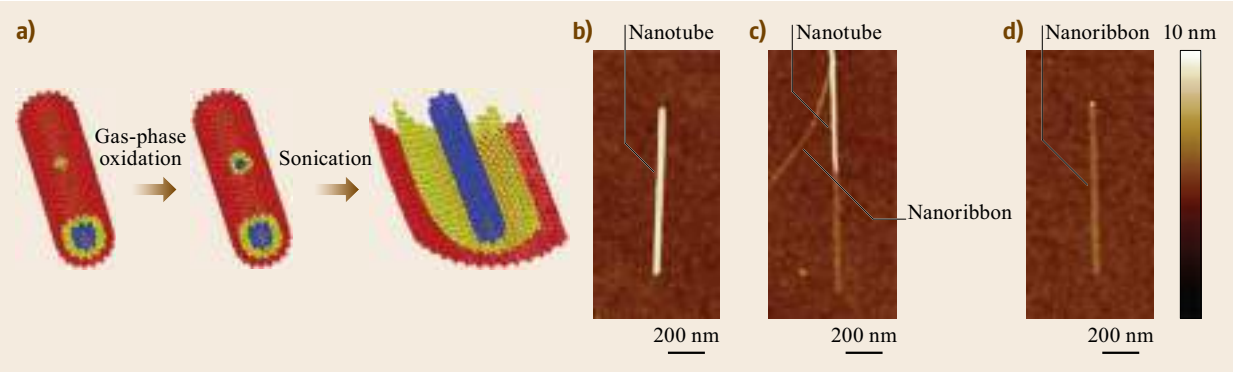


Fig. 10.23a–d Unzipping of nanotubes using a two-step process. **(a)** Schematic of the unzipping processes. In the mild gas-phase oxidation step, oxygen reacts with preexisting defects on nanotubes to form etch pits on sidewalls of nanotubes. In the solution-phase sonication step, sonochemistry and hot gas bubbles enlarge the pits, unzipping the tubes. **(b–d)** AFM images of pristine, partially and fully unzipped nanotubes, respectively (after [10.50])

electron correlation). With this method, relatively high-quality GNRs can be obtained; however, the control of the ribbon width and length is difficult.

Methods of fabricating GNRs from carbon nanotubes (CNTs) have also been proposed [10.51, 52]. In one method, GNRs were fabricated by cutting the side walls of multiwalled CNTs in the lengthwise direction with the treatment of multiwalled CNTs (MWCNTs) with a strong oxidant [10.52]. Here, the obtained GNRs consisted of graphene oxide; they were basically an insulator and required reduction treatment for use as graphene. Somewhat later, another method of fabricating GNRs by mildly oxidizing MWCNTs in a gas phase and dispersing the resulting MWCNTs in a DCE solution of PmPV was reported (Fig. 10.23) [10.50]. The obtained GNRs had smooth edges. A mobility of 1500 cm^2/Vs was obtained for the GNRs with a width of 14 nm. At low temperature, the GNRs exhibited coherent transport and Fabry–Perot interference. The methods of fabricating GNRs from CNTs were originally successful only with relatively thick multiwalled CNTs, leading to the formation of relatively wide GNRs. However, recently narrower GNRs with widths of 2–3 nm were also obtained from single-walled CNTs [10.53].

It was also reported that graphene can be preferentially formed on narrow twin crystal regions of Cu films by carefully tuning the CVD conditions [10.54]. Such twin crystal regions are often obtained when a Cu film is annealed at high temperature. Figure 10.24 shows SEM images of graphene ribbons formed on the twin crystal regions. Such regions have Cu(001) or high index surfaces, in contrast to the other area having a Cu(111) surface. In fact, simulations revealed that graphene nucleation is more likely to occur on Cu(001) than Cu(111). Although GNRs obtained by this approach are still about 100 nm wide, it was reported that much narrower and denser twin crystal regions can be obtained [10.55]. So, it seems possible to obtain narrower GNRs by this approach. Furthermore, *Jacobberger* et al. [10.56] has recently demonstrated that GNRs can be formed on Ge(001) by CVD. The GNRs were self-aligning 3° from the Ge(110) directions, and predominantly had armchair edges. The widths of GNRs were as narrow as 10 nm and the aspect ratio was as high as 70. The mechanism of the preferential growth was discussed in the manuscript, but does not seem to be conclusive yet. Further studies would be necessary for understanding such interesting growth manner.

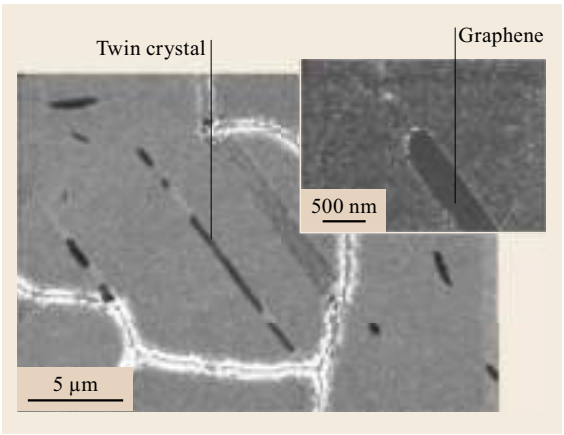


Fig. 10.24 SEM images of graphene nanoribbons formed on Cu twin crystals. (Reprinted by permission from [10.54], copyright 2010 American Chemical Society)

Graphene nanoribbons can also be formed on SiC surface, as shown in Fig. 10.25 [10.57]. In this technique, controlled facets are achieved by patterning

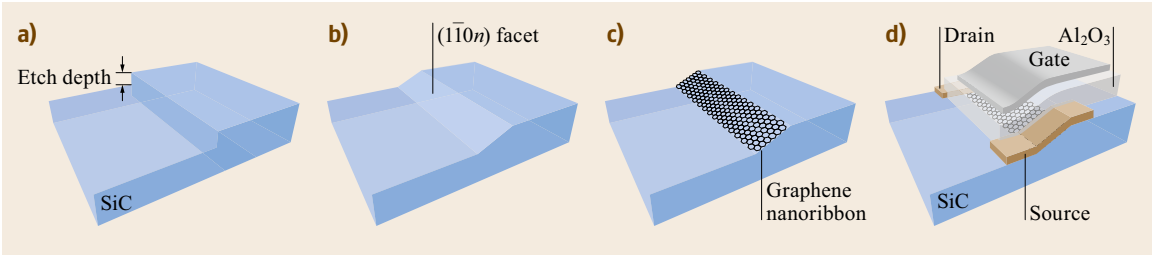


Fig. 10.25a–d Process for selective graphene nanoribbon growth. **(a)** A nanometer-scale step is etched into SiC by fluorine-based reactive ion etching. **(b)** The SiC is heated to 1200–1300 °C (at low vacuum), inducing step flow and relaxation to the (1–10n) facet. **(c)** Upon further heating to 1450 °C, graphene nanoribbon forms on the facet in a self-organized manner. **(d)** Device structure with source and drain contacts, GNR channel, Al_2O_3 gate dielectric and top gate (after [10.57])

nickel lines on a SiC substrate perpendicular to the $< 1{-}100 >$ direction; these lines are transferred into the SiC by fluorine-based reactive ion etching (RIE), controlling the etch depth in nanometer precision. As shown in Fig. 10.25, the etch depth ultimately defines the width of the GNRs. Etch depths of 20 nm were easily achieved, resulting in a ribbon width (40 nm) sufficiently narrow to give a sizable bandgap at room temperature. After removal of the nickel mask and cleaning, the crystal is heated to temperatures of 1200–1300 °C at 10^{-4} torr for 30 min, inducing SiC step flow. The abrupt step relaxes to a (110 n) facet, and the temperature is then elevated to 1450 °C within 1.5 min, maintained for 10 min for graphene growth, then allowed to cool naturally, falling below 1300 °C within 15 s. With this technique, GNRs can be obtained all over a large substrate. Forming narrower GNRs is, however, required for application to transistor channels.

10.5 Potential Applications

10.5.1 Application to Transistors

Graphene nanoribbons have intrinsically excellent and interesting properties, and thus there are many potential applications. One of the promising applications is transistors. As described in the previous sections, the band structure varies with the ribbon width and edge structure. According to theoretical calculations, the bandgap changes from ≈ 5 eV down to ≈ 0 eV by controlling the structure, which is amazing and almost impossible with materials other than carbon. Not only logic circuit applications with small-bandgap GNRs but also high-power applications with large-bandgap GNRs may be possible. In fact, large-scale integrated circuits (LSIs), whose performance has continuously been improved by scaling (miniaturization), is now facing the physical limitation of scaling. Therefore, the semiconductor devices will have to be improved by methods other than scaling in the near future. Employing new channel materials with high mobility, such as graphene and CNTs, is one of the promising candidates. Like carbon nanotubes (CNTs), long ballistic lengths in GNRs can make a ballistic transistor with excellent transport properties realistic. It should be noted here that, although it is generally understood that there is a trade-off between the mobility and the bandgap for nonballistic transport, this understanding does not simply apply to the ballistic regime: transport properties should be understood based on the first-principles calculations [10.58].

As for characteristics of GNR transistors, several theoretical studies have been performed [10.59–62], typically based on the nonequilibrium Green's function formalism. Overall understanding is that the performance of GNR transistors is comparable with that of CNT transistors, and better than that of silicon complementary metal–oxide–semiconductors (CMOSs) or nanowire transistors. A typical transistor structure is illustrated in Fig. 10.26. A top-gate structure is normally used in real-world applications, where the electronic band in the channel is electrostatically modulated by the top-gate bias in order to turn on the channel current from the source to drain electrodes. Experimentally, however, a bottom-gate structure is often employed for simplicity. Simulation results by *Fiori* and *Iannaccone* [10.61] are shown in Fig. 10.27. In this figure, transfer characteristics of both GNR and CNT transistors are shown. As for the GNR transistor, an array of GNRs is assumed. The width (W) of each GNR is assumed to be 1.37 nm, and the length (L), 15 nm. The space between the GNRs is 4 nm. The gate-oxide thickness is assumed to be 1 nm. In this case, the bandgap was calculated to be 0.6 eV using the tight-

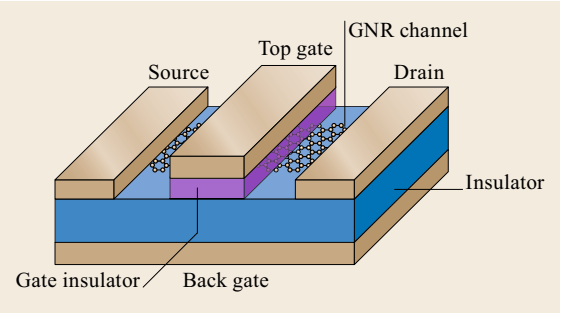


Fig. 10.26 Typical transistor structure with a GNR as a channel

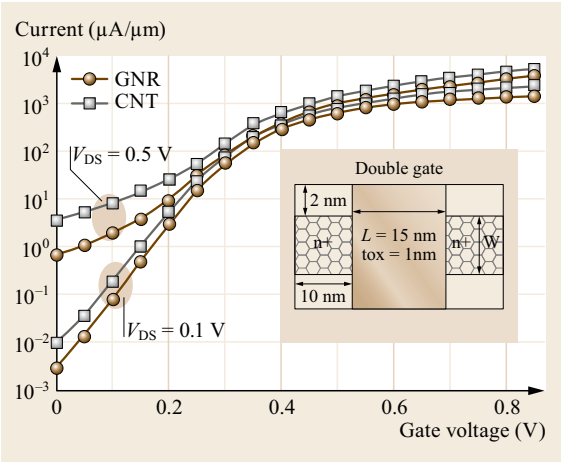


Fig. 10.27 Transfer characteristics of double-gated CNT and GNR field-effect transistors (FETs), with channel length of 15 nm, gate-oxide thickness (tox) of 1 nm, and channel of 1.37 nm. *Inset*: Illustration of the GNR-FET. The source and drain extensions are 10 nm long. The spacing between parallel GNRs is 4 nm, the double of 2 nm shown in the illustration (after [10.61])

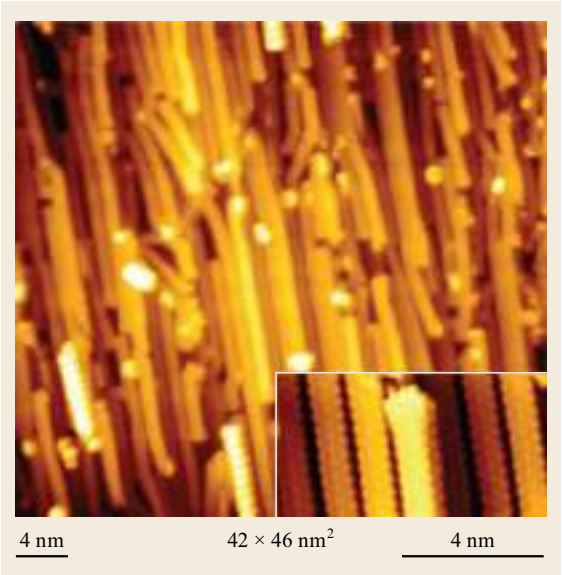


Fig. 10.28 STM images of aligned 7-AGNRs on Au(788) surfaces. Scale bars: 4 nm. (Reprinted with permission from [10.63], copyright 2012 American Physical Society)

binding approach, which was smaller than that obtained by calculations using a first-principles many-electron Green's function approach within the GW approximation [10.17]. Because of this small bandgap, the off-current is relatively large when the drain voltage, V_{DS}, is 0.5 V. This is caused by band-to-band tunneling. However, the off-current is suppressed when the drain voltage is 0.1 V. Surprisingly, even when the drain voltage is 0.1 V, an on-current larger than 1 mA/μm is obtained, which is a requirement for high-performance transistors. Since the power consumption of a transistor during operation is proportional to the square of the supply voltage, this implies that GNR transistors can greatly reduce the power consumption of electronics devices, such as large-scale integrated circuits. Of course, if the edge roughness exists, the transistor performance is degraded, as also shown by *Fiori* and *Iannaccone* [10.61]. However, atomically precise bottom-up GNRs explained in the former section can be a solution to this problem. Moreover, *Linden* et al. demonstrated densely aligned growth of AGNRs on vicinal Au(778) surface as shown in Fig. 10.28 [10.63], which indicates an array of GNRs can actually be available as transistor channels.

It may be noticed that, in Fig. 10.27, the gate voltage required to obtain a large drain current (> 1 mA/μm) is still as large as 0.5 V or so. This means that, even if the drain voltage is lowered to 0.1 V, the supply voltage should remain high. This is mainly caused by the operation principles of the transistors described above. In fact, for MOSFETs (metal-oxide-semiconductor field-effect transistor), subthreshold swing (voltage required to increase the drain current by one decade in the subthreshold region) has to be equal to or larger than 60 meV/decade at room temperature. Therefore, for example, in order to obtain the current on-off ratio larger than 10^5, the gate voltage of 0.3 V or larger would be required. This is also a problem for the conventional MOSFETs using silicon. As a solution to this problem, a new-principle transistor, tunnel FET (TFET), is actively studied, which makes use of band-to-band tunneling to turn on the current [10.65]. GNRs can also be applied to TFETs and theoretical simulations show that the on-current of GNR TFETs can be more than two orders of magnitude higher than that of Si TFETs [10.66, 67]. In fact, the low on-current of Si TFETs, which is

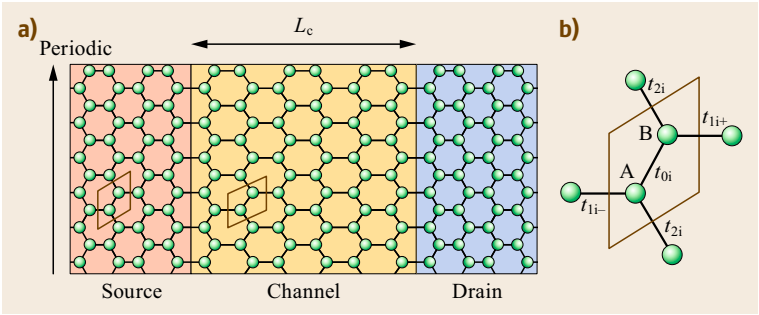


Fig. 10.29 **(a)** A schematic illustration of the model structure, where the graphene is stretched along the armchair direction over the gated channel region. The *left* and the *right* unstrained graphene regions act as the source and drain electrodes. **(b)** shows the graphene unit cell consisting of two independent sublattice sites A and B interacting through transfer integrals t. Stretching makes t_{1i+} and t_{1i-} along the stretching direction weakened (after [10.64])

caused by the indirect bandgap of Si, is a serious obstacle for practical realization [10.65] and GNR TFETs can be a solution.

Another interesting new-principle transistor using graphene is illustrated in Fig. 10.29 [10.64]. In this transistor, an axially strained graphene is used as a channel, and unstrained graphene is used for the source and drain. The strained graphene has the Dirac point in a different k-space position. When the gate voltage is applied, the conduction bands of the strained graphene and unstrained graphene can overlap in the k space, thus turning on the transistor. This transistor behavior is similar to TFETs described above, and subthreshold swing less than 60 meV/decade can be achieved. Although this transistor is designed for gapless graphene, this concept can also work for relatively wide GNRs.

Experimental studies on GNR transistors have also been performed. So far, the best performance was obtained with GNRs described in the beginning of Sect. 10.4.3 [10.13]. In fact, an on-off current ratio of 10^5–10^7 was obtained for the GNR channels with a width < 10 nm at room temperature. Typical mobility for such GNRs was 100–200 cm^2/Vs. This combination of the on-off current ratio (10^5–10^7) and mobility (100–200 cm^2/Vs) is possibly the best reported combination for GNRs, although the mobility is not yet as good as expected. Although GNRs with widths narrower than 10 nm have been obtained by top-down approaches as described in Sect. 10.4.1, the typical on-off current ratio of the transistors with such GNRs was only 10–100. The drain current per channel width is also greatly degraded by the etching process, resulting in poor mobility. This poor performance is probably caused by the rough edges of such GNRs. Possible residues from the top-down processes can also be a reason. On the other hand, the GNR array demonstrated excellent gas sensing properties [10.36], detecting NO_2 with concentrations down to ppb levels, which is much better than graphene-based gas sensors [10.68].

Transistors using atomically precise bottom-up GNRs have also been fabricated recently [10.69, 70]. Transistors using 7-AGNR show an on-current of ≈ 1 nA/GNR and an on-off current ratio of $\approx 10^3$ [10.69]. These results are again poorer than expected, which may be caused by high Schottky barriers at the source and drain contacts and/or possible contamination on the channel. Transistors using 9-AGNRs and 13-AGNRs as channels have also been fabricated. These transistors exhibited on-currents as high as 100 nA at a drain bias of 1 V and on-off ratios of 10^2–10^5 [10.70]. The improvement in on-current is probably due to lower Schottky barriers originated from the smaller bandgaps of 9-AGNRs and 13-AGNRs. Further improvement is expected by improving the device fabrication processes, which is still ongoing.

10.5.2 Interconnect Application

GNR is also a good candidate for interconnects for scaled LSI and molecular devices. In fact, along with the miniaturization of devices, LSI interconnects are facing two problems: decreased reliability and increased effective resistivity of Cu interconnects. According to the International Technology Roadmap for Semiconductors 2013, the resistivity of Cu interconnects is estimated to exceed 10 $\mu\Omega$ cm (1.7 $\mu\Omega$ cm for bulk Cu) by about 2020 when the linewidth is predicted to have become narrower than 10 nm. GNRs along with CNTs can be a solution to these problems.

It has been reported that the reliability of graphene interconnects, as well as CNT interconnects, is superior to that of Cu interconnects. The breakdown current densities of both exfoliated graphene and CVD graphene are of the order of 10^8–10^9 A/cm^2 [10.30, 71, 72]. In the study of *Murali* et al. [10.71], GNRs with widths

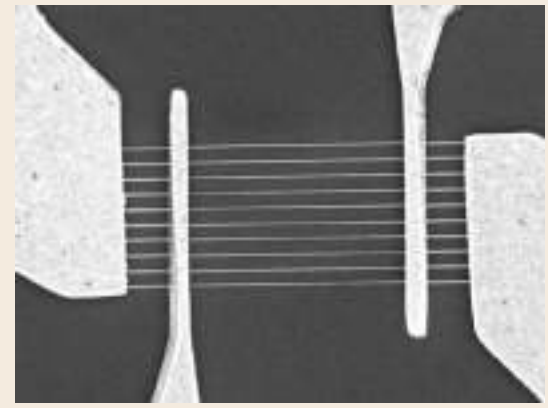


Fig. 10.30 SEM image of ten GNRs with electrodes. The GNRs below hydrogen silsesquioxane (HSQ) lines are 21 nm-wide between the middle electrode pair. (Reprinted with permission from [10.71], copyright 2009 AIP Publishing)

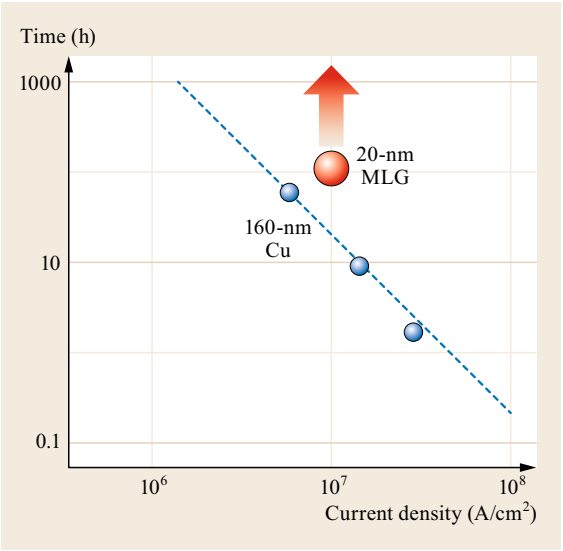


Fig. 10.31 Results for reliability tests. The *vertical axis* shows the time to failure. The *horizontal axis* shows the current density applied. The *dotted line* is a fitting line for Cu interconnects. The 20 nm-wide multilayer graphene (MLG) interconnect did not break after 100 h under a current density of 10^7 A/cm^2 (after [10.30, 73])

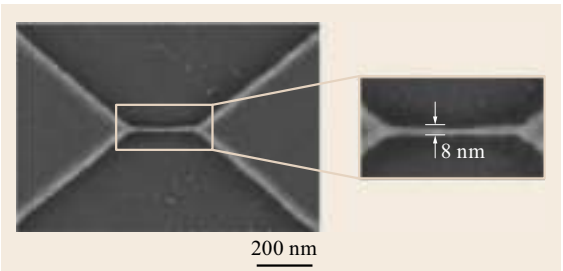


Fig. 10.32 Multilayer (ten-layer) graphene interconnect patterned by electron beam lithography. Resist for electron beam lithography, HSQ, is left on the graphene interconnects (after [10.30])

as narrow as 16 nm were fabricated by electron beam lithography and exhibited a high breakdown current (Fig. 10.30). Figure 10.31 shows the results of a reliability test conducted for 20 nm-wide GNR interconnects made of CVD graphene [10.30, 73]. The current density durability of the CVD graphene was superior to that of Cu. Although the results explained here are for relatively wide GNRs, the reliability of narrower GNRs is also expected to be good.

The transport properties of GNR interconnects have theoretically been predicted [10.74, 75]. Results are somewhat different from each other possibly due to the differences in assumption. The general understanding is that the resistivity of GNRs can be better than that of scaled Cu interconnects if the ballistic transport is realized. Doping to graphene can improve the resistivity further. However, in fact, the fabrication of graphene interconnects having resistance lower than Cu interconnects is not an easy task. Interconnects with a low resistivity were not obtained simply by patterning graphene by lithography. In such cases, electron transport tends to be diffusive, and the resistance becomes comparable with that of bulk graphite, which is about 40 $\mu\Omega$ cm, one order of magnitude higher than that of Cu (1.7 $\mu\Omega$ cm).

Doping by intercalation is an option to reduce the resistivity of GNRs, as mentioned above. Lowering the resistivity of graphite by intercalation was first reported many years ago [10.77]. Several materials have been reported to achieve resistivity lower than that of Cu by intercalation to graphite. *Xu* et al. theoretically demonstrated that AsF_5-doped multilayer GNRs can have resistivity lower than that of Cu [10.74]. Murali experimentally demonstrated that an electrostatically doped exfoliated GNR can have a resistivity comparable to that of Cu. *Kondo* et al. patterned CVD-grown graphene by electron beam lithography to fabricate GNRs as narrow as 8 nm, as shown in Fig. 10.32 and obtained a resistivity of 3.2 $\mu\Omega$ cm by the intercalation of $FeCl_3$ [10.30, 72, 73]. This value is actually better than the predicted resistivity of Cu interconnects with similar dimensions.

As mentioned above, obtaining GNRs exhibiting ballistic transport has been awaited, but is not an easy task. Use of atomically precise bottom-up

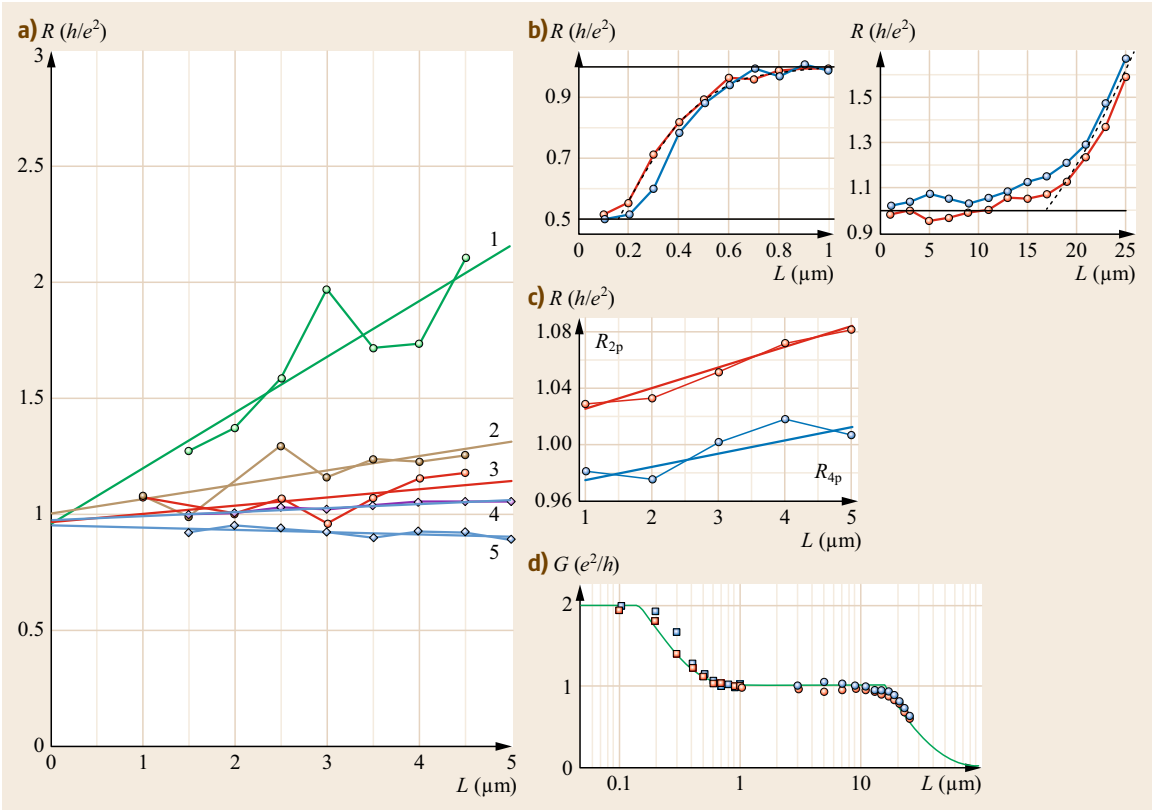


Fig. 10.33 **(a)** Multiprobe in situ transport measurements of graphene nanoribbons, showing resistance versus probe spacing, L. Linear fits, shown as five straight lines, extrapolate to $R_0 = h/e^2$. The slopes of the fitting lines 1–4 correspond to $R' = 6.2,\ 1.6,\ 0.92,\ 0.44\,\mathrm{k}\Omega/\mu\mathrm{m}$, corresponding to the mean free path, $\lambda_0 = 4.2, 28, 16, 58\,\mu\mathrm{m}$ (estimated by the following relationship: transmission, $\mathrm{Tr} \approx (1 + L/\lambda_0)^{-1}$). Fitting line 5 is consistent with zero slope. Numbered traces are as follows: 1, UHV annealed at 1100 °C for 15 min; 2, UHV annealed at 1150 °C for 15 min; 3–5, re-annealed at 1150 °C for 15 min. **(c)** Comparison of two-probe (2p) and four-probe (4p) measurements. **(b)** Nonlinear resistance increases observed at $L = 160\,\mu\mathrm{m}$ and at $L = 16\,\mu\mathrm{m}$ in two different ribbons measured at room temperature, presented as conductance, $G(L)$ **(d)** (after [10.76])

GNRs is one of the most promising options, and it is actually on the way. However, recently, ballistic transport at room temperature has been observed for GNRs formed on SiC [10.76]. The GNR structures were grown selectively by sublimation epitaxy on mesastructured 6H-SiC(0001) surfaces, as illustrated in Fig. 10.25 [10.57]. GNRs with a width of 40 nm exhibited a ballistic length greater than 10 μm (Fig. 10.33), which was equivalent to sheet resistances below 1 Ω/square. This is really a surprising, but promising, result for future nanoelectronics using GNRs.

10.5.3 Other Potential Applications

Graphene nanoribbons have other promising applications. One of them is application to spintronics. As described in the previous sections, ZGNRs have spin-polarized edge states. The orientations of the spins localized at the two edges are opposite to each other as shown in Fig. 10.34 and ZGNRs have antiferromagnetic properties overall. *Son* et al. [10.78] showed theoretically that ZGNRs can be half-metallic by applying an in-plane electric field perpendicular to the edge directions. Later, it was also predicted that the half metallicity can be obtained by modifying each edge with different chemical functional groups, such as NO_2 and CH_3 [10.79]. If GNRs are half-metallic, they can be used to inject spin-polarized electrons for spintronics. Furthermore, *Magada* et al. experimentally showed that ZGNR exhibited transition from antiferromagnetic to ferromagnetic configurations when they became wider than about 7 nm [10.80]. This implies that ZGNRs along with AGNRs can be used as the main

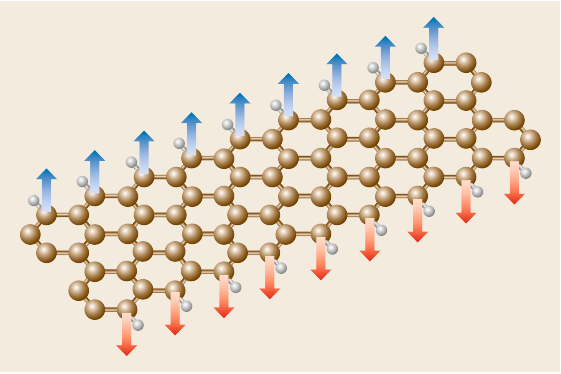


Fig. 10.34 ZGNR with edge-localized spins. The polarized spins at the two edges are opposite to each other

building blocks for spintronic devices, including magnetroresistive random access memory (MRAM), spin FETs and spin MOSFETs.

Graphene nanoribbons are also a candidate for future computing devices. In fact, *Trauzettel* et al. proposed to form spin qubits in graphene quantum dots [10.81]. In their proposal, the quantum dots are made of a gated AGNR. Because the spin-orbit coupling is weak and the hyperfine interaction is small in graphene, the spin coherence time can be large. Moreover, spin qubits in graphene cannot only be coupled between nearest-neighbor quantum dots but also over long distances, according to their explanation. A schematic diagram of the proposed graphene quantum dots is shown in Fig. 10.35. Such quantum dots need AGNRs with perfect edges, which can be formed using the bottom-up technique described in the former section. Quantum computers using AGNRs may be realized in the future.

Graphene nanoribbons can also be used for photovoltaic devices. *Konabe* and *Okada* theoretically predicted that multiple-exciton generation occurs in AGNRs wherein the efficiency of photoelectric conversion exceeds 100% [10.82, 83]. This suggests that high-efficiency and flexible solar cells can be realized with AGNRs. For this purpose, mass production of AGNRs in liquid chemicals may be necessary. Another application might be as additives to electrodes of secondary batteries and fuel cells. Functionalized GNRs could also be used as catalysts. If the technology to form GNRs with arbitrary structures is established in the future, many more applications are expected to be realized.

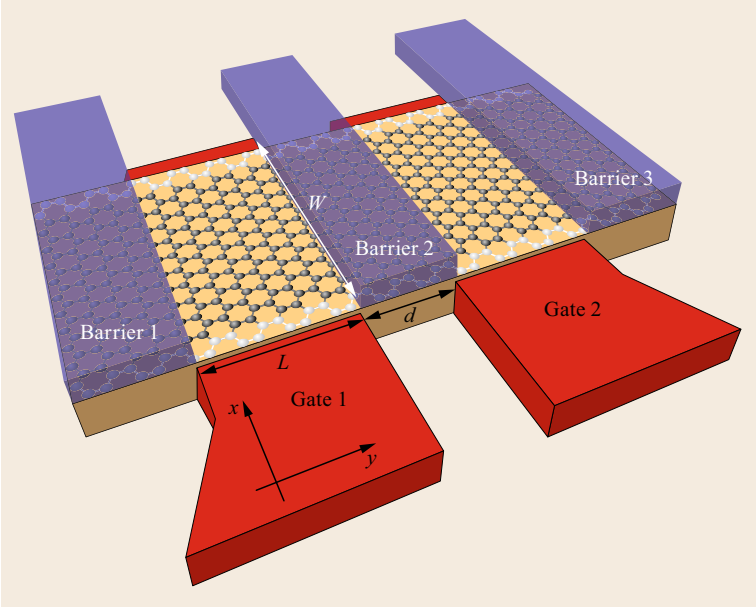


Fig. 10.35 Schematic diagram of graphene quantum dots (after [10.81])

10.6 Conclusions

Graphene is a one-atom-thick 2-D sheet, in which carbon atoms are arranged in a hexagonal honeycomb lattice. When an infinite graphene sheet is cut into nanofragments, graphene nanostructures are created. Among graphene nanostructures created by cutting, we have ribbon-shaped nanostructures called graphene nanoribbons. What makes graphene nanostructures different from infinite graphene and also from carbon nanotubes and fullerenes is the presence of open edges in the former. Owing to the presence of open edges, these graphene nanostructures have electronic structures strongly modified from an infinite graphene having features of a zero-gap semiconductor. There are two fundamental edge geometries; that is, armchair edge and zigzag edge, and how the electronic structure is modified depends crucially on the edge geometry. The modifications can be understood on the basis of the boundary condition, which is given by the open edges, for the electrons having the features of massless Dirac fermion in graphene. In the armchair edge, electron wave interference takes place, resulting in energetic stabilization, whereas the broken symmetry of the pseudospin of the Dirac fermion is responsible for the creation of a nonbonding edge state in the zigzag edge. This can also be interpreted in terms of aromatic stability and creation of unpaired electrons for armchair and zigzag edges respectively, in chemistry, taking into account that graphene nanofragments are considered as nanosized polycyclic aromatic hydrocarbon molecules.

Accordingly the electronic structures of graphene nanoribbons having armchair and zigzag edges are owing to these geometry-dependent modifications. Namely, armchair-edged graphene nanoribbons, which are energetically stable, have energy gaps, which depend on the ribbon widths, while zigzag-edged nanoribbons have nonbonding edge states, which are spin polarized and localized in the vicinity of the zigzag edges. The zigzag-edged nanoribbons become electronically, magnetically and chemically active at the expense of aromatic stability. It should be noted that how edge carbon atoms are terminated with functional groups also modifies the electronic structures in addition to the edge-geometry-originating modifications. Therefore the combination of chemistry details and edge geometry gives rise to a large variety of electronic properties.

The properties of graphene nanoribbons and also the structures of graphene edges can be characterized using various experimental techniques and also first-principle theoretical calculations. Microprobe experiments such as STM/STS and AFM together with DFT calculations give important local information on the electronic structures and also the atomic arrangements in atomic scale precision. The transmission electron microscope is also a useful tool in getting information not only of atomic arrangements but also of local electronic structures. Raman spectroscopy is particularly important in characterizing the edge geometry, as the resonant Raman signals, which are sensitive even for a single graphene nanostructure, have distinct edge geometry dependence. Electron transport measurements are indispensable for electronic structure characterizations.

Synthesis of graphene nanoribbons can be carried out basically using top-down and bottom-up methods. In the top-down methods, electron beam lithography is most important, though future improvement is required for fabricating sub-10 nm-width nanoribbons. Direct patterning of graphene nanoribbons by focused helium ion beams in a helium ion microscope can reach a 5 nm-wide graphene nanoribbon array. STM lithography can go to a width of 2.5 nm. The drawback in the top-down technique is difficulty in obtaining atomically precise edges, which theoretical predictions on the geometry-dependent electronic structures require. Bottom-up methods are important in fabricating graphene nanoribbons that have atomically precise edges. Polymerization of precursor monomer molecules with the aid of the catalytic activity of metal substrates gives a variety of armchair- and zigzag-edged graphene nanoribbons and also graphene nanoribbons having wiggled edges and chiral edges. We can create also nanoribbons in which heteroatoms are doped.

The presence of bandgap in graphene nanoribbons, armchair-edged graphene nanoribbons in particular, is important in transistor device applications. The bandgap, which varies depending on the ribbon width, is advantageous in fabricating electronics devices. Graphene nanoribbons are a good candidate for interconnects for scaled LSI and molecular devices. Zigzag-edged graphene nanoribbons, which have spin-polarized edge states in the vicinity of the edges, are expected to be useful in spintronics applications.

References

10.1 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov: Electric field effect in atomically thin carbon films, Science **306**, 666–669 (2004)

10.2 M.S. Dresselhaus, G. Dresselhaus, P. Eklund: *Science of Fullerenes and Carbon Nanotubes: Their Properties and Applications* (Academic, San Diego 1996)

10.3 R. Saito, M.S. Dresselhaus, G. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College Press, London 1998)

10.4 A.H.C. Neto, F. Guinea, N.M. Peres, K.S. Novoselov, A.K. Geim: The electronic properties of graphene, Rev. Mod. Phys. **81**, 109–162 (2009)

10.5 T. Enoki, T. Ando: *Physics and Chemistry of Graphene; Graphene to Nanographene* (Pan Stanford, Singapore 2013)

10.6 S. Fujii, T. Enoki: Nanographene and graphene edges: Electronic structure and nanofabrication, Acc. Chem. Res. **46**, 2202–2210 (2013)

10.7 E. Clar: *The Aromatic Sextet* (Wiley, London 1972)

10.8 T. Wassmann, A.P. Seitsonen, A.M. Saitta, M. Lazzeri, F. Mauri: Clar's theory, π-electron distribution and geometry of graphene nanoribbons, J. Am. Chem. Soc. **132**, 3440–3451 (2010)

10.9 M. Fujita, K. Wakabayashi, K. Nakada, K. Kusakabe: Peculiar localized state at zigzag graphite edge, J. Phys. Soc. Jpn. **65**, 1920–1923 (1996)

10.10 S. Fujii, M. Ziatdinov, M. Ohtsuka, K. Kusakabe, M. Kiguchi, T. Enoki: Role of edge geometry and chemistry in the electronic properties of graphene nanostructures, Faraday Discuss. **173**, 173–199 (2014)

10.11 D. Bischoff, J. Güttinger, S. Dröscher, T. Ihn, K. Ensslin, C. Stampfer: Raman spectroscopy on etched graphene nanoribbons, J. Appl. Phys. **109**, 073710 (2011)

10.12 L. Tapasztó, G. Dobrik, P. Lambin, L.P. Biró: Tailoring the atomic structure of graphene nanoribbons by scanning tunneling microscope lithography, Nat. Nanotechnol. **3**, 397–401 (2008)

10.13 X. Li, X. Wang, L. Zhang, S. Lee, H. Dai: Chemically derived, ultrasmooth graphene nanoribbon semiconductors, Science **319**, 1229–1232 (2008)

10.14 L. Chen, Y. Hernandez, X. Feng, K. Müllen: Bottom up from nanographene and graphene nanoribbons to graphene sheets: Chemical synthesis, Angew. Chem. Int. Ed. **51**, 7640–7654 (2012)

10.15 K. Sasaki, R. Saito: Pseudospin and deformation-induced gauge field in graphene, Prog. Theor. Phys. Suppl. **176**, 253–278 (2008)

10.16 E. Lieb: Two theorems on the Hubbard model, Phys. Rev. Lett. **62**, 1201–1204 (1989)

10.17 L. Yang, C.-H. Park, Y.-W. Son, M.L. Cohen, S.G. Louie: Quasiparticle energies and bandgaps in graphene nanoribbons, Phys. Rev. Lett. **99**, 186801 (2007)

10.18 Y.-W. Son, M.L. Cohen, S.G. Louie: Energy gaps in graphene nanoribbons, Phys. Rev. Lett. **97**, 216803 (2006)

10.19 Y.-C. Chen, D.G. de Oteyza, Z. Pedramrazi, C. Chen, F.R. Fischer, M.F. Crommie: Tuning the bandgap of graphene nanoribbons synthesized from molecular precursors, ACS Nano **7**, 6123–6128 (2013)

10.20 P. Ruffieux, J. Cai, N.C. Plumb, L. Patthey, D. Prezzi, A. Ferretti, E. Molinari, X. Feng, K. Müllen, C.A. Pignedoli, R. Fasel: Electronic structure of atomically precise graphene nanoribbons, ACS Nano **8**, 6930–6935 (2012)

10.21 A.A. Jorio, M.S. Dresselhaus, R. Saito, G. Dresselhaus: *Raman Spectroscopy in Graphene Related System* (Wiley, Weinheim 2011)

10.22 L.G. Cançado, M.A. Pimenta, B.R.A. Neves, M.S.S. Dantas, A. Jorio: Influence of the atomic structure on the Raman spectra of graphite edges, Phys. Rev. Lett. **93**, 247401 (2004)

10.23 L.G. Cançado, M.A. Pimenta, B.R.A. Neves, G. Medeiros-Ribeiro, T. Enoki, Y. Kobayashi, K. Takai, K. Fukui, M.S. Dresselhaus, R. Saito, A. Jorio: Anisotropy of the Raman spectra of nanographite ribbons, Phys. Rev. Lett. **93**, 047403 (2004)

10.24 K. Sasaki, R. Saito, K. Wakabayashi, T. Enoki: Identifying the orientation of edge of graphene using G-band Raman spectra, J. Phys. Soc. Jpn. **79**, 044603 (2010)

10.25 M. Kiguchi, K. Takai, V.L.J. Joly, T. Enoki, R. Sumii, K. Amemiya: Magnetic edge state and dangling bond state on nanographene in activated carbon fibers, Phys. Rev. B **84**, 045421 (2011)

10.26 Z. Hou, X. Wang, T. Ikeda, S.-F. Huang, K. Terakura, M. Boero, M. Oshima, M. Kakimoto, S. Miyata: Effect of hydrogen termination on carbon K-edge x-ray absorption spectra of nanographene, J. Phys. Chem. C **115**, 5392–5403 (2011)

10.27 K. Suenaga, M. Koshino: Atom-by-atom spectroscopy at graphene edge, Nature **468**, 1088–1090 (2010)

10.28 K. Sugawara, T. Sato, S. Souma, T. Takahashi, H. Suematsu: Fermi surface and edge-localized states in graphite studied by high-resolution angle-resolved photoemission spectroscopy, Phys. Rev. B **73**, 045124 (2006)

10.29 M.Y. Han, B. Özyilmaz, Y. Zhang, P. Kim: Energy band-gap engineering of graphene nanoribbons, Phys. Rev. Lett. **98**, 206805 (2007)

10.30 D. Kondo, H. Nakano, B. Zhou, I.A.K. Hayashi, M. Takahashi, S. Sato, N. Yokoyama: Sub-10-nm-wide intercalated multi-layer graphene interconnects with low resistivity. In: *IEEE Int. Interconnect Technol./Adv. Metallization Conf.* (2014) pp. 189–192

10.31 P. Gallagher, K. Todd, D. Goldhaber-Gordon: Disorder-induced gap behavior in graphene nanoribbons, Phys. Rev. B **81**, 115409 (2010)

10.32 M.Y. Han, J.C. Brant, P. Kim: Electron transport in disordered graphene nanoribbons, Phys. Rev. Lett. **104**, 056801 (2010)

10.33 X. Liang, S. Wi: Transport characteristics of multichannel transistors made from densely aligned sub-10 nm half-pitch graphene nanoribbons, ACS Nano **6**, 9700–9710 (2012)

10.34 H. Suzuki, T. Kaneko, Y. Shibuta, M. Ohno, Y. Maekawa, T. Kato: Wafer-scale fabrication and growth dynamics of suspended graphene nanoribbon arrays, Nat. Commun. **7**, 11797 (2016)

10.35 S. Nakaharai, T. Iijima, S. Ogawa, H. Miyazaki, S. Li, K. Tsukagoshi, S. Sato, N. Yokoyama: Gating operation of transport current in graphene nanoribbon fabricated by helium ion microscope. In: *Int. Conf. Solid State Devices Mater., Nagoya* (2011) p. 1300

10.36 A.N. Abbas, G. Liu, B. Liu, L. Zhang, H. Liu, D. Ohlberg, W. Wu, C. Zhou: Patterning, characterization and chemical sensing applications of graphene nanoribbon arrays down to 5 nm using helium ion beam lithography, ACS Nano **8**, 1538–1546 (2014)

10.37 J. Cai, P. Ruffieux, R. Jaafar, M. Bieri, T. Braun, S. Blankenburg, M. Muoth, A.P. Seitsonen, M. Saleh, X. Feng, K. Mullen, R. Fasel: Atomically precise bottom-up fabrication of graphene nanoribbons, Nature **466**, 470–473 (2010)

10.38 P. Ruffieux, S. Wang, B. Yang, C. Sánchez-Sánchez, J. Liu, T. Dienel, L. Talirz, P. Shinde, C.A. Pignedoli, D. Passerone, T. Dumslaff, X. Feng, K. Müllen, R. Fasel: On-surface synthesis of graphene nanoribbons with zigzag edge topology, Nature **531**, 489–492 (2016)

10.39 P. Han, K. Akagi, F. Federici Canova, H. Mutoh, S. Shiraki, K. Iwaya, P.S. Weiss, N. Asao, T. Hitosugi: Bottom-up graphene-nanoribbon fabrication reveals chiral edges and enantioselectivity, ACS Nano **8**, 9181–9187 (2014)

10.40 M. Koch, F. Ample, C. Joachim, L. Grill: Voltage-dependent conductance of a single graphene nanoribbon, Nat. Nano **7**, 713–717 (2012)

10.41 J. Björk, S. Stafström, F. Hanke: Zipping up: Cooperativity drives the synthesis of graphene nanoribbons, J. Am. Chem. Soc. **133**, 14884–14887 (2011)

10.42 J. Björk, F. Hanke, S. Stafström: Mechanisms of halogen-based covalent self-assembly on metal surfaces, J. Am. Chem. Soc. **135**, 5768–5775 (2013)

10.43 K.A. Simonov, N.A. Vinogradov, A.S. Vinogradov, A.V. Generalov, E.M. Zagrebina, N. Mårtensson, A.A. Cafolla, T. Carpy, J.P. Cunniffe, A.B. Preobrajenski: Effect of substrate chemistry on the bottom-up fabrication of graphene nanoribbons: Combined core-level spectroscopy and STM study, J. Phys. Chem. C **118**, 12532–12540 (2014)

10.44 C. Bronner, J. Björk, P. Tegeder: Tracking and removing Br during the on-surface synthesis of a graphene nanoribbon, J. Phys. Chem. C **119**, 486–493 (2015)

10.45 J. Liu, B.-W. Li, Y.-Z. Tan, A. Giannakopoulos, C. Sanchez-Sanchez, D. Beljonne, P. Ruffieux, R. Fasel, X. Feng, K. Müllen: Toward cove-edged low bandgap graphene nanoribbons, J. Am. Chem. Soc. **137**, 6097–6103 (2015)

10.46 R.R. Cloke, T. Marangoni, G.D. Nguyen, T. Joshi, D.J. Rizzo, C. Bronner, T. Cao, S.G. Louie, M.F. Crommie, F.R. Fischer: Site-specific substitutional boron doping of semiconducting armchair graphene nanoribbons, J. Am. Chem. Soc. **137**, 8872–8875 (2015)

10.47 S. Kawai, S. Saito, S. Osumi, S. Yamaguchi, A.S. Foster, P. Spijker, E. Meyer: Atomically controlled substitutional boron-doping of graphene nanoribbons, Nat. Commun. **6**, 8098 (2015)

10.48 H. Zhang, H. Lin, K. Sun, L. Chen, Y. Zagranyarski, N. Aghdassi, S. Duhm, Q. Li, D. Zhong, Y. Li, K. Müllen, H. Fuchs, L. Chi: On-surface synthesis of rylene-type graphene nanoribbons, J. Am. Chem. Soc. **137**, 4022–4025 (2015)

10.49 A. Basagni, F. Sedona, C.A. Pignedoli, M. Cattelan, L. Nicolas, M. Casarin, M. Sambi: Molecules–oligomers–nanowires–graphene nanoribbons: A bottom-up stepwise on-surface covalent synthesis preserving long-range order, J. Am. Chem. Soc. **137**, 1802–1808 (2015)

10.50 L. Jiao, X. Wang, G. Diankov, H. Wang, H. Dai: Facile synthesis of high-quality graphene nanoribbons, Nat. Nano **5**, 321–325 (2010)

10.51 L.Y. Jiao, L. Zhang, X.R. Wang, G. Diankov, H.J. Dai: Narrow graphene nanoribbons from carbon nanotubes, Nature **458**, 877–880 (2009)

10.52 D.V. Kosynkin, A.L. Higginbotham, A. Sinitskii, J.R. Lomeda, A. Dimiev, B.K. Price, J.M. Tour: Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons, Nature **458**, 872–876 (2009)

10.53 Y. Gong, M. Long, G. Liu, S. Gao, C. Zhu, X. Wei, X. Geng, M. Sun, C. Yang, L. Lu, L. Liu: Electronic transport properties of graphene nanoribbon arrays fabricated by unzipping aligned nanotubes, Phys. Rev. B **87**, 165404 (2013)

10.54 K. Hayashi, S. Sato, M. Ikeda, C. Kaneta, N. Yokoyama: Selective graphene formation on copper twin crystals, J. Am. Chem. Soc. **134**, 12492–12498 (2012)

10.55 L. Lu, Y.F. Shen, X.H. Chen, L.H. Qian, K. Lu: Ultrahigh strength and high electrical conductivity in copper, Science **304**, 422–426 (2004)

10.56 R.M. Jacobberger, B. Kiraly, M. Fortin-Deschenes, P.L. Levesque, K.M. McElhinny, G.J. Brady, R. Rojas Delgado, S. Singha Roy, A. Mannix, M.G. Lagally, P.G. Evans, P. Desjardins, R. Martel, M.C. Hersam, N.P. Guisinger, M.S. Arnold: Direct oriented growth of armchair graphene nanoribbons on germanium, Nat. Commun. **6**, 8006 (2015)

10.57 M. Sprinkle, M. Ruan, Y. Hu, J. Hankinson, M. Rubio-Roy, B. Zhang, X. Wu, C. Berger, W.A. de Heer: Scalable templated growth of graphene nanoribbons on SiC, Nat. Nano **5**, 727–731 (2010)

10.58 S. Datta: *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge 1995)

10.59 Q. Yan, B. Huang, J. Yu, F. Zheng, J. Zang, J. Wu, B.-L. Gu, F. Liu, W. Duan: Intrinsic current-voltage characteristics of graphene nanoribbon transistors and effect of edge doping, Nano Lett. **7**, 1469–1473 (2007)

10.60 Y. Ouyang, Y. Yoon, J. Guo: Scaling behaviors of graphene nanoribbon FETs: A three-dimensional quantum simulation study, IEEE Trans. Electron Devices **54**, 2223–2231 (2007)

10.61 G. Fiori, G. Iannaccone: Simulation of graphene nanoribbon field-effect transistors, IEEE Electron

Device Lett. **28**, 760–762 (2007)
10.62 N. Harada, S. Sato, N. Yokoyama: Theoretical investigation of graphene nanoribbon field-effect transistors designed for digital applications, Jpn. J. Appl. Phys. **52**, 094301 (2013)
10.63 S. Linden, D. Zhong, A. Timmer, N. Aghdassi, J.H. Franke, H. Zhang, X. Feng, K. Müllen, H. Fuchs, L. Chi, H. Zacharias: Electronic structure of spatially aligned graphene nanoribbons on Au(788), Phys. Rev. Lett. **108**, 216801 (2012)
10.64 S. Souma, M. Ueyama, M. Ogawa: Simulation-based design of a strained graphene field effect transistor incorporating the pseudo magnetic field effect, Appl. Phys. Lett. **104**, 213505 (2014)
10.65 A.M. Ionescu, H. Riel: Tunnel field-effect transistors as energy-efficient electronic switches, Nature **479**, 329–337 (2011)
10.66 K.T. Lam, D. Seah, S.K. Chin, S.B. Kumar, G. Samudra, Y.C. Yeo, G. Liang: A simulation study of graphene-nanoribbon tunneling FET with heterojunction channel, IEEE Electron Device Lett. **31**, 555–557 (2010)
10.67 R. Grassi, A. Gnudi, S. Reggiani, E. Gnani, G. Baccarani: Simulation study of graphene nanoribbon tunneling transistors including edge roughness effects. In: *10th Int. Conf. Ultimate Integr. Silicon* (2009) pp. 57–60
10.68 F. Schedin, A.K. Geim, S.V. Morozov, E.W. Hill, P. Blake, M.I. Katsnelson, K.S. Novoselov: Detection of individual gas molecules adsorbed on graphene, Nat. Mater. **6**, 652–655 (2007)
10.69 P.B. Bennett, Z. Pedramrazi, A. Madani, Y.-C. Chen, D.G. de Oteyza, C. Chen, F.R. Fischer, M.F. Crommie, J. Bokor: Bottom-up graphene nanoribbon field-effect transistors, Appl. Phys. Lett. **103**, 253114 (2013)
10.70 J.P. Llinas, A. Fairbrother, G. Barin, P. Ruffieux, W. Shi, K. Lee, B.Y. Choi, R. Braganza, N. Kau, W. Choi, C. Chen, Z. Pedramrazi, T. Dumslaff, A. Narita, X. Feng, F. Fischer, K. Müllen, A. Zettl, M. Crommie, R. Fasel, J. Bokor: Short-channel field effect transistors with 9-atom and 13-atom wide graphene nanoribbons, arXiv:1605.06730 (2016)
10.71 R. Murali, Y. Yang, K. Brenner, T. Beck, J.D. Meindl: Breakdown current density of graphene nanoribbons, Appl. Phys. Lett. **94**, 243114 (2009)
10.72 D. Kondo, H. Nakano, B. Zhou, I. Kubota, K. Hayashi, K. Yagi, M. Takahashi, M. Sato, S. Sato, N. Yokoyama: Intercalated multi-layer graphene grown by CVD for LSI interconnects. In: *IEEE Int. Interconnect Technol. Conf.-IITC* (2013) pp. 1–3
10.73 S. Sato: Nanocarbon interconnects: Demonstration of properties better than Cu and remaining issues. In: *IEEE Int. Interconnect Technol. Conf./IEEE Mater. Adv. Metallization Conf. (IITC/MAM)* (2015) pp. 313–316
10.74 C. Xu, H. Li, K. Banerjee: Modeling, analysis and design of graphene nano-ribbon interconnects, IEEE Trans. Electron Devices **56**, 1567–1578 (2009)
10.75 A. Naeemi, J.D. Meindl: Compact physics-based circuit models for graphene nanoribbon interconnects, IEEE Trans. Electron Devices **56**, 1822–1833 (2009)
10.76 J. Baringhaus, M. Ruan, F. Edler, A. Tejeda, M. Sicot, I. Taleb, A.-P. Li, Z. Jiang, E.H. Conrad, C. Berger, C. Tegenkamp, W.A. de Heer: Exceptional ballistic transport in epitaxial graphene nanoribbons, Nature **506**, 349–354 (2014)
10.77 M.S. Dresselhaus, G. Dresselhaus: Intercalation compounds of graphite, Adv. Phys. **51**, 1–186 (2002)
10.78 Y.-W. Son, M.L. Cohen, S.G. Louie: Half-metallic graphene nanoribbons, Nature **444**, 347–349 (2006)
10.79 E. Kan, Z. Li, J. Yang, J.G. Hou: Half-metallicity in edge-modified zigzag graphene nanoribbons, J. Am. Chem. Soc. **130**, 4224–4225 (2008)
10.80 G.Z. Magda, X. Jin, I. Hagymasi, P. Vancso, Z. Osvath, P. Nemes-Incze, C. Hwang, L.P. Biro, L. Tapaszto: Room-temperature magnetic order on zigzag edges of narrow graphene nanoribbons, Nature **514**, 608–611 (2014)
10.81 B. Trauzettel, D.V. Bulaev, D. Loss, G. Burkard: Spin qubits in graphene quantum dots, Nat. Phys. **3**, 192–196 (2007)
10.82 S. Konabe, N.T. Cuong, M. Otani, S. Okada: High-efficiency photoelectric conversion in graphene–diamond hybrid structures: Model and first-principles calculations, Appl. Phys. Express **6**, 045104 (2013)
10.83 S. Konabe, S. Okada: Multiple exciton generation by a single photon in single-walled carbon nanotubes, Phys. Rev. Lett. **108**, 227401 (2012)

11. Nanoparticles and Their Applications

Seyedsina Moeinzadeh, Esmaiel Jabbari

Nanoparticles (NPs) are synthesized from several classes of materials including inorganic, organic, hybrid and biological materials. Inorganic NPs are synthesized by ball milling, vapor deposition, electrospraying, reduction of metal salts, sol–gel, coprecipitation and thermal decomposition. Organic NPs are synthesized by microemulsion, nanoprecipitation, dialysis and rapid expansion of supercritical solutions. Hybrid NPs are synthesized from both organic and inorganic materials. There are a number of naturally occurring biological NPs including lipoproteins, exosomes, ferritin, and viruses. Further, NPs can be synthesized from biomolecules including proteins, peptides and polysaccharides. The surface to volume ratio, superparamagnetism, hardness, Coulomb energy and catalytic activity of NPs are generally higher than those of bulk materials. Due to their unique structural, magnetic, mechanical and electrical properties, NPs are used in a wide range of applications including biosensing, drug delivery, bioimaging, catalysis, nanomanufacturing, lubrication, electronics, textile manufacturing, and water treatment systems. This chapter covers the classification, synthesis, properties and applications of NPs.

11.1 **Overview** 335
11.2 **NPs Classification and Synthesis** 336
11.2.1 NPs Classification 336
11.2.2 NPs Synthesis 336
11.3 **Properties of NPs** 348
11.3.1 Size of NPs 348
11.3.2 Shape of NPs 349
11.3.3 Magnetic Properties of NPs 349
11.3.4 Mechanical Properties of NPs 350
11.3.5 Electrical Properties of NPs 351
11.4 **Applications of NPs** 351
11.5 **Summary** 353
References 353

11.1 Overview

Nanoparticles (NPs) are solid colloidal particles with dimensions in the range of a few nanometers to a few hundred nanometers [11.1]. Due to their small size, NPs have significantly higher surface to volume ratios, hence different physicochemical properties compared with bulk materials. NPs have applications in a wide range of areas including [11.2]:

- Biosensing
- Drug delivery
- Bioimaging
- Catalysis
- Nanomanufacturing
- Lubrication
- Electronics
- Textile industry
- Water treatment.

In the first part of this chapter, various methods for the synthesis of inorganic, organic, hybrid, and biological NPs are described. The properties of NPs including size and morphology as well as their unique magnetic, mechanical and electrical properties are discussed next. The major applications of NPs are discussed in the last part of this chapter.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_11

11.2 NPs Classification and Synthesis

11.2.1 NPs Classification

Multiple approaches are used to classify nanoparticles into the following: two-dimensional (e.g., nanotubes) and three-dimensional (e.g., spherical NPs, dendrimers). NPs can be synthesized from several classes of materials including inorganic (e.g., metal oxide NPs), organic (e.g., polymeric NPs), hybrid and biological materials [11.3]. Hybrid NPs are comprised of an inorganic core with organic shell or an organic core with inorganic shell. NP synthesis falls into two major categories namely top-down and bottom-up synthetic approaches. In the top-down approach, NPs are synthesized from bulk materials using traditional solid-state techniques (e.g., milling, machining) or lithographic techniques. In the bottom-up approach, NPs are synthesized from single molecules via chemical reactions, self-assembly, aggregation or chemical vapor deposition [11.1].

11.2.2 NPs Synthesis

Inorganic NPs

Depending on the type of material and the desired NP size, different physical or chemical strategies can be used to synthesize inorganic NPs [11.1]. Here, a number of physical techniques (ball milling, vapor deposition and electrospraying) and chemical methods (reduction of metal salts, sol-gel, coprecipitation and thermal decomposition) are described for the synthesis of inorganic NPs. The microemulsion technique that is extensively used for the synthesis of both organic and inorganic NPs is described in the section on organic NPs.

The ball milling method is a physical top-down synthesis approach based on breaking of solid particulate materials into smaller particles using high-energy balls rotating within a closed chamber [11.1]. Metal and metal oxide NPs have been synthesized using the ball milling method [11.4, 5]. The size and physicochemical characteristics of NPs synthesized via ball milling depend on milling time, rotation speed, ball to powder weight ratio as well as temperature and pressure [11.1]. For example, magnetite (Fe_3O_4) NPs with sizes ranging from 12 to 20 nm were synthesized via ball milling of iron powder in the presence of distilled water [11.5]. The size of Fe_3O_4 NPs decreased and the purity of NPs increased with increasing milling time from 10 to 96 h [11.5].

Vapor deposition is a physical approach for the synthesis of NPs based on vaporization of the material from a solid source using an ion (sputtering), electron (electron beam evaporation) or laser (laser ablation) beam followed by nucleation and growth of the vaporized material on a substrate (Fig. 11.1) [11.1]. Metal and metal oxide NPs have been synthesized via the vapor deposition method. For example, pure metallic (Fe, Ag, Au, Y, Co), nonmetallic (C) and metal oxide (TiO_2) NPs were synthesized via sputtering-based vapor deposition on a direct current (DC) magnetron cathode [11.6]. The size and physicochemical characteristics of NPs synthesized by vapor deposition are affected by [11.6]:

- Target to substrate distance
- Type of substrate
- Substrate temperature
- Deposition time
- Chamber pressure
- Electrical power.

In the laser-based vapor deposition method, the number of pulses, deposition time, gas pressure and substrate temperature significantly affects the size and morphology of NPs [11.7, 8]. For instance, the average size of FeCo NPs synthesized by pulsed laser vapor deposition increased from 4.8 to 8.6 nm as the argon

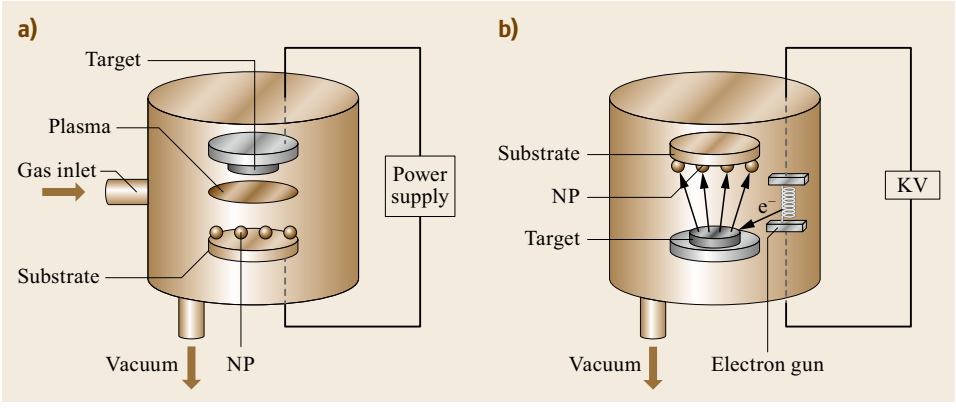


Fig. 11.1a,b Schematic representation of the physical vapor deposition method using **(a)** ion and **(b)** electron beams

gas pressure inside the chamber increased from 10 to 90 kPa [11.7]. Further, agglomeration of the FeCo NPs was reduced by increasing the temperature gradient [11.7].

In the electrospraying method, nanodroplets of a solution of the selected material are sprayed from a syringe to a collector within a high electric field (Fig. 11.2) [11.9]. Metal, metal oxide and polymer NPs have been synthesized via electrospraying [11.9–11]. A number of solvents including toluene, ethanol, methanol and isopropyl alcohol have been used for electrospraying of inorganic NPs [11.1]. For example, Sn NPs with an average size of 3 nm were synthesized by electrospraying a solution of $SnCl_2 \cdot 2H_2O$ in isopropyl alcohol into a reductive bath [11.10]. Monodisperse gold NPs with 5.7 nm diameter were synthesized by electrospraying $HAuCl_4 \cdot H_2O$ solution in isopropanol into a bath containing the reducing agent octadecylaminomethanol and the nonpolar solvent cyclohexane [11.11].

The most common chemical method for the synthesis of metallic NPs is reduction of metal ions in a metal salt solution using a reducing agent. Sodium borohydride ($NaBH_4$) and hydrazine (N_2H_4) are commonly used as reducing agents for the reduction of metal ions [11.12]. For example, the ferric ions in the solution of $FeCl_3 \cdot 6H_2O$ salt were reduced using sodium borohydride to synthesize iron NPs [11.13]. The size and morphology of NPs synthesized by the reduction method depends on temperature and pH of the reaction medium as well as the reactant concentration [11.14, 15].

The sol-gel technique has been used extensively for the synthesis of metal oxide NPs. The sol-gel method, can be divided into three major steps [11.12, 16]. In the first step (sol formation), the metal alkoxide or the metal salt is hydrolyzed in water/ethanol solution to form metal hydroxide (11.1). The metal hydroxide then undergoes condensation to create metal–O–metal bridges (11.2). For example, the hydrolysis and condensation reactions of tetraethylorthosilicate ($Si(OC_2H_5)_4$) in the synthesis of silica NPs are as follows

$$Si(OC_2H_5)_4 + 4(H_2O) \xrightarrow{\text{hydrolysis}} Si(OH)_4 + 4C_2H_5OH \quad (11.1)$$

$$Si(OH)_4 + Si(OH)_4 \xrightarrow{\text{condensation}} (OH)_3Si{-}O{-}Si(OH)_3 + H_2O \quad (11.2)$$

Colloidal particles or sol is formed when sufficient metal–O–metal bridges are created locally. Usually a mineral base (e.g., NH_3) or a mineral acid (e.g., HCl) is used as a catalyst. In the second step, colloidal particles link together with time to create a three-dimensional network at the nanoscale. In the third step, the nanoparticles are dried by heating, freeze drying or supercritical drying [11.16]. After drying and dehydration, NPs are densified by heat treatment at high temperatures [11.16].

Coprecipitation is another approach for the synthesis of inorganic and hybrid NPs [11.12, 17]. In the coprecipitation method, two or more water soluble salts react in aqueous solution to form a water insoluble salt that precipitates and forms NPs [11.12]. For example, hydrated ferrous chloride ($FeCl_2 \cdot 4H_2O$) and ferric chloride ($FeCl_3 \cdot 6H_2O$) undergo oxidative reaction following deprotonation in aqueous solution in the presence of $NH_3 \cdot H_2O$ as the precipitating agent [11.12]. The overall reaction can be written as

$$2Fe^{3+} + Fe^{2+} + 8OH^- \rightarrow Fe_3O_4 + 4H_2O\,. \quad (11.3)$$

When the concentration of Fe_3O_4 product in the solution increases above the solubility limit, the product precipitates by nucleation in the liquid phase. The growth of Fe_3O_4 crystals continues with precipitation reaction on the nucleating surfaces. Finally, the small particles undergo agglomeration via Ostwald ripening and form Fe_3O_4 NPs [11.12].

The thermal decomposition technique is used to synthesize metal oxide NPs from metal salts [11.12, 17]. In this method, a transition metal (e.g., Fe, Zn, Zr, and Co) salt is stabilized in an organic solvent/surfactant solution and mixed with an organic compound to form an organometallic complex. The organometallic complex is decomposed at high temperatures to create metal oxide NPs. A wide range of surfactants including oleylamine [11.18], oleic acid [11.19], benzylamine [11.20], and pentanedione [11.21] have been used to stabilize metal salts in the thermal decomposition process.

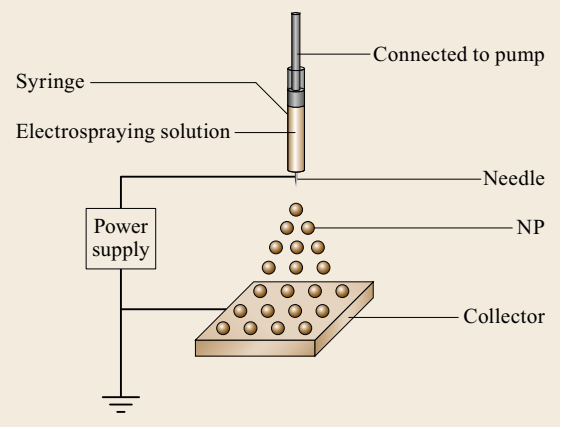


Fig. 11.2 Schematic representation of electrospraying method for the synthesis of NPs

Organic NPs

Several synthetic approaches for the synthesis of organic NPs are used. These include microemulsion, nanoprecipitation, dialysis and rapid expansion of a supercritical solution. The microemulsion method, used extensively for the synthesis of organic NPs, involves two major steps [11.22]. In the first step (emulsification), the organic compound (e.g., the polymer) is dissolved in a selective solvent and the polymer solution (dispersed phase) is emulsified in a second solvent (continuous phase) that is fully or partially immiscible in the dispersed phase solvent (Fig. 11.3). Nanodroplets of the disperse phase with diameters ranging from 10–500 nm are formed in the emulsification step [11.22]. There are two methods for the emulsification of the dispersed phase namely high-energy and low-energy methods. High-energy methods usually involve sonication or high-shear mixing [11.23]. The low-energy methods are divided into two groups: microemulsification based on spontaneous diffusion and microemulsification based on phase inversion stimulated by a change in temperature or composition [11.23, 24]. In the spontaneous emulsification of oil nanodroplets in water, the oily phase is composed of a partially immiscible solvent and a water soluble solvent, usually acetone or ethanol. When the oily and aqueous solutions are mixed, the water soluble solvent rapidly diffuses from the oily phase to the aqueous phase. This rapid diffusion between the two phases produces surface tension gradients that lead to the formation of a microemulsion [11.25].

The phase inversion techniques for microemulsification are classified into two groups: The so-called phase inversion temperature (PIT) method is centered on the hydrophilic-lipophilic transition of polyethylene glycol (PEG)-based surfactants at high temperatures as a result of disruption of the hydrogen bonding [11.24]. As shown in Fig. 11.4, an oil (dispersed phase)/water (continuous phase) emulsion at constant composition inverts to a water/oil emulsion when temperature rises above the inversion temperature [11.26]. The affinity of surfactants for oil and water phase is similar at the inversion temperature leading to a minimal interfacial tension and bicontinuous microemulsion in the transition region. Therefore, kinetically stable nanodroplets can be formed by rapid cooling or dilution of the aforementioned emulsion systems at invasion temperature [11.24] (Fig. 11.4). The second phase inversion technique for microemulsification, the so-called emulsion inversion point (EIP) method, employs emulsion composition change for phase transition. The water/oil emulsion is formed by successive addition of water to an oil solution. The emulsion undergoes phase transition from water/oil to oil/water with increasing volume fraction of water. A bicontinuous microemulsion with minimal interfacial tension and nanosized droplets is formed at the phase inversion point [11.26].

Following emulsification, the NPs are formed in the second step of the microemulsion technique from the emulsified nanodroplets via solvent removal, gelation, or emulsion polymerization (Fig. 11.3a). Different approaches including solvent evaporation and solvent displacement methods have been employed for solvent removal and precipitation of the NPs in

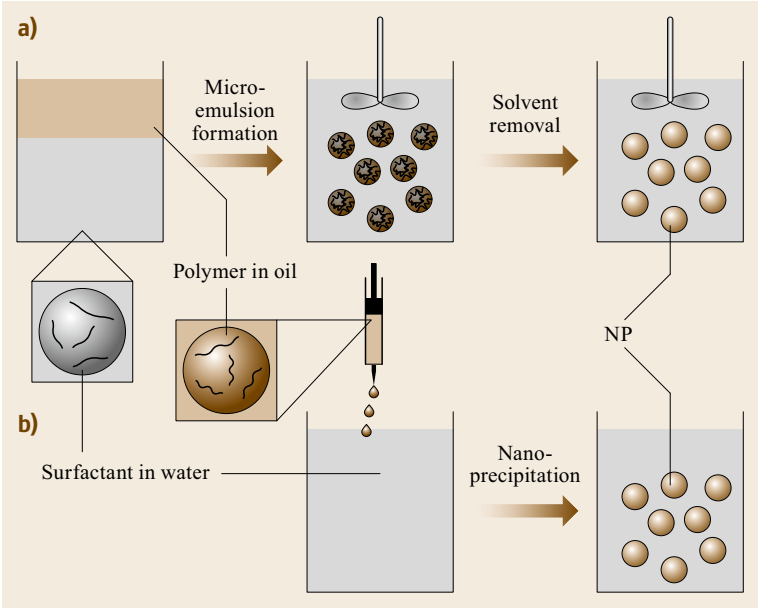


Fig. 11.3a,b Schematic representation of **(a)** microemulsion and **(b)** nanoprecipitation techniques for the synthesis of organic NPs

microemulsion systems [11.24]. In the solvent evaporation method, a volatile solvent (e.g., chloroform or dichloromethane) is used as the solvent in the disperse phase. The volatile solvent is evaporated under vacuum leaving the polymer in the nanodroplets. During solvent evaporation, the size of the nanodroplets initially decreases dramatically due to solvent loss and then increases significantly due to droplet coalescence [11.24]. Organic NPs from numerous polymers and copolymers including poly(lactide) (PLA), poly(lactide-co-glycolide) (PLGA), poly(ε-caprolactone) (PCL), and PEG-co-PLA were synthesized via the solvent evaporation method [11.27–29]. In the solvent displacement method, a solvent with partial miscibility in water is used in the disperse phase. NPs are formed by solvent displacement from the dispersed phase to the continuous aqueous phase upon successive addition of water or dialysis [11.24].

The gelation of macromolecules in the nanodroplets is another approach for the synthesis of NPs in microemulsion systems. Several gelation mechanisms depending on physicochemical properties of the macromolecules within the nanodroplets are employed for NP formation. These include temperature-sensitive physical gelation, pH-sensitive gelation and chemical crosslinking [11.30, 31]. For example, nanogels based on polyethylene glycol (PEG) macromers chain-extended with short lactide (L) and glycolide (G) segments were synthesized via dialysis method and they were used for grafting proteins (Fig. 11.5a) [11.32]. The size of synthesized nanogels ranged from 118 to 190 nm depending on the molecular weight of PEG and L/G molar ratio [11.32]. Further, the rate of BSA protein release from the nanogels was between 51% release after 24 days and complete release in 11 days depending on the molecular weight of PEG and L/G molar ratio (Fig. 11.5b) [11.32].

Polymerization of monomers in nanodroplets is used extensively for the synthesis of NPs from microemulsions. In the first step, nanodroplets of monomers with low water solubility are formed in the aqueous solution containing a surfactant and a water soluble initiator (Fig. 11.6). Free radical or ionic initiators are used depending on the type of polymerization reaction. The collision of initiator molecules and monomer molecules starts the polymerization reaction within the nanodroplets. Solid NPs are formed with the progress of polymerization reaction and increase in the molecular weight of the polymer molecules in the nanodroplets. NPs of different polymers including polystyrene (PS), poly(methyl methacrylate) (PMMA) and poly(vinylcarbazole) were synthesized using emulsion polymerization [11.2, 33, 34]. Further, surfactant-free emulsion polymerization has been employed for the synthesis of NPs based on acrylic or vinyl monomers [11.24]. For example, PMMA, PS and poly(vinyl acetate) (PVA)-based NPs with diameter ranging from 155 to 345 nm were synthesized using surfactant-free emulsion polymerization and sodium persulfate as the initiator [11.35]. The surfactant removal, an indisputable step in conventional emulsion polymerization, is absent in the surfactant-free emulsion polymerization technique, which is a significant advantage in large-scale manufacturing of organic NPs.

In addition to the conventional and surfactant-free emulsion polymerization, interfacial polymerization has been used for the synthesis of polymeric nanocap-

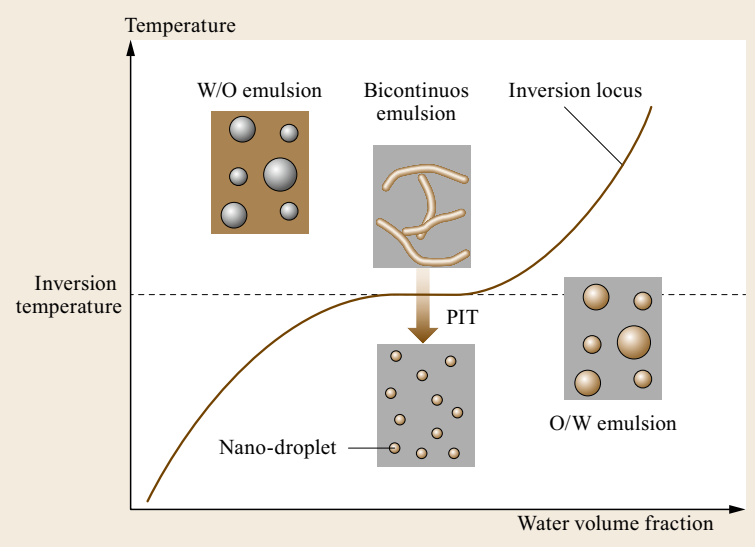


Fig. 11.4 Schematic representation of the phase inversion temperature (PIT) method for microemulsification. Nanodroplets are formed with rapid cooling of the bicontinuous emulsion. Reprinted with permission from [11.26]

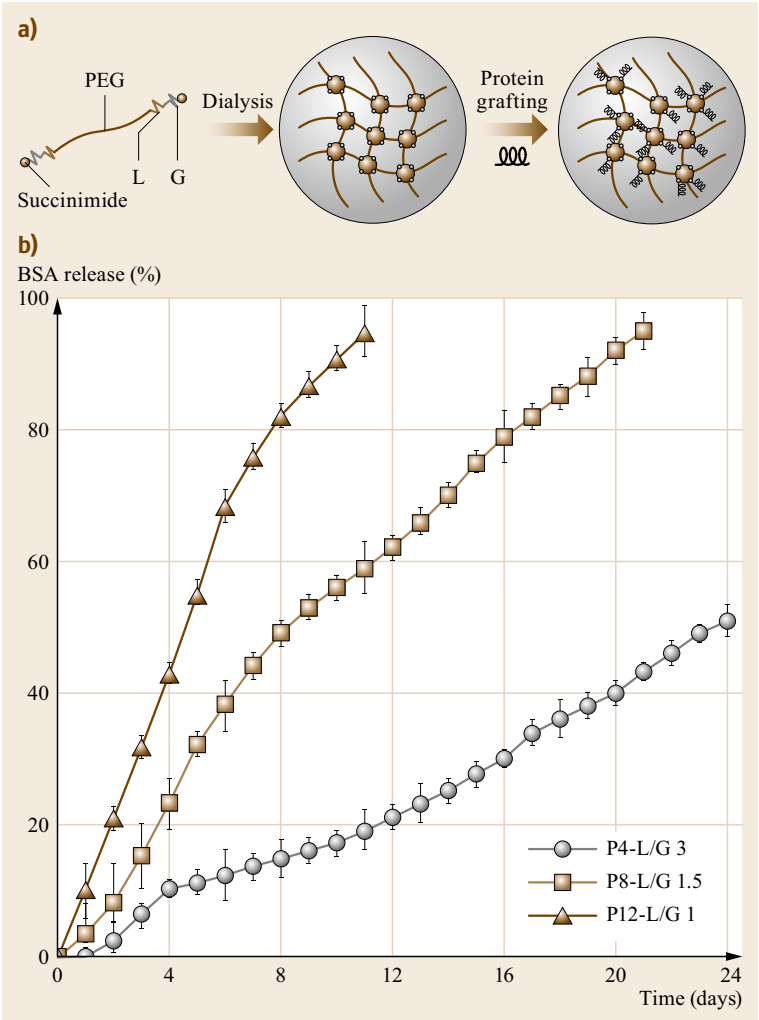


Fig. 11.5 **(a)** Schematic diagram for the assembly of nanogels from PEG macromers chain-extended with short lactide (*L*) and glycolide (*G*) segments and protein grafting to nanogels. **(b)** The cumulative BSA release from nanogels in PBS with incubation time. The abbreviations *P*4, *P*8, and *P*12 represent nanogels made from macromers with PEG $M_W = 4, 8$ and 12 kDa. The abbreviations $L/G\,3$, $L/G\,1.5$ and $L/G\,1$ represent lactide/glycolide molar ratio of 3, 1.5 and 1, respectively. Reprinted with permission from [11.32]

sules (Fig. 11.6) [11.36]. In interfacial polymerization, a monomer is dissolved in the emulsified nanodroplets and a second monomer (comonomer) is dissolved in the continuous phase. The two monomers undergo polymerization at the oil/water interface on the surface of the nanodroplets to form nanocapsules [11.36]. For example, poly(alkylcyanoacrylate)-based nanocapsules are synthesized using the interfacial polymerization method [11.37]. The alkylcyanoacrylate monomer dissolved in the oil nanodroplets reacts with hydroxyl ions in water and undergo interfacial polymerization at the oil/water interface [11.37]. Polybutylcyanoacrylate nanocapsules prepared via the method of interfacial polymerization had diameters of less than 150 nm, and the nanocapsules were used for encapsulation of the D-cycloserine drug [11.38]. Poly(ethyl 2-cyanoacrylate) nanocapsules synthesized using interfacial polymerization had ≈ 150 nm size and the nanocapsules were used for encapsulation of insulin [11.39]. Nanocapsules encapsulating hydrophilic drugs were synthesized by inverse emulsion (water/oil) using a thiol-isocyanate interfacial polymerization [11.40].

Conventional radical polymerization suffers from poor control over the molecular weight and polydispersity of the polymer chains which significantly influences the uniformity of polymeric NPs synthesized using radical polymerization [11.36]. Therefore, living radical polymerization has recently been employed to synthesize NPs with narrow size distributions [11.41]. For example, styrene monomer underwent reversible addition and fragmentation transfer (RAFT) polymerization in poly(*N*-isopropylacrylamide-*b*-dimethylacrylamide) based emulsions to form polystyrene NPs [11.42]. The resulting polystyrene NPs had 38–110 nm diameter and < 0.13 polydispersity index (derived from the standard deviation

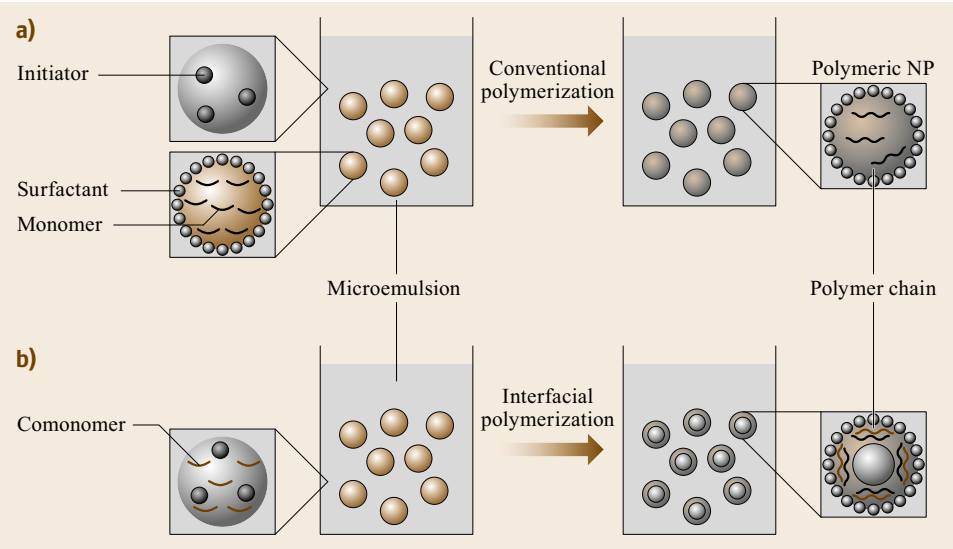


Fig. 11.6a,b Schematic representation of **(a)** conventional and **(b)** interfacial microemulsion polymerization

of the hypothetical Gaussian distribution) [11.42]. Nitroxide mediated polymerization (NMP), a well-known living polymerization technique, was employed to synthesize polystyrene, poly(*n*-butyl acrylate) or poly(methyl methacrylate) NPs with narrow size distributions [11.43–45]. Further, NPs based on poly(methyl methacrylate), poly(*n*-butyl methacrylate) and poly(ethylene oxide) were synthesized using the method of atom transfer radical polymerization (ATRP) [11.46–48].

In the nanoprecipitation method, the polymer is first dissolved in a water-miscible organic solvent. The polymer solution is then injected into a stirred aqueous solution containing a stabilizer (Fig. 11.3b). The spontaneous nanodrop formation followed by rapid diffusion of the organic solvent to the aqueous phase leads to polymer precipitation at the nanoscale and the formation of NPs stabilized by the stabilizer [11.36] (Fig. 11.3b). After NP formation, the organic solvent is removed by evaporation and the NPs are collected via centrifugation. The organic solvent in the nanoprecipitation technique should dissolve the polymer, be soluble in water, and have a low boiling point to facilitate its removal by evaporation. To satisfy the aforementioned criteria, acetone and tetrahydrofuran (THF) are widely used as solvents for nanoprecipitation [11.49, 50]. NPs of different polymers including PLA [11.50, 51], PLGA [11.49, 52], PCL [11.53, 54] and gelatin [11.55] were synthesized using the nanoprecipitation technique.

The dialysis method for the synthesis of NPs is based on solvent displacement similar to the nanoprecipitation method with the difference that there is a membrane barrier between the two phases in the dialysis method. In a typical procedure for the dialysis method, the polymer is dissolved in a water-miscible organic solvent, the organic solution is placed in a dialysis tube and dialyzed against distilled water [11.32]. The molecular weight cut off of the dialysis tube is usually smaller than the molecular weight of the polymer molecules but significantly larger than the solvents molecular weight. Therefore, the solvent diffuses out and water diffuses inside the dialysis tube while polymer molecules stay in the tube during dialysis. The solvent quality for the polymer in the dialysis tube decreases with time which causes polymer aggregation and precipitation at the nanoscale [11.24]. The formed NPs are stabilized by either a stabilizer or a hydrophilic block on the polymer molecule.

In the method of rapid expansion of a supercritical solution, the polymer is dissolved in a supercritical fluid (e.g., carbon dioxide) at high pressure and then the solution is expanded through an orifice or a nozzle into ambient air or a liquid. NPs are formed in the expansion jet due to a rapid pressure reduction [11.36]. The expansion of the supercritical solution into a liquid leads to the formation of uniform nanosized particles as opposed to expansion into air wherein both nanosized and microsized particles form [11.36]. For example, tetrahydrocurcumin (THC) loaded PLA NPs with 80–110 nm size and spherical morphology were synthesized by rapid expansion of supercritical solution of PLA and THC into water [11.56]. Further, NPs made of pure hydrophobic drugs including Naproxen and Ibuprofen have been synthesized using the rapid expansion of the supercritical solution technique [11.57, 58].

Organic–Inorganic Hybrid NPs

Hybrid NPs are composed of both organic and inorganic materials. The inorganic part may form the core or the shell of the hybrid NPs. Here, the methods for synthesis of hybrid NPs with inorganic core are discussed first followed by the discussion for the synthesis of hybrid NPs with organic core.

Hybrid NPs with Inorganic Core. Polymers can be grafted to the surface of inorganic NPs through covalent or noncovalent bonds (Fig. 11.7). Non-covalent attachment usually employs the electrostatic interaction between a charged polymer and an oppositely charged inorganic NP (Fig. 11.7a) [11.59]. For example, positively charged poly(allylamine hydrochloride) polymer was deposited on the negatively charged citrate-coated gold NPs using a layer-by-layer approach [11.60]. In another example, positively charged Fe_3O_4 NPs with 5 nm diameter were coated with negatively charged poly(acrylic acid)-*b*-poly(ethylene oxide) polymer. The coated NPs interacted with oppositely charged pristine NPs and formed aggregates with hydrodynamic diameter ranging from 50 to 100 nm [11.61].

The functional groups present on the surface of inorganic NPs (e.g., hydroxyl groups on the surface of metal oxide NPs) can be used to conjugate polymer chains to NPs (Fig. 11.7b) [11.62]. For example, polymethacrylic acid (PMAA) was conjugated to the surface of ZnO NPs through the reaction of −OH groups on the NP surface with carboxyl groups of PMAA [11.63]. In order to improve dispersion properties of silica NPs, poly(ethylene glycol) methacrylate (PEGMA) or poly(propylene glycol) methacrylate (PPGMA) were conjugated on the surface of vinyl-functionalized silica NPs via UV photo-polymerization [11.63]. Polystyrene was conjugated to TiO_2 NPs via a three-step process [11.64]. First, the surface of TiO_2 NPs was functionalized with phosphonic acid. Next, the surface of phosphonic acid modified NPs was functionalized with azide groups. Then, polystyrene was conjugated to azide functionalized NPs via the click reaction [11.64]. In another study, amine-terminated silica-coated Fe_3O_4 NPs were synthesized via sol-gel method and functionalized with alkyne groups using Michael addition of propargyl acrylate [11.65]. Then, poly(ester amine) was grafted to the alkyne-functionalized NPs using the click reaction [11.65].

In addition to grafting of preformed polymers to inorganic NPs, in situ polymerization of monomers on the surface of NPs (grafting from the surface) has been used extensively for the synthesis of organic–inorganic hybrid NPs (Fig. 11.7c) [11.66]. For example, densely grafted polystyrene and PMMA chains were grown on the surface of silica NPs by incorporation of atom transfer radical polymerization (ATRP) initiator on the surface of NPs [11.67]. A technique similar to ATRP was

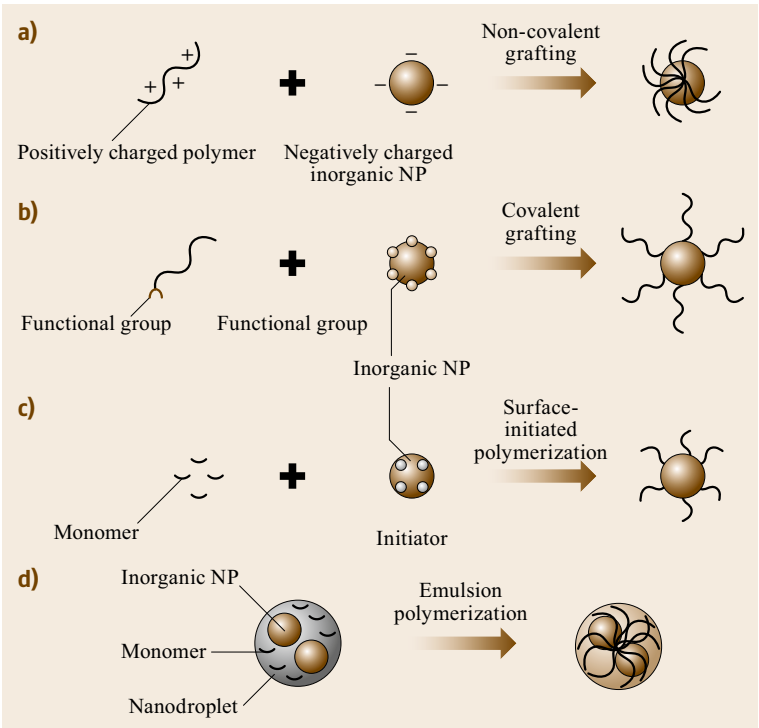


Fig. 11.7a–d Schematic representation of methods for the synthesis of hybrid NPs with inorganic core. **(a)** Non-covalent grafting of polymer chains to inorganic NPs via electrostatic complexation. **(b)** Covalent grafting of polymer chains to functionalized inorganic NPs. **(c)** Surface-initiated polymerization of monomers on inorganic NPs. **(d)** Emulsion polymerization of monomers in the presence of inorganic NPs

recently applied for the synthesis of PMMA-Al_2O_3 hybrid NPs [11.68]. In an attempt to synthesize thermoresponsive NPs, poly(*N*-isopropylacrylamide) (PNIPAm) chains were grown on the surface of Fe_3O_4 NPs using surface-initiated reversible addition-fragmentation transfer (RAFT) reaction [11.69]. The hydrodynamic diameter of the thermoresponsive hybrid NPs decreased from 37 to 24 nm when temperature increased from 28 to 36 °C [11.69]. PMMA-Fe_3O_4 hybrid NPs were synthesized using a thiol-lactam initiated radical polymerization technique [11.66].

Hybrid NPs can also be synthesized via emulsion polymerization of monomers in the presence of inorganic NPs (Fig. 11.7d). The surface of inorganic NPs is typically modified to increase compatibility with the polymer matrix [11.70]. For example, silica NPs with 20 or 78 nm diameter and hydrophobically-modified with γ-methacryloxypropyl trimethoxysilane (γ-MPS) were encapsulated in poly(methyl methacrylate-co-butyl acrylate) polymer NPs via emulsion polymerization [11.71]. Untreated silica NPs aggregated in the miniemulsion system and did not uniformly distribute in the poly(methyl methacrylate-co-butyl acrylate) NPs [11.71]. Polystyrene-silica hybrid NPs were synthesized via a miniemulsion polymerization technique using octaphenyl polyoxyethylene as a surfactant [11.72]. The hybrid polystyrene-silica NPs had a raspberry-like morphology with 130–370 nm diameter and 17–31% inorganic silica content [11.72]. Hybrid NPs with an inorganic TiO_2 core and organic polyacrylic shell were synthesized via emulsion polymerization of methyl methacrylate-butyl acrylate (inner shell) and dimethylaminoethyl methacrylate-butyl acrylate-acrylic acid (outer shell) in the presence of TiO_2 NPs modified with glycidyl methacrylate [11.73]. The TiO_2-polyacrylic hybrid NPs had 85–125 nm diameter [11.73]. Polystyrene nodules were formed on the surface of silica NPs via emulsion polymerization of styrene in the presence of silica NPs modified with an oxyethylene-based macromonomer [11.74]. The morphology of silica-polystyrene hybrid NPs depended on the ratio between silica seeds and growing polystyrene nodules and changed from snowman-like to daisy-like and raspberry-like morphology when increasing the ratio [11.74].

Hybrid NPs with Organic Core. Preformed inorganic NPs can be positioned on the surface of organic NPs using surface treatment followed by aggregation of inorganic–organic NPs (hetero-coagulation) (Fig. 11.8a) [11.70]. For example, negatively charged inorganic gold or silica NPs were deposited on the surface of polystyrene NPs functionalized with positively charged amine groups [11.75]. The position of inorganic NPs within the polymeric NPs can be adjusted by proper surface treatment. For example, gold NPs modified with mercaptoundecanol and dodecanethiol ligands were positioned in the shell of polystyrene-*b*-poly(acrylic acid) NPs whereas gold NPs modified with only hydrophobic dodecanethiol were distributed

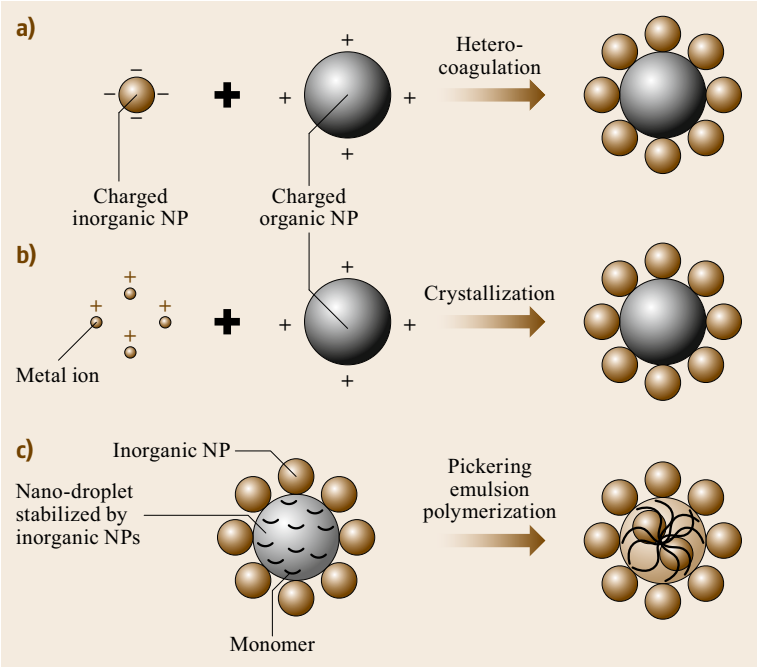


Fig. 11.8a–c Schematic representation of methods for the synthesis of hybrid NPs with organic core. **(a)** Hetero-coagulation of preformed inorganic and organic NPs. **(b)** Crystallization of inorganic NPs on the surface of organic NPs. **(c)** Pickering emulsion polymerization of monomers in the nanodroplets stabilized by inorganic NPs

in the core of the block-copolymer NPs [11.76]. Further, inorganic NPs can be covalently attached to the surface of organic NPs by conjugation. For instance, in an attempt to prepare fluorescent thermosensitive hybrid NPs, organic PNIPAM NPs were covered with inorganic CdTe nanocrystals via an amide coupling agent [11.77]. The photoluminescence (PL) intensity of the hybrid PNIPAM-CdTe NPs was temperature dependent [11.77].

Another approach for the synthesis of hybrid NPs is crystallization of inorganic materials on the surface of polymeric NPs (Fig. 11.8b) [11.70]. In this approach, the surface of polymeric particles should be charged to induce nucleation of the inorganic metals or metal oxides on the polymeric surface.

Several techniques including noncovalent assembly of a polyelectrolyte, covalent conjugation of a polyelectrolyte, and incorporation of a functional comonomer have been used to modify the surface of organic NPs for the synthesis of hybrid NPs [11.70]. For example, gold and platinum NPs (2−3 nm) were formed on the surface of polystyrene NPs (≈ 100 nm) covered with cationic 2-aminoethylmethacrylate polyelectrolytes [11.78]. In another example, styrene and a surface active comonomer (surfmer) with negatively charged phosphate groups underwent miniemulsion polymerization to form polystyrene NPs [11.79]. Negatively charged polystyrene NPs were then used for controlled crystallization of cerium, iron, and zinc oxide nanocrystals in aqueous or alcohol medium [11.79]. In addition, poly(styrene-co-acrylic acid) NPs surface-functionalized with negatively charged carboxyl groups were prepared via the miniemulsion technique and then covered with inorganic hydroxyapatite nanocrystals [11.80]. Calcium phosphate (CaP) NPs were deposited on the surface of PLA and PLGA nanofibers functionalized with negatively charged Glu-Glu-Gly-Gly-Cys peptide in a modified simulated body fluid (mSBF) (Fig. 11.9) [11.81, 82]. The deposition of CaP NPs increased and the size of NPs decreased when a carboxylate-rich organic acid such as citric acid, hydroxycitric acid, tartaric acid, malic acid, ascorbic acid, or salicylic acid was added to mSBF [11.81].

Hybrid NPs can also be synthesized via particle-stabilized emulsion polymerization (Pickering emulsion polymerization) technique (Fig. 11.8c). In Pickering emulsion polymerization, the surfactants are replaced with particle stabilizers [11.83]. Inorganic NPs used in the Pickering emulsion polymerization need to be wettable by both phases in the emulsion, be dispersible in the continuous phase and be able to accumulate at the interface of two phases [11.83]. The most widely used inorganic NPs as Pickering stabilizers are silica NPs [11.83]. For example, hybrid

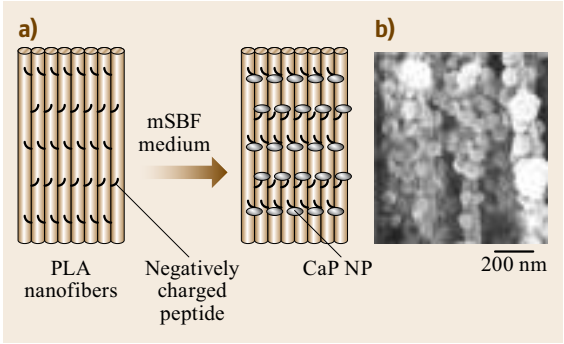


Fig. 11.9 **(a)** Schematic representation of the synthesis of CaP NPs on the surface of peptide functionalized PLA nanofibers. **(b)** SEM image of CaP NPs crystallized on the surface of peptide functionalized PLA nanofibers. Reproduced with permission from [11.81]

NPs were synthesized via Pickering emulsion polymerization of styrene, methyl methacrylate, *n*-butyl acrylate, or *n*-butyl methacrylate as a monomer and 4-vinylpyridine as a comonomer in the presence of silica NPs [11.84]. It was shown that a strong acid-base interaction between the copolymer and silica NPs is essential for the formation of hybrid NPs and hybrid NPs did not form in the absence of 4-vinylpyridine comonomer [11.84]. Silica-polystyrene hybrid NPs were synthesized via Pickering emulsion polymerization of styrene using a cationic azobisisobutyronitrile initiator (AIBN), glycerol functionalized silica and without using any comonomer [11.85]. The hybrid NPs had 200−400 nm diameter and 22−28 wt% silica [11.85]. In addition to silica, zinc oxide, titania, magnetite and graphene oxide NPs have been used as inorganic Pickering stabilizers in the synthesis of hybrid NPs [11.83]. For instance, ZnO-polystyrene hybrid NPs were prepared via emulsion polymerization of styrene in the presence of ZnO particles using AIBN as the initiator. The polystyrene formed the core while ZnO particles were positioned in the shell of the hybrid NPs [11.86].

Biological NPs

Biological NPs can be classified into two categories: Those that are naturally formed in biological systems from organic or inorganic materials (naturally occurring NPs) and those that are synthesized from biomolecules (e.g., proteins, peptides and polysaccharides).

Naturally Occurring NPs. Naturally occurring NPs are assembled in biological systems and include extracellular NPs (e.g., lipoproteins and viruses) and intracellular NPs (e.g., ferritin). Lipoproteins are self-assembled naturally occurring NPs made of lipids

and specialized proteins. Lipoproteins have a micellar structure with a core composed of triglycerides and cholesterol esters and a shell of phospholipids and amphipathic proteins [11.87, 88]. Lipoproteins are synthesized mainly by the liver and intestines and can be separated from plasma via ultracentrifugation [11.87, 88]. There are several classes of lipoproteins depending on size and density. The size of lipoproteins increases from 7−13 nm for high density lipoprotein (HDL) to 22−27 nm for low density lipoprotein (LDL), 27−30 nm for intermediate density lipoprotein (IDL), 35−80 nm for very low density lipoprotein (VLDL) and 80−1200 nm for chylomicrons [11.88]. Hydrophobic core of the lipoproteins can be loaded with water insoluble drugs for therapeutic applications. For example, valrubicin, an anticancer drug, was encapsulated in spherical HDL NPs [11.89]. The toxicity of valrubicin drug against ovarian and prostate cancer cell lines increased while its toxicity against normal prostate and ovarian epithelial cell lines decreased when the drug was encapsulated in HDL NPs compared with the drug alone [11.89].

Exosomes are 30−120 nm vesicles released by B cells, dendritic cells, mast cells, mesenchymal stem cells, adipocytes, neurons, endothelial cells and cancer cells in vivo [11.87, 90]. Further, exosomes are secreted by most cells in vitro [11.87]. In addition, exosomes can be isolated from extracellular fluids including blood, urine, saliva, and amniotic fluid [11.90]. The outer surface of exosomes is a lipid bilayer made of cholesterol, sphingolipids, ceramide and phosphoglycerides with saturated fatty-acyl chains [11.87]. The functions of exosomes in vivo include encapsulation and removal of excess membrane proteins and nucleotides from cells, encapsulation and delivery of signaling proteins and regulation of immune response through antigen presentation [11.91–93]. Proteins and nucleotides encapsulated in exosomes are protected from degradation and endocytosis by macrophages hence having a long circulation time [11.87]. Therefore, exosomes have been studied for the delivery of proteins and RNA in cancer treatment and tissue regeneration [11.90, 94].

Ferritin is an intracellular hollow nanoparticle with an outer diameter of 12 nm and interior cavity diameter of 8 nm [11.87]. Ferritin protein shell is composed of 24 subunits made up of the heavy-chain ferritin and the light-chain ferritin [11.87]. The interior cavity is naturally used to store iron oxides and to sequester the undesirable effects of iron ions. Ferritin NPs have been used as nanoreactors to synthesize ferrimagnetic iron oxide NPs in a constrained volume [11.95]. After the formation of inorganic core in the ferritin cavity, the protein shell can be removed to separate the inorganic NPs. In addition to iron oxide, inorganic cobalt oxide, chromium oxide and nickel hydroxide NPs have been synthesized in the cavity of ferritin NPs [11.96]. Further, semiconductor NPs including CdSe, ZnSe, CdS and ZnS NPs have been synthesized using ferritin [11.96]. It is shown that ferritin-iron oxide hybrid NPs bind to tumor cells that overexpress transferrin receptor 1 (TfR1) and undergo a color reaction that can be used to visualize tumor tissues [11.97].

Viruses are naturally occurring NPs (20−300 nm) with capsid proteins forming the shell and viral DNA or RNA forming the core [11.87]. Viruses have a wide range of sizes and morphologies including spherical, rod-shape, and icosahedral morphologies [11.87]. Virus-like particles with a native protein shell but without nucleic acid do not cause infection and can be used as carriers for drugs or proteins. Further, bioactive ligands can be conjugated to lysine or cysteine residues on the capsid protein [11.98]. For example, cell penetrating peptides conjugated to Cowpea mosaic virus (CPMV) increased the uptake efficiency of the virus by human cervical cancer cells [11.99]. In another example, conjugation of nucleoprotein of influenza to Bacteriophage P22 virus-like particles resulted in a vaccine that induced a protective $CD8^+$ T cell response against lethal doses of influenza while the virus-like particles alone did not provide immunity [11.100].

Protein NPs. Several techniques including desolvation, emulsion–solvent extraction and polyelectrolyte complexation have been employed for the synthesis of NPs from proteins [11.101]. The desolvation method is based on drop-wise addition of a desolvation agent such as alcohols, salts or other solvents to a protein solution with stirring to trigger precipitation of protein NPs [11.102]. Since the solubility and conformation of proteins in aqueous solution depends on pH, solvent polarity, and ionic strength of the solution, a change in these factors leads to conformational changes and nanoprecipitation of proteins [11.101]. The precipitation step is often followed by crosslinking of NPs with glutaraldehyde in order to stabilize the particles [11.102]. NPs of different proteins including albumin, gelatin and elastin have been synthesized via the desolvation method [11.101]. For example, human serum albumin (HSA) NPs were prepared via the slow addition of ethanol to HSA solution at room temperature followed by glutaraldehyde crosslinking [11.103]. The size of HSA NPs varied between 200−300 nm depending on pH of the desolvation solution [11.103]. The size of the albumin NPs prepared via desolvation decreased with increasing pH and increasing the ratio of antisolvent (e.g., ethanol) to solvent [11.101, 103]. Bovine serum albumin (BSA) NPs synthesized via desolvation and surface modified with cationic polyethylenimine were

employed for the encapsulation of bone morphogenic protein-2 (BMP-2) protein [11.104]. The NPs exhibited a sustained release of BMP-2 over 10 days and high retention of protein activity in vitro [11.104]. In another example, gelatin NPs were synthesized by slow addition of ethanol to the aqueous solution of gelatin with 65 : 35 ethanol/water ratio at 37 °C and pH = 7.0 followed by particle crosslinking with glyoxal [11.105]. The gelatin NPs had a 17 h half-life and 30 h mean residence time in the plasma of Lewis lung carcinoma-bearing mice [11.105].

The emulsion–solvent extraction method for the synthesis of protein NPs is based on emulsification of the aqueous solution of the protein in oil using a high-speed homogenizer or a sonicator in the presence of a surfactant such as phosphatidylcholine or Span 80 [11.101]. The oil phase is then removed using an organic solvent and NPs are separated via centrifugation [11.101]. NPs can be stabilized using a physical or chemical crosslinking technique. For example, NPs based on casein, a milk protein, were synthesized via oil-in-water emulsification in the presence of Span 80 [11.106]. Casein NPs were stabilized via ionic crosslinking using a polyanionic crosslinker sodium tripolyphosphate [11.106]. Casein NPs had 60−100 nm diameter and exhibited a sustained release of anticancer drug flutamide for four days in vitro [11.106]. NPs based on albumin protein were also synthesized using the emulsion–solvent extraction method [11.101, 107].

Polyelectrolyte complexation is typically used for the synthesis of DNA-laden protein NPs [11.101]. In this method, proteins are first charged cationic by adjusting the pH to below their isoelectric point (pI) (Fig. 11.10). Cationic proteins then form a polyelectrolyte complex with negatively charged DNA that undergo desolvation to form DNA-laden NPs [11.101]. For example, HSA-DNA-polyethylenimine (PEI) NPs were synthesized in a two-step process [11.108]. First, a solution of PEI and positively charged HSA was prepared at pH 4.1. Then, NPs were formed by adding sodium sulfate solution containing DNA to the HSA-PEI solution [11.108]. Positively charged poly(lysine) and chitosan were also used for the synthesis of NPs via the polyelectrolyte complexation method [11.24].

Peptide NPs. Several classes of peptides including surfactant-like peptides and polypeptides have been used for the synthesis of NPs [11.109, 110]. The surfactant-like peptides have a hydrophilic head and a hydrophobic tail similar to an organic surfactant. The peptides self-assemble in aqueous solution to form micelles with different morphologies including nanotubes, nanovesicles and nanofibers [11.111]. For instance, surfactant-like peptides with a tail of six hydrophobic amino acids and a head of two positively charged amino acids (e.g., V_6K_2, V=Valine, K=Lysine) self-assembled to nanotubes and spherical nanovesicles in aqueous solution [11.112]. The V_6K_2 peptide conjugated to a low molecular weight PLA molecule (PLA-V6K2) was self-assembled to spherical NPs with an average size of 100 nm (Fig. 11.11a) [11.113]. The PLA-V6K2 NPs had sustained release of an anticancer drug Doxorubicin (Dox) over 35 days (Fig. 11.11b) [11.113]. Further, 49% of the PLA-V6K2 NPs were taken up by 4T1 mouse breast carcinoma cells after 24 h (Fig. 11.11c) [11.113]. The release rate of Dox was slower and the percent uptake by 4T1 cells was higher for PLA-V6K2 NPs compared with PEG stabilized PLA (PLA-EG) NPs (Fig. 11.11b,c) [11.113].

When the hydrophobic tail of a surfactant-like peptide is an alkyl chain, the structure is called peptide amphiphile (PA). It is shown that PAs spontaneously self-assemble into a range of nanostructures including nanofibers, spherical micelles, nanovesicles and nanotubes [11.114, 115]. For example, a PA with the structure of C_{16}-KKFFVLK (F= Phenylalanine, L= Leucine) self-assembled into nanotubes and helical ribbons at room temperature [11.116].

A number of polypeptides have been used for synthesis of NPs [11.117]. For example, an elastin-like polypeptide (ELP) with a repeating unit of VPAVG (P=Proline, A=Alanine, G=Glycine) self-assembled to NPs when the temperature of polypeptide aqueous solution was raised from 4 to 37 °C [11.118]. The elastin-like polypeptide NPs exhibited a sustained release of encapsulated dexamethasone phosphate for 30 days in vitro at 37 °C [11.118]. In a recent example, NPs with 600 nm diameter were synthesized from an ELP conjugated to stromal cell-derived growth factor-1 (SDF1) for wound healing applications [11.119]. The wounds of diabetic mice treated with ELP-SDF1 NPs were closed after 28 days while those wounds treated with SDF1 alone were closed after 42 days [11.119]. pH sensitive polypeptide nanovesicles were synthesized from zwitterionic diblock poly(L-glutamic acid)-*b*-poly(L-lysine) copolymers. The hydrodynamic radius of the nanovesicles was 110 and 175 nm in acidic (pH < 4) and basic (pH > 10) aqueous solutions while the polypeptides did not self-assemble to nanovesicles at intermediate pH values (5 < pH < 9) [11.120].

Polysaccharide NPs. Polysaccharides are naturally derived polymers with monosaccharide repeat units joined together by glycosidic bonds [11.110]. Positively charged chitosan and negatively charged alginate, heparin, hyaluronic acid, dextran sulfate and der-

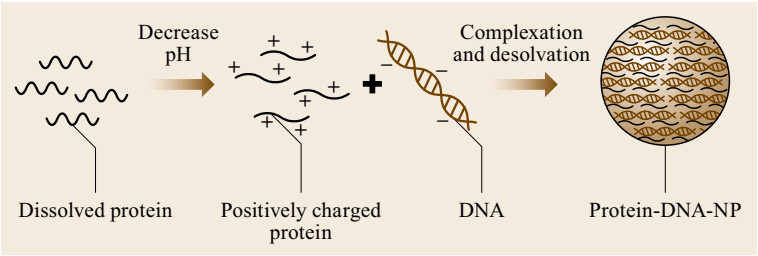


Fig. 11.10 Schematic representation of the polyelectrolyte complexation method for synthesis of protein–DNA NPs

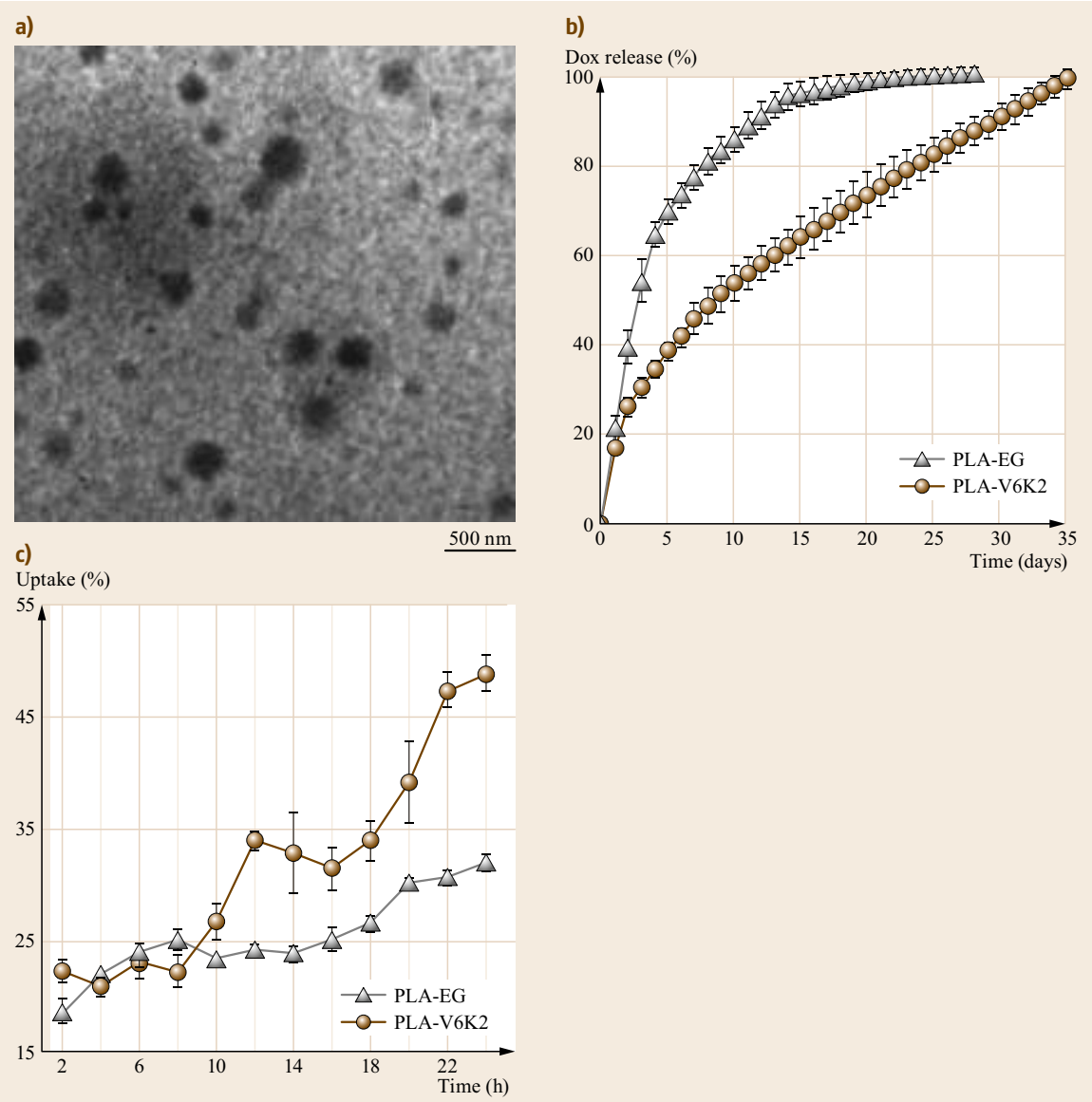


Fig. 11.11 **(a)** TEM image of the uranyl acetate stained PLA-V6K2 NPs. **(b)** Release profile of Dox encapsulated in PLA-V6K2 and PLA-EG NPs incubated in PBS. **(c)** The effect of incubation time on percent uptake of PLA-V6K2 and PLA-EG NPs by 4T1 breast tumor cells. Reprinted with permission from [11.113]

matan sulfate are the most widely used polysaccharides for synthesis of NPs [11.110]. Different approaches including polyelectrolyte complexation, crosslinking, and self-assembly have been employed for the synthesis of polysaccharide NPs [11.110]. For example, chitosan-alginate NPs were prepared by polyelectrolyte complexation between cationic chitosan and anionic alginate pregel particles [11.121]. Polysaccharide NPs prepared by polyelectrolyte complexation between dermatan and chitosan had 730 nm radius and were selectively taken up by H5V murine endothelial cells in the presence of Raw 264.7 murine macrophage cells [11.122]. Chitosan NPs (250 nm diameter) synthesized via water-in-oil emulsion method, stabilized by glutaraldehyde crosslinking, and coated with anionic polysaccharides were used for encapsulation of the anticancer drug 5-fluorouracil (5-FU) [11.123]. Self-assembling amphiphilic polysaccharides generated by hydrophobic modification of polysaccharides' backbone have also been used for the synthesis of NPs and micelles [11.110]. For instance, hydrophobically modified chitosan prepared via grafting *N*-acyl groups to the backbone of chitosan self-assembled to NPs with < 550 nm diameter at 38 °C and pH 7.4 [11.124]. In another example, polyglycolic acid (PGA) conjugated alginate self-assembled to NPs with 200–250 nm diameter through hydrophobic interactions between the PGA segments [11.125].

11.3 Properties of NPs

11.3.1 Size of NPs

A change in NP size significantly affects the surface area as well as mechanical, magnetic, electrical and biological properties of NPs. The surface to volume ratio increases when the particle size decreases. Therefore, the fraction of surface atoms or functional groups increases with decreasing the NP size. For example, the chemical reactivity of silica NPs increases with decreasing the particle size due to an increase in the density of silanol groups (Si–OH) on the NPs surface [11.126]. The superparamagnetism of magnetic NPs increases as the size decreases [11.127]. The elastic modulus and hardness of crystalline metal NPs is inversely correlated with particle size [11.95]. The capacitance of conductive NPs decreases and Coulomb energy increases with reducing the NP size [11.128]. Anticancer drug-loaded NPs with diameters < 100 nm can pass through the leaky vasculature of the tumor tissues [11.129]. The size of NPs can be controlled by adjusting the synthesis parameters including reactant–surfactant concentration, temperature and pH of the medium [11.12].

In the ball milling method, the size of synthesized NPs depends on milling time, rotation speed, ball-to-powder weight ratio as well as temperature and pressure [11.1]. The size of NPs synthesized via the vapor deposition method depends on target to substrate distance, type of substrate, substrate temperature, deposition time, chamber pressure and the electrical power for deposition [11.6]. In the laser-based vapor deposition method, number of pulses, deposition time, gas pressure and substrate temperature significantly affects the size and morphology of NPs [11.7, 8] The size of NPs synthesized via reduction and precipitation methods can be controlled by changing pH of the synthesis medium. The adsorption of proton or hydroxide ions on the surface of growing particles decreases the surface energy of growing NPs, hence increases the stability of small particles [11.130]. Therefore, adjusting pH of the synthesis medium to values that increase the surface charge density of growing particles decreases the size of NPs [11.130]. For instance, in the synthesis of TiO_2 NPs via precipitation method, the surface charge density increased from 0 to 0.2 and the size of NPs decreased from 9 to 5 nm by decreasing pH of the synthesis medium from 5.3 to 2 [11.14].

In the microemulsion method, NPs of smaller size can be synthesized by reducing water to surfactant ratio. An increase in the water to surfactant ratio, increases the droplet size as well as the aggregation of particles, hence increases size of the synthesized NPs. For example, Pd NPs were synthesized via the microemulsion technique in a water/isooctane medium using sodium bis(2-ethylhexyl) sulfosuccinate as the surfactant and N_2H_4 as the reducing agent [11.131]. The size of Pd NPs after 10 min of reaction increased from 8.1 to 11.7 nm with increasing the water to surfactant ratio from 3 to 7 at reducing agent to metal molar ratio of 60 [11.131]. In another example, a higher concentration of PVA stabilizer in the single-emulsion synthesis of PLGA polymer NPs decreased the size of particles significantly [11.132]. Further, the size of NPs in the microemulsion technique can be reduced by using a more viscous oil phase [11.133]. In addition, high-energy mixing (e.g., sonication) of microemulsion systems decreases the size of synthesized NPs [11.12]. An increase in the concentration of the reactants in the microemulsion technique decreased the size of NPs which was due to a transition from intermicellar nucleation to intramicellar nucleation [11.134, 135].

The size of polymer NPs prepared via dialysis or nanoprecipitation depends on the polymer composition and concentration and the solvent used. The size of nanoprecipitated NPs decreases with increasing solvent–nonsolvent affinity as well as polymer-solvent affinity [11.136]. For example, the size of PLGA NPs decreased with changing the solvent from dimethylacetamide (DMAc) to dimethylsulfoxide (DMSO) and dimethylformamide (DMF) in the dialysis method [11.137]. In another example, the size of poly(lactide-*co*-glycolide fumarate) (PLGF) NPs synthesized by the dialysis method was smaller than poly(lactide fumarate) (PLAF) NPs due to the higher hydrophilicity of the PLGF macromer [11.138].

11.3.2 Shape of NPs

The shape of NPs significantly affects their physicochemical and biological properties. For example, the catalytic activity of Pt NPs is shape dependent [11.139] or the circulation time and cell uptake of drug-loaded NPs varies with the NPs' morphology [11.140]. NPs generally favor a morphology that is associated with minimal surface free energy [11.15]. For amorphous materials with an isotropic surface, a spherical morphology minimizes the surface free energy by lowering the surface to volume ratio [11.15]. In an anisotropic crystalline structure, each crystal plane has a specific surface free energy [11.130]. For example, in noble metals (e.g., Ag, Au and Pt), the surface free energy of (100), (110) and (111) planes is lower than those of high-index crystal planes. Therefore, the standard morphology of NPs made of single crystalline noble metals is a truncated octahedron that is a linear combination of the aforementioned low-index planes [11.15]. However, the shape of crystalline metal or metal oxide NPs synthesized via solution-based methods can be modified by adjusting the nucleation and growth steps. For example, the shape uniformity of NPs increases by increasing the rate of crystal nucleation in the salt reduction method [11.15]. Further, adsorbates that selectively adsorb to specific crystal planes in the growth step can be employed to control the crystal morphology [11.15]. The crystal growth continues on the plane that has weak interaction with the adsorbate. Surfactants, poly(vinyl pyrrolidone) and metal ions have been used as adsorbates in the synthesis of crystalline NPs [11.15]. For instance, single crystalline Ag NPs with several morphologies including cube, octahedral, truncated cube, truncated octahedra and cuboctahedra were synthesized via reduction of $AgNO_3$ in 1,5-pentanediol in the presence of poly(vinyl pyrrolidone) as an adsorbate at 180 °C [11.141]. The morphology of Ag NPs depended on the rate of injection of $AgNO_3$ solution in the reducing solvent [11.141]. The morphology of inorganic NPs can also be controlled by adjusting the reaction parameters such as temperature, pH and relative concentration of metal ions. For example, manganese dioxide (MnO_2) NPs with a number of morphologies including needle-like, lamellar and spherical were synthesized by varying the Mn(VII) to Mn(II) ratio and pH of the reducing medium [11.130].

Due to the semicrystalline nature of polymers, the favorable morphology of polymeric NPs is spherical [11.142]. Polymeric NPs with nonspherical morphologies have been synthesized via top-down and bottom-up approaches. For example, several morphologies of polystyrene NPs including ellipsoids, rods and disks were synthesized by embedding the spherical polystyrene NPs in a poly(vinyl alcohol) film and stretching the film [11.143]. Further, several morphologies of polyethylene glycol (PEG) based NPs including spheres, cylinders, cubes and disks were synthesized via spraying or drop-casting the polymer precursor solution on a perfluoropolyether nonwetting template [11.144]. Polymer and peptide based NPs with a range of morphologies have been synthesized via the self-assembly method. For example, worm-like micelles were synthesized via self-assembly of poly(ethylene glycol)-poly(ethylethylene) diblock copolymers [11.145]. Surfactant-like peptides with a tail of six hydrophobic amino acids and a head of two positively charged amino acids (e.g., V_6K_2) self-assembled to nanotubes and spherical nanovesicles in aqueous solution [11.112]. Furthermore, it was shown that peptide amphiphiles spontaneously self-assemble into a variety of nano-objects including spherical micelles, nanovesicles and nanotubes [11.114, 115].

11.3.3 Magnetic Properties of NPs

Magnetic NPs have been synthesized with several compositions including iron oxides, pure metals, alloys (e.g., FePt) and spinel-type compounds (e.g., $MnFe_2O_4$ or $CoFe_2O_4$) [11.146]. Further, a number of approaches including coprecipitation, microemulsion, thermal decomposition and hydrothermal synthesis have been applied for the synthesis of magnetic NPs [11.146]. The majority of magnetic NPs are classified into either ferromagnetic or superparamagnetic materials [11.127]. In ferromagnetic materials, atomic magnetic dipoles are aligned and there exists a net magnetic moment in the absence of an external magnetic field. Paramagnetic materials have magnetic dipoles but these dipoles are not aligned and fluctuate randomly, hence the net magnetic moment is zero in the absence of an external magnetic field above a material-specific temperature called the Curie temperature. Ferromagnetic or fer-

rimagnetic NPs with sizes below a material-specific threshold have zero net magnetic moment in the absence of an external field and they become magnetic and their magnetic moment rapidly increases with the application of an external magnetic field (Fig. 11.12). This property of magnetic NPs which is called superparamagnetism occurs at temperatures above a blocking temperature (T_B) wherein thermal fluctuations exceed the energy barrier for random flipping of magnetic moments within NPs [11.127].

The T_B of NPs correlates with particle volume and rapidly increases with particle diameter [11.148]. The maximum on the magnetization curve (Fig. 11.12) is called saturation magnetization (M_s) which happens when all magnetic dipoles are aligned with the external magnetic field [11.127, 147]. The remaining magnetization after the removal of an external magnetic field (the intercept of y-axis and the magnetization curve in Fig. 11.12) is called remanent magnetization (M_r) [11.147]. The strength of an external field required to force the magnetization to zero (the intercept of x-axis and the magnetization curve in Fig. 11.12) is called coercivity (H_c) [11.147]. The M_r and H_c are both zero for paramagnetic and superparamagnetic materials as opposed to ferromagnetic materials (Fig. 11.12) [11.147]. The magnetization of superparamagnetic materials increases rapidly upon applying an external field as opposed to the linear trend of paramagnetic materials (Fig. 11.12) [11.147]. The M_s increases linearly with NP size and reaches a plateau corresponding to the bulk M_s [11.127]. Therefore, smaller NPs have lower M_s but higher superparamagnetism. For example the M_s of $Fe_{2.66}O_4$ NPs decreased from 77 to 29 emu/g and T_B decreased from 100 to 10 K when the NP size was reduced from 14 to 2.5 nm [11.149]. In addition to size, the morphology of NPs affects magnetization. For example, the T_B of Fe_3O_4 NPs decreased when the NPs morphology was changed from spherical to cubic [11.150]. Further, the M_s of Fe_3O_4 NPs increased from 18 to 40 and 80 emu/g when the NPs morphology was changed from rods to cubes and spheres, respectively [11.107]. The composition and method of synthesis of NPs also significantly affects magnetic properties of NPs. For instance, the size at which a transition to superparamagnetism happens decreased when the composition of NPs was changed from Fe_2O_3 to Fe_3O_4 and FePt [11.151]. Alloyed magnetic NPs generally have higher M_s than the corresponding pure metals or iron oxide based NPs [11.127]. In another example, metal-doped NPs with the composition of MFe_2O_4 (M = Mn, Fe, Co or Ni) were synthesized via a nonhydrolytic reaction between iron tris-2,4-pentadionate and divalent metal chloride (MCl_2). The M_s of MFe_2O_4 NPs with the same size of 12 nm varied from 85 to 99, 101 and 110 emu/g by changing the metal (M) from Ni to Co, Fe and Mn, respectively [11.152]. The highest M_s of $MnFe_2O_4$ was related to a mixed spinel structure as opposed to an inverse structure for $NiFe_2O_4$, $CoFe_2O_4$ and $FeFe_2O_4$ NPs [11.152]. Particle coating generally decreases the magnetization of NPs [11.127]. For example, M_s of Fe_3O_4 NPs decreased from 75 to 51 emu/g and coercivity increased from 1.1 to 5 when the NPs were coated with PNIPAM polymer [11.153].

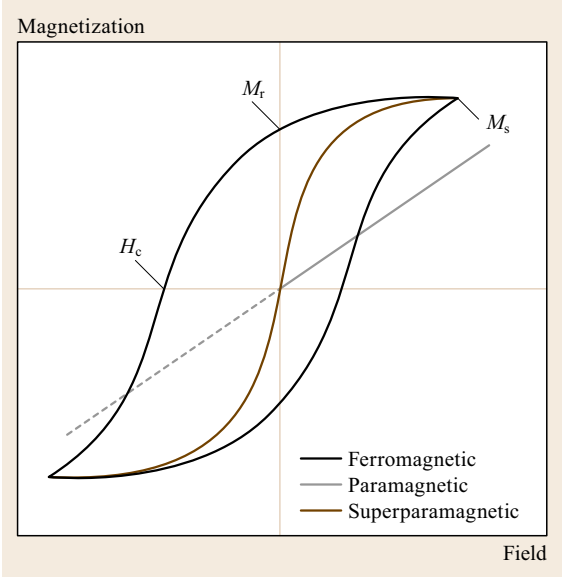


Fig. 11.12 Hysteresis loops of ferromagnetic, paramagnetic, and superparamagnetic materials (after [11.147])

11.3.4 Mechanical Properties of NPs

The mechanical properties of NPs including elastic modulus, hardness, and frication are significantly different from those of bulk materials [11.154]. Results of several studies have shown that the hardness, yield strength, and elastic modulus of crystalline metal NPs are higher than those of bulk metals and inversely correlated with particle size [11.155–158]. For example, the mechanical properties of gold NPs with a sixfold icosahedral symmetry and 22 nm size were measured via the nano-indentation method [11.157]. The hardness and elastic modulus of gold NPs were 1.72 and 100 GPa, respectively, which were significantly higher than the hardness (216 MPa) and elastic modulus (79 GPa) of gold [11.157]. The increase in yield strength or hardness with decreasing size for single crystalline NPs is due to the absence of dislocations in the crystalline structure leading to higher resistance to plastic deformation [11.154]. For polycrystalline NPs, the enhanced mechanical properties are due to an increase in the den-

sity and entanglement of dislocations with decreasing the grain size that limits the motion of dislocations (Hall–Petch effect) [11.159]. However, the hardness and yield strength of crystalline NPs decrease significantly if the NPs are smaller than a critical size due to the inverse Hall–Petch effect [11.154]. In this case, the grain is too small to support a high density of dislocations hence the deformation is dominated by grain boundary sliding [11.154]. For example, the modulus of nanocrystalline titania NPs increased with reducing the particle size down to 15 nm [11.160]. The modulus of titania NPs then decreased with decreasing particle size below 15 nm [11.160]. The mechanical properties of polymeric NPs depend on the glass transition temperature, crystallinity of the polymer as well as the surface properties of NPs [11.161]. For instance, the compressive modulus of surface-charged polystyrene NPs was slightly less than that of bulk polystyrene but the elastic modulus of polypropylene NPs was higher than that of bulk polypropylene [11.162, 163].

The friction coefficient of nano-objects is proportional to the contact area as opposed to the macroscopic dry sliding friction. Results of a number of studies showed that the friction of NPs increases with particle size [11.154, 161]. For example, an increase in the size of gold NPs from 30 to 90 nm, increased the dry friction of particles under 300 nN load by threefold [11.164]. Further, gold nanorods with 50 nm diameter and 200 nm length had a higher friction in air or water than gold NPs with 50 nm diameter [11.154]. The friction force between single polystyrene NPs (30–230 nm) and a silicon substrate was correlated with two thirds power of the particle radius [11.165]. Further, the frication of nano-objects increased logarithmically with the sliding velocity as opposed to the macroscopic frication which was independent of sliding velocity [11.166].

11.3.5 Electrical Properties of NPs

The semicontinuous electronic band structure of bulk metals changes to discrete electronic levels when decreasing the size in the nanoscale. This phenomenon is called the size quantization effect, which is responsible for the unique electrical properties of metallic and semiconductive NPs. A transition from a conductive bulk material to semiconductive NP happens due to the size quantization effect and formation of a band gap if the size of NP is sufficiently low. Metallic NPs with a diameter lower than 2 nm start to undergo transition from conductor to semiconductor [11.167]. Further, the band gap that normally exists in semiconductive materials grows with decreasing size from bulk to NPs. However, the quantization effect in semiconductor materials happens at relatively larger scales (1–10 nm) [11.167]. The electrical current in bulk materials obeys Ohm's law

$$V = IR\,, \tag{11.4}$$

where V is the applied voltage, I is current and R is resistance of the material. Ohm's law is not valid when quantization happens at the nanoscale and the Coulomb energy of NPs exceeds the thermal energy

$$k_{\mathrm{B}}T < \frac{e^2}{2C}\,, \tag{11.5}$$

where k_{B} is the Boltzmann constant, T is temperature, e is elementary charge and C is capacitance. The capacitance of a nanoparticle depends on its size as [11.128]

$$C = 4\pi r\varepsilon\varepsilon_0\,, \tag{11.6}$$

where r is the NP radius, ε and ε_0 are dielectric constant of the medium and vacuum, respectively. Therefore, capacitance decreases and Coulomb energy increases with reducing NPs size. When the NPs are sufficiently small such that Coulomb energy exceeds thermal energy (11.5), charge transfer occurs through the NPs via a single electron tunneling. In this case, the external voltage $\Delta V = e/2C$ is required to transfer a single charge to the NP leading to a tunneling current $\Delta I = e/RC$. Therefore, the current–voltage curve in quantum dots is like a staircase with height of ΔI and width of ΔV, due to tunneling of multiple single electrons [11.128]. For example, Au NPs with 5.8 nm diameter at 77 °K had a staircase current–voltage curve with a width of $\Delta V =$ 200 mV [11.168].

11.4 Applications of NPs

NPs have been used for sensing biological analytes including DNA and proteins as well as small molecules such as glucose [11.11]. In biosensing, specific receptors conjugated to the surface of NPs interact with specific biological analytes and the biological recognition event is translated into detectable optical, magnetic or electrochemical signals [11.169, 170]. The increased sensitivity and lower detection limit of NP-based biosensors compared with macroscale biosensors are due to high surface to volume ratio of NPs resulting in a larger density of biospecific receptors per unit volume of NPs [11.11]. Further, unique magnetic,

optical and electrical properties of NPs contribute to the translation of recognition event to detectable signals [11.11]. For example, Au NPs were used for colorimetric sensing of oligonucleotides [11.169]. Au NPs functionalized with single-stranded DNA aggregated in the presence of oligonucleotides leading to a change in the color of solution [11.169]. In another example, interleukin-6 (IL-6) protein in the unprocessed human sera was detected using IL-6 antibody conjugated magnetic FeCo NPs and a giant magnetoresistive (GMR) sensor functionalized with IL-6 antibody [11.171]. The IL-6 analyte and antibody-conjugated magnetic NPs were captured and formed a sandwich on the surface of the GMR sensor [11.171]. The change in electrical resistance of the GMR sensor was correlated to the amount of captured magnetic FeCo NPs, which was related to the concentration of IL-6 protein [11.171].

One of the major applications of NPs is the localized or targeted delivery of drugs and biomolecules to cells and tissues [11.172]. NPs as delivery vehicles protect drugs or biomolecules from degradation, transport them into the cells or tissues and release them in a sustained manner [11.169]. The uptake efficiency of bioactive molecules by cells is significantly higher for NPs compared with microparticles. For example, PLGA NPs (100 nm) had 15–250-fold higher gastrointestinal tissue uptake in a rat model than microparticles (10 μm) [11.173]. Anticancer-drug-loaded NPs with diameters < 100 nm can pass the leaky vasculature of tumors and enter the interstitial space of tumor tissue [11.129]. Several biodegradable NPs based on PLA, PLGA, PCL and polyacrylate polymers as well as natural alginate, chitosan and gelatin have been synthesized for the delivery of drugs or proteins [11.172]. The drug or protein may be entrapped within the NPs or conjugated on the surface of NPs. Further, targeting antibodies can be attached to the surface of NPs for recognition and binding to target cells [11.174]. The release rate of encapsulated drugs or proteins from biodegradable polymer NPs can be tuned from a few days to months by altering the composition or length of the polymer chains [11.172]. For example, anticancer drug Paclitaxel encapsulated in the more hydrophobic NPs based on poly(lactide-fumarate) (PLAF) displayed a sustained release of Paclitaxel over 28 days whereas the less hydrophobic poly(lactide-co-glycolide fumarate) (PLGF) NPs released the encapsulated drug in 15 days [11.175]. The viability of HCT116 human colon carcinoma cells significantly decreased when they were exposed to Paclitaxel encapsulated in PLGF or PLAF NPs compared with the drug alone [11.175]. In a recent example, recombinant human bone morphogenetic protein-2 (BMP2) and vascular endothelial growth factor (VEGF) were grafted to nanogels based on polyethylene glycol (PEG) macromers chain-extended with short lactide and glycolide segments [11.32]. The release rate of BMP2 and VEGF proteins from nanogels depended on PEG molecular weight, lactide/glycolide segment length and lactide–glycolide molar ratio in the polymer [11.32]. Nanogels with 12 kDa PEG molecular weight, 24 lactide-glycolide segment length, and 60 : 40 lactide/glycolide ratio released the grafted VEGF in 10 days and induced vasculogenic differentiation of human endothelial colony-forming cells while nanogels with 8 kDa PEG molecular weight, 26 lactide-glycolide segment length, and 60 : 40 lactide–glycolide ratio released the grafted BMP2 in 21 days and induced osteogenic differentiation of human mesenchymal stem cells [11.32].

Several types of NPs including noble metal NPs, quantum dots, polymer NPs and fluorescently doped silica NPs have been used for imaging of cells and tissues [11.176]. For example, self-illuminating quantum dots were synthesized by conjugation of a modified bioluminescent protein *Renilla reniformis* luciferase to fluorescent semiconductor CdSe/ZnS quantum dots [11.177]. The conjugate emitted long-wavelength bioluminescent light without a need for external excitation of quantum dots [11.177]. Further, magnetic NPs have been used as contrast agents in magnetic resonance imaging (MRI) [11.178]. MRI signal intensity is related to the spin–lattice relaxation time (T_1) and spin–spin relaxation time (T_2) of water protons [11.178]. When magnetic NPs accumulate in the tissue that is being imaged, they reduce the relaxation time which in turn enhance the image contrast [11.127]. The contrast enhancement is correlated with the M_s of the magnetic NPs [11.127]. For example, Fe_3O_4 NPs shorten the T_2 relaxation time and have been used as a contrast agent in imaging liver, spleen and bone marrow [11.179].

One of the earliest commercial applications of NPs was in catalyzed reactions such as the synthesis of reformulated gasoline and hydrosilylation reactions [11.180]. Metal NPs have been used as catalysts in quasihomogeneous catalyzed reactions including hydrogenation, cross-coupling, oxidation and electron transfer in organic or aqueous solutions [11.180]. For instance, hydrogenation of cyclooctene to cyclooctane was performed via quasihomogeneous catalysis using chitosan-stabilized Pt and Pd NPs (1.9–2.2 nm in size) as catalysts in acetic acid/methanol solution [11.181]. Further, metal NPs deposited on carbon, silica, and metal oxide supports have been used as catalysts in heterogeneous catalysis [11.180]. Several techniques including physical adsorption, grafting, flame spraying, ion exchange and the Langmuir–Blodgett method

have been employed for deposition of NPs on supports for heterogeneous catalysis [11.182–184]. The size, shape and composition of NPs significantly affect the catalytic properties of supported metal catalysts [11.180]. For example, Pt NPs of different shapes (cubic and cuboctahedral) deposited on silicon wafer supports via Langmuir–Blodgett technique were used for hydrogenation of benzene [11.183]. Selectivity of the heterogeneously catalyzed reaction depended significantly on the shape of Pt NPs. While cyclohexane and cyclohexene were formed on cuboctahedral NPs, only cyclohexane was formed on cubic NPs [11.183].

Metal, metal oxide, silicon and carbon-based NPs have been used as additives to lubricants in order to reduce the friction and wear between contacting surfaces [11.161]. The reduction of friction and wear in the presence of NPs is related to the rolling of the spherical NPs or sliding of the nonspherical NPs between tribopair surfaces [11.161]. For example, the friction and wear of PAO-6 polyalphaolefin lubricant was reduced by 20 and 50%, respectively, when 0.5% ZnO or ZrO_2 was added to the lubricant [11.185]. Abrasive NPs have been used for chemical and mechanical polishing (CMP) of silicon wafers in manufacturing of integrated circuits (IC) [11.161]. For instance, silicon wafers were polished by CeO_2-coated SiO_2 NPs (150−200 nm) with material removal rate of 454 nm/min [11.186].

Ceramic or carbon-based NPs have been added to polymer or metal matrices to make nanocomposites with improved mechanical properties [11.161]. For example, the hardness and tensile strength of an aluminum matrix increased by 115 and 67%, respectively, with the addition of 7 vol.% Al_2O_3 NPs (50 nm) to the matrix [11.187]. The tensile strength and tensile modulus of rubbery epoxy resin increased by fivefold with incorporation of 10 wt% silica NPs [11.188].

In addition of the aforementioned applications, NPs have applications in electronics, textile manufacturing, and water treatment. Semiconducting polymer NPs have been used in organic photovoltaics, field-effect transistors and light-emitting diodes [11.189]. Carbon-coated Sn NPs and Si NPs distributed on graphene sheets were used as anodes while $LiCoPO_4$ NPs were used as cathodes in lithium ion batteries [11.190–192]. Ag NPs due to their antimicrobial properties have been used to produce antimicrobial paints and coatings [11.193, 194]. ZnO and TiO_2 NPs due to their photocatalytic activity have applications in removal of chemical and biological contaminants from water [11.195].

11.5 Summary

NPs can be synthesized from several classes of materials including inorganic (e.g., metal oxide NPs), organic (e.g., polymeric NPs), hybrid and biological materials. Physical techniques (ball milling, vapor deposition and electrospraying) as well as chemical methods (reduction of metal salts, sol-gel, coprecipitation and thermal decomposition) are used for the synthesis of inorganic NPs. Organic NPs can be synthesized via microemulsion, nanoprecipitation, dialysis and rapid expansion of a supercritical solution. Hybrid NPs are composed of both organic and inorganic materials. The inorganic component can form the core or the shell of the hybrid NPs. In addition to the naturally occurring biological NPs, NPs can be synthesized from biomolecules including proteins, peptides and polysaccharides. The physico-mechanical properties of NPs are significantly different from those of bulk materials. For example, the catalytic activity, superparamagnetism, hardness and Coulomb energy of the NPs are higher than those of bulk materials. Due to their unique properties, NPs have applications in a wide range of areas including biosensing, drug delivery, bioimaging, catalysis, nanomanufacturing, lubrication, electronics, textile manufacturing, and water treatment.

Acknowledgments. This work was supported by research grants to E. Jabbari from the United States National Science Foundation under Award Numbers CBET1403545 and IIP150024 and the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number AR063745. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

11.1 C. Dhand, N. Dwivedi, X.J. Loh, A.N.J. Ying, N.K. Verma, R.W. Beuerman, R. Lakshminarayanan, S. Ramakrishna: Methods and strategies for the synthesis of diverse nanoparticles and their applications: A comprehensive overview, RSC Advances **5**, 105003–105037 (2015) doi:10.1039/c5ra19388e

11.2 M. Dossi, R. Ferrari, L. Dragoni, C. Martignoni, P. Gaetani, M. D'Incalci, M. Morbidelli, D. Moscatelli: Synthesis of fluorescent pmma-based nanoparticles, Macromol. Mater. Eng. **298**, 771–778 (2013) doi:10.1002/mame.201200122

11.3 S.L. Pal, U. Jana, P.K. Manna, G.P. Mohanta, R. Manavalan: Nanoparticle: An overview of preparation and characterization, J. Appl. Pharm. Sci. **1**, 228–234 (2011) doi:10.7897/2230-8407.04408

11.4 J.E. Munoz, J. Cervantes, R. Esparza, G. Rosas: Iron nanoparticles produced by high-energy ball milling, J. Nanopar. Res. **9**, 945–950 (2007) doi:10.1007/s11051-007-9226-6

11.5 J.F. de Carvalho, S.N. de Medeiros, M.A. Morales, A.L. Dantas, A.S. Carrico: Synthesis of magnetite nanoparticles by high energy ball milling, Appl. Surf. Sci. **275**, 84–87 (2013) doi:10.1016/j.apsusc.2013.01.118

11.6 V. Bouchat: N, Moreau, J.F. Colomer, S. Lucas: On some applications of nanoparticles synthesized in the gas phase by magnetron discharges, J. Surf. Eng. Mater. Adv. Technol. **3**, 184–189 (2013) doi:10.4236/jsemat.2013.33025

11.7 P.L. Ong, S. Mahmood, T. Zhang, J.J. Lin, R.V. Ramanujan, P. Lee, R.S. Rawat: Synthesis of FeCo nanoparticles by pulsed laser deposition in a diffusion cloud chamber, Appl. Surf. Sci. **254**, 1909–1914 (2008) doi:10.1016/j.apsusc.2007.07.186

11.8 Y.T. Jing, H.W. Wang, X. Chen, X.F. Wang, H.G. Wei, Z.H. Guo: Pulsed laser deposited Ag nanoparticles on nickel hydroxide nanosheet arrays for highly sensitive surface-enhanced Raman scattering spectroscopy, Appl. Surf. Sci. **316**, 66–71 (2014) doi:10.1016/j.apsusc.2014.07.169

11.9 A. Jaworek, A.T. Sobczyk: Electrospraying route to nanotechnology: An overview, J. Electrost. **66**, 197–219 (2008) doi:10.1016/j.elstat.2007.10.001

11.10 M. Valvo, U. Lafont, D. Munao, E.M. Kelder: Electrospraying-assisted synthesis of tin nanoparticles for Li-ion battery electrodes, J. Power Sources **189**, 297–302 (2009) doi:10.1016/j.jpowsour.2008.09.019

11.11 K. Soliwoda, M. Rosowski, E. Tomaszewska, B. Tkacz-Szczesna, G. Celichowski, M. Psarski, J. Grobelny: Synthesis of monodisperse gold nanoparticles via electrospray-assisted chemical reduction method in cyclohexane, Colloids Surf. A **482**, 148–153 (2015) doi:10.1016/j.colsurfa.2015.04.040

11.12 R.G. Chaudhuri, S. Paria: Core/shell nanoparticles: Classes, properties, synthesis mechanisms, characterization, and applications, Chem. Rev. **112**, 2373–2433 (2012) doi:10.1021/cr100449n

11.13 X.Q. Li, W.X. Zhang: Iron nanoparticles: The core-shell structure and unique properties for Ni(II) sequestration, Langmuir **22**, 4638–4642 (2006) doi:10.1021/la060057k

11.14 J.P. Jolivet, C. Froidefond, A. Pottier, C. Chaneac, S. Cassaignon, E. Tronc, P. Euzen: Size tailoring of oxide nanoparticles by precipitation in aqueous medium. A semi-quantitative modelling, J. Mater. Chem. **14**, 3281–3288 (2004) doi:10.1039/b407086k

11.15 A.R. Tao, S. Habas, P.D. Yang: Shape control of colloidal metal nanocrystals, Small **4**, 310–325 (2008) doi:10.1002/smll.200701295

11.16 I.A. Rahman, V. Padavettan: Synthesis of silica nanoparticles by sol-gel: Size-dependent properties, surface modification, and applications in silica-polymer nanocomposites – A review, J. Nanomater. **24**, 015105 (2012) doi:10.1088/0957-4484/24/1/015105

11.17 J. Gautier, E. Allard-Vannier, K. Herve-Aubert, M. Souce, I. Chourpa: Design strategies of hybrid metallic nanoparticles for theragnostic applications, Nanotechnology **24**(43), 432002 (2013) doi:10.1088/0957-4484/24/43/432002

11.18 M. Salavati-Niasari, M. Dadkhah, F. Davar: Synthesis and characterization of pure cubic zirconium oxide nanocrystals by decomposition of bis-aqua, tris-acetylacetonato zirconium(IV) nitrate as new precursor complex, Inorganica Chimica Acta **362**, 3969–3974 (2009) doi:10.1016/j.ica.2009.05.036

11.19 S.Y. Lee, M.T. Harris: Surface modification of magnetic nanoparticles capped by oleic acids: Characterization and colloidal stability in polar solvents, J. Colloid and Interface Sci. **293**, 401–408 (2006) doi:10.1016/j.jcis.2005.06.062

11.20 N. Pinna, G. Garnweitner, M. Antonietti, M. Niederberger: A general nonaqueous route to binary metal oxide nanocrystals involving a C–C bond cleavage, J. Am. Chem. Soc. **127**, 5608–5612 (2005) doi:10.1021/ja042323r

11.21 D.L. Huber, E.L. Venturini, J.E. Martin, P.P. Provencio, R.J. Patel: Synthesis of highly magnetic iron nanoparticles suitable for field structuring using a beta-diketone surfactant, J. Magnetism Magnet. Mater. **278**, 311–316 (2004) doi:10.1016/j.jmmm.2003.12.1317

11.22 R. Kumar, S. Lal: Synthesis of organic nanoparticles and their applications in drug delivery and food nanotechnology: A review, J. Nanomater. Mol. Nanotechnol. **3**, 4 (2014) doi:10.4172/2324-8777.1000150

11.23 I. Sole, A. Maestro, C. Gonzalez, C. Solans, J.M. Gutierrez: Optimization of nano-emulsion preparation by low-energy methods in an ionic surfactant system, Langmuir **22**, 8326–8332 (2006) doi:10.1021/la0613676

11.24 J. Allouche: Synthesis of organic and bioorganic nanoparticles: An overview of the preparation methods. In: *Nanomaterials: A Danger or a Promise?*, ed. by R. Brayner, F. Fiévet, T. Coradin (Springer, London 2013)

11.25 J.C. Lopez-Montilla, P.E. Herrera-Morales, S. Pandey, D.O. Shah: Spontaneous emulsification: Mechanisms, physicochemical aspects, modeling, and applications, J. Dispers. Sci. Technol. **23**, 219–268 (2002) doi:10.1080/01932690208984202

11.26 P. Fernandez, V. Andre, J. Rieger, A. Kuhnle: Nano-emulsion formation by emulsion phase inversion, Colloids Surf. A **251**, 53–58 (2004) doi:10.1016/

j.colsurfa.2004.09.029
11.27 R.L. McCall, R.W. Sirianni: PLGA nanoparticles formed by single- or double-emulsion with vitamin E-TPGS, Jove-J. Visualized Exp. **27**(82), 51015 (2013) doi:10.3791/51015
11.28 J. Singh, S. Pandit, V.W. Bramwell, H.O. Alpar: Diphtheria toxoid loaded poly-(epsilon-caprolactone) nanoparticles as mucosal vaccine delivery systems, Methods **38**, 96–105 (2006) doi:10.1016/j.ymeth.2005.11.003
11.29 C.X. Song, V. Labhasetwar, H. Murphy, X. Qu, W.R. Humphrey, R.J. Shebuski, R.J. Levy: Formulation and characterization of biodegradable nanoparticles for intravascular local drug delivery, J. Control. Release **43**, 197–212 (1997) doi:10.1016/S0168-3659(96)01484-8
11.30 J. Allouche, M. Boissiere, C. Helary, J. Livage, T. Coradin: Biomimetic core-shell gelatine/silica nanoparticles: A new example of biopolymer-based nanocomposites, J. Mater. Chem. **16**, 3120–3125 (2006) doi:10.1039/b604366f
11.31 C. Damge, C.P. Reis, N. Ubrich, P. Maincent, F. Veiga, A. Ribeiro: Alginate-based insulin nanoparticles – Dextrane for oral administration in the diabetic rat, Diabetes Metab. **33**, S67–S67 (2007) doi:10.1186/s12951-015-0136-y
11.32 D. Barati, S.R.P. Shariati, S. Moeinzadeh, J.M. Melero-Martin, A. Khademhosseini, E. Jabbari: Spatiotemporal release of BMP-2 and VEGF enhances osteogenic and vasculogenic differentiation of human mesenchymal stem cells and endothelial colony-forming cells co-encapsulated in a patterned hydrogel, J. Control. Release **223**, 126–136 (2016) doi:10.1016/j.jconrel.2015.12.031
11.33 J.N.L. De Avila, L.L.G.C. De Araujo, S. Drexler, J.D. Rodrigues, R.S.V. Nascimento: Polystyrene nanoparticles as surfactant carriers for enhanced oil recovery, J. Appl. Polymer Sci. **133**, 43789 (2016) doi:10.1002/app.43789
11.34 S.J. Yoon, H. Chun, M.S. Lee, N. Kim: Preparation of poly(N-vinylcarbazole) (PVK) nanoparticles by emulsion polymerization and PVK hollow particles, Synth. Metals **159**, 518–522 (2009) doi:10.1016/j.synthmet.2008.11.011
11.35 A.M. Oliveira, K.L. Guimaraes, N.N.P. Cerize: The role of functional monomers on producing nanostructured lattices obtained by surfactant-free emulsion polymerization – A novel approach, Eur. Polym. J. **71**, 268–278 (2015) doi:10.1016/j.eurpolymj.2015.07.049
11.36 B.V.N. Nagavarma, K.S.Y. Hemant, A. Ayaz, L.S. Vasudha, H.G. Shivakumar: Different techniques for preparation of polymeric nanoparticles – A review, Asian J. Pharm. Clin. Res. **5**(3), 16–23 (2012)
11.37 C.P. Reis, R.J. Neufeld, A.J. Ribeiro, F. Veiga: Nanoencapsulation I. Methods for preparation of drug-loaded polymeric nanoparticles, Nanomedicine **2**, 8–21 (2006) doi:10.1016/j.nano.2005.12.003
11.38 T. Musumeci, C.A. Ventura, C. Carbone, R. Pignatello, G. Puglisi: Effects of external phase on D-cycloserine loaded W/O nanocapsules prepared by the interfacial polymerization method, Eur. J. Med. Chem. **46**, 2828–2834 (2011) doi:10.1016/j.ejmech.2011.04.003
11.39 S. Watnasirichaikul, N.M. Davies, T. Rades, I.G. Tucker: Preparation of biodegradable insulin nanocapsules from biocompatible microemulsions, Pharm. Res. **17**, 684–689 (2000) doi:10.1023/A:1007574030674
11.40 S. Kuypers, S.K. Pramanik, L. D'Olieslaeger, G. Reekmans, M. Peters, J. D'Haen, D. Vanderzande, T. Junkers, P. Adriaensens, A. Ethirajan: Interfacial thiol-isocyanate reactions for functional nanocarriers: A facile route towards tunable morphologies and hydrophilic payload encapsulation, Chem. Commun. **51**, 15858–15861 (2015) doi:10.1039/c5cc05258k
11.41 M.J. Monteiro, M.F. Cunningham: Polymer nanoparticles via living radical polymerization in aqueous dispersions: Design and applications, Macromolecules **45**, 4939–4957 (2012) doi:10.1021/ma300170c
11.42 K.O. Sebakhy, S. Kessel, M.J. Monteiro: Nanoreactors to synthesize well-defined polymer nanoparticles: Decoupling particle size from molecular weight, Macromolecules **43**, 9598–9600 (2010) doi:10.1021/ma1019889
11.43 C. Dire, S. Magnet, L. Couvreur, B. Charleux: Nitroxide-mediated controlled/living free-radical surfactant-free emulsion polymerization of methyl methacrylate using a poly(methacrylic acid)-based macroalkoxyamine initiator, Macromolecules **42**, 95–103 (2009) doi:10.1021/ma802083g
11.44 R. Gonzalez-Blanco, M.F. Cunningham, E. Saldivar-Guerra: High solids TEMPO-mediated radical semibatch emulsion polymerization of styrene, J. Polym. Sci. Part A **54**, 49–62 (2016) doi:10.1002/pola.27771
11.45 J. Nicolas, A.V. Ruzette, C. Farcet, P. Gerard, S. Magnet, B. Charleux: Nanostructured latex particles synthesized by nitroxide-mediated controlled/living free-radical polymerization in emulsion, Polymer **48**, 7029–7040 (2007) doi:10.1016/j.polymer.2007.09.039
11.46 W.W. Li, K. Matyjaszewski, K. Albrecht, M. Moller: Reactive surfactants for polymeric nanocapsules via interfacially confined miniemulsion ATRP, Macromolecules **42**, 8228–8233 (2009) doi:10.1021/ma901574y
11.47 K. Min, H. Gao, J.A. Yoon, W. Wu, T. Kowalewski, K. Matyjaszewski: One-pot synthesis of hairy nanoparticles by emulsion ATRP, Macromolecules **42**, 1597–1603 (2009) doi:10.1021/ma8026244
11.48 D.J. Siegwart, A. Srinivasan, S.A. Bencherif, A. Karunanidhi, J.K. Oh, S. Vaidya, R. Jin, J.O. Hollinger, K. Matyjaszewki: Cellular uptake of functional nanogels prepared by inverse miniemulsion atrp with encapsulated proteins, carbohydrates, and gold nanoparticles, Biomacromolecules **10**, 2300–2309 (2009) doi:10.1021/bm9004904

11.49 F.Y. Cheng, S.P.H. Wang, C.H. Su, T.L. Tsai, P.C. Wu, D.B. Shieh, J.H. Chen, P.C.H. Hsieh, C.S. Yeh: Stabilizer-free poly(lactide-co-glycolide) nanoparticles for multimodal biomedical probes, Biomaterials **29**, 2104–2112 (2008) doi:10.1016/j.biomaterials.2008.01.01

11.50 P. Legrand, S. Lesieur, A. Bochot, R. Gref, W. Raatjes, G. Barratt, C. Vauthier: Influence of polymer behaviour in organic solution on the production of polylactide nanoparticles by nanoprecipitation, Int. J. Pharm. **344**, 33–43 (2007) doi:10.1016/j.ijpbarm.2007.05.054

11.51 S.K. Pandey, C. Haldar, D.K. Vishwas, P. Maiti: Synthesis and in vitro evaluation of melatonin entrapped PLA nanoparticles: An oxidative stress and T-cell response using golden hamster, J. Biomed. Mater. Res. A **103**, 3034–3044 (2015) doi:10.1002/jbm.a.35441

11.52 S.K. Pandey, D.K. Patel, A.K. Maurya, R. Thakur, D.P. Mishra, M. Vinayak, C. Haldar, P. Maiti: Controlled release of drug and better bioavailability using poly(lacticacid-co-glycolic acid) nanoparticles, Int. J. Biol. Macromol. **89**, 99–110 (2016) doi:10.1016/j.ijbiomac.2016.04.065

11.53 K. Sawant, A. Pandey, S. Patel: Aripiprazole loaded poly(caprolactone) nanoparticles: Optimization and in vivo pharmacokinetics, Mater. Sci. Eng. C **66**, 230–243 (2016) doi:10.1016/j.msec.2016.04.089

11.54 Z. Zili, S. Sfar, H. Fessi: Preparation and characterization of poly-e-caprolactone nanoparticles containing griseofulvin, Int. J. Pharm. **294**, 261–267 (2005) doi:10.1016/j.ijpharm.2005.01.020

11.55 A.O. Elzoghby: Gelatin-based nanoparticles as drug and gene delivery systems: Reviewing three decades of research, J. Control. Release **172**, 1075–1091 (2013) doi:10.1016/j.jconrel.2013.09.019

11.56 L. Songtipya, M.C. Thies, A. Sane: Effect of rapid expansion of subcritical solutions processing conditions on loading capacity of tetrahydrocurcumin encapsulated in poly(L-lactide) particles, J. Supercrit. Fluids **113**, 119–127 (2016) doi:10.1016/j.supflu.2016.03.020

11.57 S.K. Sharma, R. Jagannathan: High throughput RESS processing of sub-10 nm ibuprofen nanoparticles, J. Supercrit. Fluids **109**, 74–79 (2016) doi:10.1016/j.supflu.2015.11.019

11.58 M. Turk, D. Bolten: Formation of submicron poorly water-soluble drugs by rapid expansion of supercritical solution (RESS): Results for Naproxen, J. Supercrit. Fluids **55**, 778–785 (2010) doi:10.1016/j.supflu.2010.09.023

11.59 J.P. Chapel, J.F. Berret: Versatile electrostatic assembly of nanoparticles and polyelectrolytes: Coating, clustering and layer-by-layer processes, Curr. Opin. Colloid Interface Sci. **17**, 97–105 (2012) doi:10.1016/j.cocis.2011.08.009

11.60 G. Schneider, G. Decher: Functional core/shell nanoparticles via layer-by-layer assembly. Investigation of the experimental parameters for controlling particle aggregation and for enhancing dispersion stability, Langmuir **24**, 1778–1789 (2008) doi:10.1021/la7021837

11.61 A. Aqil, S. Vasseur, E. Duguet, C. Passirani, J.P. Benoit, A. Roch, R. Muller, R. Jerome, C. Jerome: PEO coated magnetic nanoparticles for biomedical application, Eur. Polym. J. **44**, 3191–3199 (2008) doi:10.1016/j.eurpolymj.2008.07.011

11.62 P. Froimowicz, R. Munoz-Espi, K. Landfester, A. Musyanovych, D. Crespy: Surface-functionalized particles: From their design and synthesis to materials science and bio-applications, Curr. Org. Chem. **17**, 900–912 (2013) doi:10.2174/1385272811317090004

11.63 E.J. Tang, G.X. Cheng, X.L. Ma, X.S. Pang, Q. Zhao: Surface modification of zinc oxide nanoparticle by PMAA and its dispersion in aqueous system, Appl. Surf. Sci. **252**, 5227–5232 (2006) doi:10.1016/j.apsusc.2005.08.004

11.64 M.N. Tchoul, S.P. Fillery, H. Koerner, L.F. Drummy, F.T. Oyerokun, P.A. Mirau, M.F. Durstock, R.A. Vaia: Assemblies of titanium dioxide-polystyrene hybrid nanoparticles for dielectric applications, Chem. Mater. **22**, 1749–1759 (2010), doi:10.1021/cm903182n

11.65 M. Li, L.Q. Xu, L. Wang, Y.P. Wu, J. Li, K.G. Neoh, E.T. Kang: Clickable poly(ester amine) dendrimer-grafted Fe304 nanoparticles prepared via successive Michael addition and alkyne-azide click chemistry, Polym. Chem. **2**, 1312–1321 (2011) doi:10.1039/c1py00084e

11.66 S. Kango, S. Kalia, A. Celli, J. Njuguna, Y. Habibi, R. Kumar: Surface modification of inorganic nanoparticles for development of organic-inorganic nanocomposites-A review, Prog. Polym. Sci. **38**, 1232–1261 (2013) doi:10.1016/j.progpolymsci.2013.02.003

11.67 T. von Werne, T.E. Patten: Atom transfer radical polymerization from nanoparticles: A tool for the preparation of well-defined hybrid nanostructures and for understanding the chemistry of controlled/living radical polymerizations from surfaces, J. Am. Chem. Soc. **123**, 7497–7505 (2001) doi:10.1021/Ja010235q

11.68 A. Khabibullin, K. Bhangaonkar, C. Mahoney, Z. Lu, M. Schmitt, A.K. Sekizkardes, M.R. Bockstaller, K. Matyjaszewski: Grafting PMMA brushes from alpha-alumina nanoparticles via SI-ATRP, ACS Appl. Mater. Interfaces **8**, 5458–5465 (2016) doi:10.1021/acsami.5b12311

11.69 H.D. Wang, W.Q. Luo, J.C. Chen: Fabrication and characterization of thermoresponsive Fe304@PNIPAM hybrid nanomaterials by surface-initiated RAFT polymerization, J. Mater. Sci. **47**, 5918–5925 (2012) doi:10.1007/s10853-012-6493-0

11.70 M.A. Hood, M. Mari, R. Munoz-Espi: Nanodroplet, Materials **7**, 4057–4087 (2014) doi:10.3390/ma7054057

11.71 E. Bourgeat-Lami, G.A. Farzi, L. David, J.L. Putaux, T.F.L. McKenna: Silica encapsulation by miniemulsion polymerization: distribution and localization of the silica particles in droplets and latex particles, Langmuir **28**, 6021–6031 (2012) doi:10.1021/la300587b

11.72 Z.X. Chen, Y.H. Zhang, L.L. Duan, Z.G. Wang, Y.L. Li, P.X. He: Preparation of PSt/SiO_2 nanoparticles with raspberry-like structure via nonionic surfactant miniemulsion polymerization, J. Adhes. Sci. Technol. **29**, 2117–2129 (2015) doi:10.1080/01694243.2015.1057396

11.73 M.M. Rahim-Abadi, A.R. Mandavian, A. Gharieh, H. Salehi-Mobarakeh: Chemical modification of Ti02 nanoparticles as an effective way for encapsulation in polyacrylic shell via emulsion polymerization, Prog. Org. Coat. **88**, 310–315 (2015) doi:10.1016/j.porgcoat.2015.07.013

11.74 A. Perro, S. Reculusa, E. Bourgeat-Lami, E. Duguet, S. Ravaine: Synthesis of hybrid colloidal particles: From snowman-like to raspberry-like morphologies, Colloids Surf. A **284**, 78–83 (2006) doi:10.1016/j.colsurfa.2005.11.073

11.75 C.S. Wagner, S. Shehata, K. Henzler, J.Y. Yuan, A. Wittemann: Towards nanoscale composite particles of dual complexity, J. Colloid Interface Sci. **355**, 115–123 (2011) doi:10.1016/j.jcis.2010.12.017

11.76 Q.J. Luo, R.J. Hickey, S.J. Park: Controlling the location of nanoparticles in colloidal assemblies of amphiphilic polymers by tuning nanoparticle surface chemistry, ACS Macro Lett. **2**, 107–111 (2013) doi:10.1021/mz3006044

11.77 M. Agrawal, J. Rubio-Retama, N.E. Zafeiropoulos, N. Gaponik, S. Gupta, V. Cimrova, V. Lesnyak, E. Lopez-Cabarcos, S. Tzavalas, R. Rojas-Reyna, A. Eychmuller, M. Stamm: Switchable photoluminescence of CdTe nanocrystals by temperature-responsive microgels, Langmuir **24**, 9820–9824 (2008) doi:10.1021/la801347d

11.78 M. Schrinner, M. Ballauff, Y. Talmon, Y. Kauffmann, J. Thun, M. Moller, J. Breu: Single nanocrystals of platinum prepared by partial dissolution of Au-Pt nanoalloys, Science **323**, 617–620 (2009) doi:10.1126/science.1166703

11.79 V. Fischer, I. Lieberwirth, G. Jakob, K. Landfester, R. Munoz-Espi: Metal oxide/polymer hybrid nanoparticles with versatile functionality prepared by controlled surface crystallization, Adv. Funct. Mater. **23**, 451–466 (2013) doi:10.1002/adfm.201201839

11.80 A. Ethirajan, U. Ziener, K. Landfester: Surface-functionalized polymeric nanoparticles as templates for biomimetic mineralization of hydroxyapatite, Chem. Mater. **21**, 2218–2225 (2009) doi:10.1021/cm9001724

11.81 D. Barati, J.D. Walters, S.R.P. Shariati, S. Moeinzadeh, E. Jabbari: Effect of organic acids on calcium phosphate nucleation and osteogenic differentiation of human mesenchymal stem cells on peptide functionalized nanofibers, Langmuir **31**, 5130–5140 (2015) doi:10.1021/acs.langmuir.5b00615

11.82 O. Karaman, A. Kumar, S. Moeinzadeh, X.Z. He, T. Cui, E. Jabbari: Effect of surface modification of nanofibres with glutamic acid peptide on calcium phosphate nucleation and osteogenic differentiation of marrow stromal cells, J. Tissue Eng. Regen. Med. **10**, E132–E146 (2016) doi:10.1002/term.1775

11.83 A. Schrade, K. Landfester, U. Ziener: Pickering-type stabilized nanoparticles by heterophase polymerization, Chem. Soc. Rev. **42**, 6823–6839 (2013) doi:10.1039/c3cs60100e

11.84 M.J. Percy, C. Barthet, J.C. Lobb, M.A. Khan, S.F. Lascelles, M. Vamvakaki, S.P. Armes: Synthesis and characterization of vinyl polymer-silica colloidal nanocomposites, Langmuir **16**, 6913 (2000) doi:10.1021/la0004294

11.85 A. Schmid, S.P. Armes, C.A.P. Leite, F. Galembeck: Efficient preparation of polystyrene/silica colloidal nanocomposite particles by emulsion polymerization using a glycerol-functionalized silica sol, Langmuir **25**, 2486–2494 (2009)

11.86 J.H. Chen, C.Y. Cheng, W.Y. Chiu, C.F. Lee, N.Y. Liang: Synthesis of ZnO/polystyrene composites particles by Pickering emulsion polymerization, Eur. Polym. J. **44**, 3271–3279 (2008) doi:10.1016/j.eurpolymj.2008.07.023

11.87 S. Stanley: Biological nanoparticles and their influence on organisms, Curr. Opin. Biotechnol. **28**, 69–74 (2014) doi:10.1016/j.copbio.2013.11.014

11.88 C.S. Thaxton, J.S. Rink, P.C. Naha, D.P. Cormode: Lipoproteins and lipoprotein mimetics for imaging and drug delivery, Adv. Drug Deliv. Rev. **106**, 116–131 (2016)

11.89 N. Sabnis, M. Nair, M. Israel, W.J. McConathy, A.G. Lacko: Enhanced solubility and functionality of valrubicin (AD-32) against cancer cells upon encapsulation into biocompatible nanoparticles, Int. J. Nanomed. **7**, 975–983 (2012) doi:10.2147/ijn.S28029

11.90 K.B. Johnsen, J.M. Gudbergsson, M.N. Skov, L. Pilgaard, T. Moos, M. Duroux: A comprehensive overview of exosomes as drug delivery vehicles – Endogenous nanocarriers for targeted cancer therapy, Biochim. Biophys. Acta **1846**, 75–87 (2014) doi:10.1016/j.bbcan.2014.04.005

11.91 B.S. Batista, W.S. Eng, K.T. Pilobello, K.D. Hendricks-Munoz, L.K. Mahal: Identification of a conserved glycan signature for microvesicles, J. Proteome Res. **10**, 4624–4633 (2011) doi:10.1021/pr200434y

11.92 F.T. Borges, S.A. Melo, B.C. Ozdemir, N. Kato, I. Revuelta, C.A. Miller, V.H. Gattone, V.S. LeBleu, R. Kalluri: TGF-beta 1-containing exosomes from injured epithelial cells activate fibroblasts to initiate tissue regenerative responses and fibrosis, J. Am. Soc. Nephrol. **24**, 385–392 (2013) doi:10.1681/Asn.2012101031

11.93 G. Raposo, H.W. Nijman, W. Stoorvogel, R. Leijendekker, C.V. Harding, C.J.M. Melief, H.J. Geuze: B lymphocytes secrete antigen-presenting vesicles, J. Exp. Med. **183**, 1161–1172 (1996) doi:10.1084/jem.183.3.1161

11.94 T.N. Lamichhane, S. Sokic, J.S. Schardt, R.S. Raiker, J.W. Lin, S.M. Jay: Emerging roles for extracellular vesicles in tissue engineering and regenerative medicine, Tissue Eng. Part B Rev. **21**, 45–54 (2015) doi:10.1089/ten.teb.2014.0300

11.95 M. Allen, D. Willits, J. Mosolf, M. Young, T. Douglas: Protein cage constrained synthesis of ferrimagnetic iron oxide nanoparticles, Adv. Mater. **14**, 1562 (2002) doi:10.1002/1521-4095(20021104)14:21<1562::Aid-Adma1562>3.0.Co;2-D

11.96 I. Yamashita, K. Iwahori, S. Kumagai: Ferritin in the field of nanodevices, Biochim. Biophys. Acta-Gen. Subjects **1800**, 846–857 (2010) doi:10.1016/j.bbagen.2010.03.005

11.97 K.L. Fan, C.Q. Cao, Y.X. Pan, D. Lu, D.L. Yang, J. Feng, L.N. Song, M.M. Liang, X.Y. Yan: Magnetoferritin nanoparticles for targeting and visualizing tumour tissues, Nat. Nanotechnol. **7**, 459–464 (2012) doi:10.1038/Nnano.2012.90

11.98 D. Yan, Y.Q. Wei, H.C. Guo, S.Q. Sun: The application of virus-like particles as vaccines and biological vehicles, Appl. Microbiol. Biotechnol. **99**, 10415–10432 (2015) doi:10.1007/s00253-015-7000-8

11.99 Z.J. Wu, K.V. Chen, I. Yildiz, A. Dirksen, R. Fischer, P.E. Dawson, N.F. Steinmetz: Development of viral nanoparticles for efficient intracellular delivery, Nanoscale **4**, 3567–3576 (2012) doi:10.1039/c2nr30366c

11.100 D.P. Patterson, A. Rynda-Apple, A.L. Harmsen, A.G. Harmsen, T. Douglas: Biomimetic antigenic nanoparticles elicit controlled protective immune response to influenza, ACS Nano **7**, 3036–3044 (2013) doi:10.1021/nn4006544

11.101 W. Lohcharoenkal, L.Y. Wang, Y.C. Chen, Y. Rojanasakul: Protein nanoparticles as drug delivery carriers for cancer therapy, Biomed Res. Int. **2014**, 180549 (2014)

11.102 L.P.H. Estrada, J.A. Champion: Protein nanoparticles for therapeutic protein delivery, Biomater. Sci. **3**, 787–799 (2015) doi:10.1039/c5bm00052a

11.103 K. Langer, M.G. Anhorn, I. Steinhauser, S. Dreis, D. Celebi, I. Schrickel, S. Faust, V. Vogel: Human serum albumin (HSA) nanoparticles: Reproducibility of preparation process and kinetics of enzymatic degradation, Int. J. Pharm. **347**, 109–117 (2008) doi:10.1016/j.ijpharm.2007.06.028

11.104 G.L. Wang, K. Siggers, S.F. Zhang, H.X. Jiang, Z.H. Xu, R.F. Zernicke, J. Matyas, H. Uludag: Preparation of BMP-2 containing bovine serum albumin (BSA) nanoparticles stabilized by polymer coating, Pharm. Res. **25**, 2896–2909 (2008) doi:10.1007/s11095-008-9692-2

11.105 G. Kaul, M. Amiji: Biodistribution and targeting potential of poly(ethylene glycol)-modified gelatin nanoparticles in subcutaneous murine tumor model, J. Drug Target. **12**, 585–591 (2004) doi:10.1080/10611860400013451

11.106 A.O. Elzoghby, M.W. Helmy, W.M. Samy, N.A. Elgindy: Novel ionically crosslinked casein nanoparticles for flutamide delivery: formulation, characterization, and in vivo pharmacokinetics, Int. J. Nanomed. **8**, 1721–1732 (2013) doi:10.2147/Ijn.S40674

11.107 L. Yang, F. Cui, D.M. Cun, A. Tao, K. Shi, W.H. Lin: Preparation, characterization and biodistribution of the lactone form of 10-hydroxycamptothecin (HCPT)-loaded bovine serum albumin (BSA) nanoparticles, Int. J. Pharm. **340**, 163–172 (2007) doi:10.1016/j.ijpharm.2007.03.028

11.108 S. Rhaese, H. von Briesen, H. Rubsamen-Waigmann, J. Kreuter, K. Langer: Human serum albumin-polyethylenimine nanoparticles for gene delivery, J. Control. Release **92**, 199–208 (2003) doi:10.1016/S0168-3659(03)00302-X

11.109 I.W. Hamley: Self-assembly of amphiphilic peptides, Soft Matter **7**, 4122–4138 (2011) doi:10.1039/c0sm01218a

11.110 S.K. Nitta, K. Numata: Biopolymer-based nanoparticles for drug/gene delivery and tissue engineering, Int. J. Mol. Sci. **14**, 1629–1654 (2013) doi:10.3390/ijms14011629

11.111 S. Santoso, W. Hwang, H. Hartman, S.G. Zhang: Self-assembly of surfactant-like peptides with variable glycine tails to form nanotubes and nanovesicles, Nano Lett. **2**, 687–691 (2002) doi:10.1021/nl025563i

11.112 G. von Maltzahn, S. Vauthey, S. Santoso, S.U. Zhang: Positively charged surfactant-like peptides self-assemble into nanostructures, Langmuir **19**, 4332–4337 (2003) doi:10.1021/la026526+

11.113 E. Jabbari, X.M. Yang, S. Moeinzadeh, X.Z. He: Drug release kinetics, cell uptake, and tumor toxicity of hybrid VVVVVKK peptide-assembled polylactide nanoparticles, Eur. J. Pharm. Biopharm. **84**, 49–62 (2013) doi:10.1016/j.ejpb.2012.12.012

11.114 H.G. Cui, M.J. Webber, S.I. Stupp: Self-assembly of peptide amphiphiles: From molecules to nanostructures to biomaterials, Biopolymers **94**, 1–18 (2010) doi:10.1002/bip.21328

11.115 A. Dehsorkhi, V. Castelletto, I.W. Hamley: Self-assembling amphiphilic peptides, J. Pept. Sci. **20**, 453–467 (2014) doi:10.1002/psc.2633

11.116 I.W. Hamley, A. Dehsorkhi, V. Castelletto, S. Furzeland, D. Atkins, J. Seitsonen, J. Ruokolainen: Reversible helical unwinding transition of a self-assembling peptide amphiphile, Soft Matter **9**, 9290–9293 (2013) doi:10.1039/c3sm51725j

11.117 S. Eskandari, T. Guerin, I. Toth, R.J. Stephenson: Recent advances in self-assembled peptides: Implications for targeted drug delivery and vaccine engineering, Adv. Drug Deliv. Rev. (2016) doi:10.1016/j.addr.2016.06.013, in press

11.118 R. Herrero-Vanrell, A.C. Rincon, M. Alonso, V. Reboto, I.T. Molina-Martinez, J.C. Rodriguez-Cabello: Self-assembled particles of an elastin-like polymer as vehicles for controlled drug release, J. Control. Release **102**, 113–122 (2005) doi:10.1016/j.jconrel.2004.10.001

11.119 A. Yeboah, R.I. Cohen, R. Faulknor, R. Schloss, M.L. Yarmush, F. Berthiaume: The development and characterization of SDF1 alpha-elastin-like-peptide nanoparticles for wound healing, J. Control. Release **232**, 238–247 (2016) doi:10.1016/j.jconrel.2016.04.020

11.120 J. Rodriguez-Hernandez, S. Lecommandoux: Reversible inside-out micellization of pH-responsive and water-soluble vesicles based on polypeptide diblock copolymers, J. Am. Chem.

Soc. **127**, 2026–2027 (2005) doi:10.1021/ja043920g
11.121 B. Sarmento, A.J. Ribeiro, F. Veiga, D.C. Ferreira, R.J. Neufeld: Insulin-loaded nanoparticles are prepared by alginate ionotropic pre-gelation followed by chitosan polyelectrolyte complexation, J. Nanosci. Nanotechnol. **7**, 2833–2841 (2007) doi:10.1166/jnn.2007.609
11.122 R.Y. Rasente, J.C. Imperiale, J.M. Lazaro-Martinez, L. Gualco, R. Oberkersch, A. Sosnik, G.C. Calabrese: Dermatan sulfate/chitosan polyelectrolyte complex with potential application in the treatment and diagnosis of vascular disease, Carbohydr. Polym. **144**, 362–370 (2016) doi:10.1016/j.carbpol.2016.02.046
11.123 K.A. Janes, P. Calvo, M.J. Alonso: Polysaccharide colloidal particles as delivery systems for macromolecules, Adv. Drug Deliv. Rev. **47**, 83–97 (2001) doi:10.1016/S0169-409x(00)00123-X
11.124 Y. Cho, J.T. Kim, H.J. Park: Preparation, characterization, and protein loading properties of N-acyl chitosan nanoparticles, J. Appl. Polym. Sci. **124**, 1366–1371 (2012) doi:10.1002/app.34931
11.125 G. Shi, Y.X. Che, Y.M. Zhou, X. Bai, C.H. Ni: Synthesis of polyglycolic acid grafting from sodium alginate through direct polycondensation and its application as drug carrier, J. Mater. Sci. **50**, 7835–7841 (2015) doi:10.1007/s10853-015-9363-8
11.126 C.J. Brinker, G.W. Scherer: *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing* (Academic, San Diego 1990)
11.127 A.G. Kolhatkar, A.C. Jamison, D. Litvinov, R.C. Willson, T.R. Lee: Tuning the magnetic properties of nanoparticles, Int. J. Mol. Sci. **14**, 15977–16009 (2013) doi:10.3390/ijms140815977
11.128 C. Burda, X.B. Chen, R. Narayanan, M.A. El-Sayed: Chemistry and properties of nanocrystals of different shapes, Chem. Rev. **105**, 1025–1102 (2005) doi:10.1021/cr030063a
11.129 F. Alexis, E. Pridgen, L.K. Molnar, O.C. Farokhzad: Factors affecting the clearance and biodistribution of polymeric nanoparticles, Mol. Pharm. **5**, 505–515 (2008) doi:10.1021/mp800051m
11.130 J.P. Jolivet, S. Cassaignon, C. Chaneac, D. Chiche, O. Durupthy, D. Portehault: Design of metal oxide nanoparticles: Control of size, shape, crystalline structure and functionalization by aqueous chemistry, C.R. Chimie **13**, 40–51 (2010) doi:10.1016/j.crci.2009.09.012
11.131 A.M. Perez-Coronado, L. Calvo, N. Alonso-Morales, F. Heras, J.J. Rodriguez, M.A. Gilarranz: Multiple approaches to control and assess the size of Pd nanoparticles synthesized via water-in-oil microemulsion, Colloids Surf. A **497**, 28–34 (2016) doi:10.1016/j.colsurfa.2016.02.012
11.132 D. Lemoine, V. Preat: Polymeric nanoparticles as delivery system for influenza virus glycoproteins, J. Control. Release **54**, 15–27 (1998) doi:10.1016/S0168-3659(97)00241-1
11.133 I. Lisiecki, M.P. Pileni: Copper metallic particles synthesized in-situ in reverse micelles - influence of various parameters on the size of the particles, J. Phys. Chem. **99**, 5077–5082 (1995) doi:10.1021/J100014a030
11.134 M. Husein, E. Rodil, J.H. Vera: Formation of silver bromide precipitate of nanoparticles in a single microemulsion utilizing the surfactant counterion, J. Colloid Interface Sci. **273**, 426–434 (2004) doi:10.1016/j.jcis.2004.02.057
11.135 N.M. Husein, E. Rodil, J.H. Vera: Preparation of AgBr nanoparticles in microemulsions via reaction of AgNO3 with CTAB counterion, J. Nanoparticle Res. **9**, 787–796 (2007) doi:10.1007/s11051-006-9107-4
11.136 S. Schubert, J.T. Delaney, U.S. Schubert: Nanoprecipitation and nanoformulation of polymers: from history to powerful possibilities beyond poly(lactic acid), Soft Matter **7**, 1581–1588 (2011) doi:10.1039/c0sm00862a
11.137 H.J. Jeon, J.I. Jeong, M.K. Jang, Y.H. Park, J.W. Nah: Effect of solvent on the preparation of surfactant-free poly(DL-lactide-co-glycolide) nanoparticles and norfloxacin release characteristics, Int. J. Pharm. **207**, 99–108 (2000) doi:10.1016/S0378-5173(00)00537-8
11.138 A.E. Mercado, J. Ma, X. He, E. Jabbari: Release characteristics and osteogenic activity of recombinant human bone morphogenetic protein-2 grafted to novel self-assembled poly(lactide-co-glycolide fumarate) nanoparticles, J. Control. Release **140**, 148–156 (2009) doi:10.1016/j.jconrel.2009.08.009
11.139 H. Mistry, F. Behafarid, E. Zhou, L.K. Ono, L. Zhang, B.R. Cuenya: Shape-dependent catalytic oxidation of 2-butanol over Pt nanoparticles supported on gamma-Al203, ACS Catalysis **4**, 109–115 (2014) doi:10.1021/cs400888n
11.140 N.P. Truong, M.R. Whittaker, C.W. Mak, T.P. Davis: The importance of nanoparticle shape in cancer drug delivery, Expert Opin. Drug Deliv. **12**, 129–142 (2015) doi:10.1517/17425247.2014.950564
11.141 A. Tao, P. Sinsermsuksakul, P.D. Yang: Polyhedral silver nanocrystals with distinct scattering signatures, Angew. Chem. Int. Ed. **45**, 4597–4601 (2006) doi:10.1002/anie.200601277
11.142 L.A. Renna, C.J. Boyle, T.S. Gehan, D. Venkataraman: Polymer nanoparticle assemblies: A versatile route to functional mesostructures, Macromolecules **48**, 6353–6368 (2015) doi:10.1021/acs.macromol.5b00375
11.143 J.A. Champion, Y.K. Katare, S. Mitragotri: Particle shape: A new design parameter for micro- and nanoscale drug delivery carriers, J. Control. Release **121**, 3–9 (2007) doi:10.1016/j.jconrel.2007.03.022
11.144 Y.P. Wang, T.J. Merkel, K. Chen, C.A. Fromen, D.E. Betts, J.M. DeSimone: Generation of a library of particles having controlled sizes and shapes via the mechanical elongation of master templates, Langmuir **27**, 524–528 (2011) doi:10.1021/la1045095
11.145 P. Dalhaimer, A.J. Engler, R. Parthasarathy, D.E. Discher: Targeted worm micelles, Biomacromolecules **5**, 1714–1719 (2004) doi:10.1021/bm049884v

11.146 A.H. Lu, E.L. Salabas, F. Schuth: Magnetic nanoparticles: Synthesis, protection, functionalization, and application, Angew. Chem. Int. Ed. **46**, 1222–1244 (2007) doi:10.1002/anie.200602866

11.147 M. Arruebo, R. Fernandez-Pacheco, M.R. Ibarra, J. Santamaria: Magnetic nanoparticles for drug delivery, Nano Today **2**, 22–32 (2007) doi:10.1016/S1748-0132(07)70084-1

11.148 M. Knobel, W.C. Nunes, L.M. Socolovsky, E. De Biasi, J.M. Vargas, J.C. Denardin: Superparamagnetism and other magnetic features in granular materials: A review on ideal and real systems, J. Nanosci. Nanotechnol. **8**, 2836–2857 (2008) doi:10.1166/Jnn.2008.017

11.149 A. Demortiere, P. Panissod, B.P. Pichon, G. Pourroy, D. Guillon, B. Donnio, S. Begin-Colin: Size-dependent properties of magnetic iron oxide nanocrystals, Nanoscale **3**, 225–232 (2011) doi:10.1039/c0nr00521e

11.150 G.L. Zhen, B.W. Muir, B.A. Moffat, P. Harbour, K.S. Murray, B. Moubaraki, K. Suzuki, I. Madsen, N. Agron-Olshina, L. Waddington, P. Mulvaney, P.G. Hartley: Comparative study of the magnetic behavior of spherical and cubic superparamagnetic iron oxide nanoparticles, J. Phys. Chem. C **115**, 327–334 (2011) doi:10.1021/jp104953z

11.151 K.M. Krishnan: Biomedical nanomagnetics: A spin through possibilities in imaging, diagnostics, and therapy, IEEE Trans. Magn. **46**, 2523–2558 (2010) doi:10.1109/Tmag.2010.2046907

11.152 J.H. Lee, Y.M. Huh, Y. Jun, J. Seo, J. Jang, H.T. Song, S. Kim, E.J. Cho, H.G. Yoon, J.S. Suh, J. Cheon: Artificially engineered magnetic nanoparticles for ultra-sensitive molecular imaging, Nat. Med. **13**, 95–99 (2007) doi:10.1038/nm1467

11.153 N. Shamim, L. Hong, K. Hidajat, M.S. Uddin: Thermosensitive polymer (N-isopropylacrylamide) coated nanomagnetic particles: Preparation and characterization, Colloids Surf. B **55**, 51–58 (2007) doi:10.1016/j.colsurfb.2006.11.007

11.154 D. Maharaj, B. Bhushan: Friction, wear and mechanical behavior of nano-objects on the nanoscale, Mater. Sci. Eng. R. **95**, 1–43 (2015) doi:10.1016/j.mser.2015.07.001

11.155 D. Maharaj, B. Bhushan: Scale effects of nanomechanical properties and deformation behavior of Au nanoparticle and thin film using depth sensing nanoindentation, Beilstein J. Nanotechnol. **5**, 822–836 (2014) doi:10.3762/bjnano.5.94

11.156 D. Mordehai, S.W. Lee, B. Backes, D.J. Srolovitz, W.D. Nix, E. Rabkin: Size effect in compression of single-crystal gold microparticles, Acta Mater. **59**, 5202–5215 (2011) doi:10.1016/j.actamat.2011.04.057

11.157 M. Ramos, L. Ortiz-Jordan, A. Hurtado-Macias, S. Flores, J.T. Elizalde-Galindo, C. Rocha, B. Torres, M. Zarei-Chaleshtori, R.R. Chianelli: Hardness and elastic modulus on six-fold symmetry gold nanoparticles, Materials **6**, 198–205 (2013) doi:10.3390/ma6010198

11.158 D.R. Saha, A. Mandal, S. Mitra, M.R. Mada, P. Boughton, S. Bandyopadhyay, D. Chakravorty: Nanoindentation studies on silver nanoparticles, Proc. Int. Conf. Recent Trends in Appl. Phys. Mater. Sci. **1536**, 257–258 (2013) doi:10.1063/1.4810198

11.159 H. Conrad, K. Jung: Effect of grain size from millimeters to nanometers on the flow stress and deformation kinetics of Ag, Mater. Sci. Eng. A **391**, 272–284 (2005) doi:10.1016/j.msea.2004.08.073

11.160 B. Chen, H. Zhang, K.A. Dunphy-Guzman, D. Spagnoli, M.B. Kruger, D.V.S. Muthu, M. Kunz, S. Fakra, J.Z. Hu, Q.Z. Guo, J.F. Banfield: Size-dependent elasticity of nanocrystalline titania, Phys. Rev. B **79**, 125406 (2009) doi:10.1103/PhysRevB.79.125406

11.161 D. Guo, G.X. Xie, J.B. Luo: Mechanical properties of nanoparticles: Basics and applications, J. Phys. D-Appl. Phys. **47**(1), 3001 (2014) doi:10.1088/0022-3727/47/1/013001

11.162 P. Paik, K.K. Kar, D. Deva, A. Sharma: Measurement of mechanical properties of polymer nanospheres by atomic force microscopy: Effects of particle size, Micro Nano Lett. **2**, 72–77 (2007) doi:10.1049/mnl:20070030

11.163 S.S. Tan, R.L. Sherman, W.T. Ford: Nanoscale compression of polymer microspheres by atomic force microscopy, Langmuir **20**, 7015–7020 (2004) doi:10.1021/la049597c

11.164 D. Maharaj, B. Bhushan: Effect of spherical Au nanoparticles on nanofriction and wear reduction in dry and liquid environments, Beilstein J. Nanotechnol. **3**, 759–772 (2012) doi:10.3762/bjnano.3.85

11.165 D. Guo, J.N. Li, L. Chang, J.B. Luo: Measurement of the friction between single polystyrene nanospheres and silicon surface using atomic force microscopy, Langmuir **29**, 6920–6925 (2013) doi:10.1021/la400984d

11.166 E. Gnecco, R. Bennewitz, T. Gyalog, C. Loppacher, M. Bammerlin, E. Meyer, H.J. Guntherodt: Velocity dependence of atomic friction, Phys. Rev. Lett. **84**, 1172–1175 (2000) doi:10.1103/PhysRevLett.84.1172

11.167 U. Simon, G. Schon, G. Schmid: The application of Au-55 clusters as quantum dots, Angewandte Chemie **32**, 250–254 (1993) doi:10.1002/anie.199302501

11.168 D.L. Klein, P.L. McEuen, J.E.B. Katari, R. Roth, A.P. Alivisatos: An approach to electrical studies of single nanocrystals, Appl. Phys. Lett. **68**, 2574–2576 (1996) doi:10.1063/1.116188

11.169 M. De, P.S. Ghosh, V.M. Rotello: Applications of nanoparticles in biology, Adv. Mater. **20**, 4225–4241 (2008) doi:10.1002/adma.200703183

11.170 T.A.P. Rocha-Santos: Sensors and biosensors based on magnetic nanoparticles, Trac-Trends Anal. Chem. **62**, 28–36 (2014) doi:10.1016/j.trac.2014.06.016

11.171 Y.P. Li, B. Srinivasan, Y. Jing, X.F. Yao, M.A. Hugger, J.P. Wang, C.G. Xing: Nanomagnetic competition assay for low-abundance protein biomarker quantification in unprocessed human sera, J. Am. Chem. Soc. **132**, 4388–4392 (2010) doi:10.1021/ja910406a

11.172 J. Panyam, V. Labhasetwar: Biodegradable nanoparticles for drug and gene delivery to cells and tissue, Adv. Drug Delivery Rev. **64**, 61–71 (2012) doi:10.1016/j.addr.2012.09.023

11.173 M.P. Desai, V. Labhasetwar, G.L. Amidon, R.J. Levy: Gastrointestinal uptake of biodegradable microparticles: Effect of particle size, Pharm. Res. **13**, 1838–1845 (1996) doi:10.1023/A:1016085108889

11.174 D. Peer, J.M. Karp, S. Hong, O.C. Faro Kzad, R. Margalit, R. Langer: Nanocarriers as an emerging platform for cancer therapy, Nat. Nanotechnol. **2**, 751–760 (2007) doi:10.1038/nnano.2007.387

11.175 X.Z. He, J.Y. Ma, A.E. Mercado, W.J. Xu, E. Jabbari: Cytotoxicity of paclitaxel in biodegradable self-assembled core-shell poly(lactide-co-glycolide ethylene oxide fumarate) nanoparticles, Pharm. Res. **25**, 1552–1562 (2008) doi:10.1007/s11095-007-9513-z

11.176 O.S. Wolfbeis: An overview of nanoparticles commonly used in fluorescent bioimaging, Chem. Soc. Rev. **44**, 4743–4768 (2015) doi:10.1039/c4cs00392f

11.177 M.K. So, C.J. Xu, A.M. Loening, S.S. Gambhir, J.H. Rao: Self-illuminating quantum dot conjugates for in vivo imaging, Nat. Biotechnol. **24**, 339–343 (2006) doi:10.1038/nbt1188

11.178 J. Estelrich, M.J. Sanchez-Martin, M.A. Busquets: Nanoparticles in magnetic resonance imaging: from simple to dual contrast agents, Int. J. Nanomed. **10**, 1727–1741 (2015) doi:10.2147/IJN.S76501

11.179 H.B. Na, I.C. Song, T. Hyeon: Inorganic nanoparticles for MRI contrast agents, Adv. Mater. **21**, 2133–2148 (2009) doi:10.1002/adma.200802366

11.180 C.J. Jia, F. Schuth: Colloidal metal nanoparticles as a component of designed catalyst, Phys. Chem. Chem. Phys. **13**, 2457–2487 (2011) doi:10.1039/c0cp02680h

11.181 M. Adlim, M.A. Bakar, K.Y. Liew, J. Ismail: Synthesis of chitosan-stabilized platinum and palladium nanoparticles and their hydrogenation activity, J. Mol. Catalysis A **212**, 141–149 (2004) doi:10.1016/j.molcata.2003.08.012

11.182 M. Ruta, N. Semagina, L. Kiwi-Minsker: Monodispersed Pd nanoparticles for acetylene selective hydrogenation: Particle size and support effects, J. Phys. Chem. C **112**, 13635–13641 (2008) doi:10.1021/jp803800w

11.183 K.M. Bratlie, H. Lee, K. Komvopoulos, P.D. Yang, G.A. Somorjai: Platinum nanoparticle shape effects on benzene hydrogenation selectivity, Nano Lett. **7**, 3097–3101 (2007) doi:10.1021/nl0716000

11.184 N. Dimitratos, J.A. Lopez-Sanchez, J.M. Anthonykutty, G. Brett, A.F. Carley, R.C. Tiruvalam, A.A. Herzing, C.J. Kiely, D.W. Knight, G.J. Hutchings: Oxidation of glycerol using gold-palladium alloy-supported nanocrystals, Phys. Chem. Chem. Phys. **11**, 4952–4961 (2009) doi:10.1039/b904317a

11.185 A.H. Battez, R. Gonzalez, J.L. Viesca, J.E. Fernandez, J.M.D. Fernandez, A. Machado, R. Chou, J. Riba: CuO, ZrO_2 and ZnO nanoparticles as antiwear additive in oil lubricants, Wear **265**, 422–428 (2008) doi:10.1016/j.wear.2007.11.013

11.186 X.B. Zhao, R.W. Long, Y. Chen, Z.G. Chen: Synthesis, characterization of Ce0(2)@Si0(2) nanoparticles and their oxide CMP behavior, Microelectron. Eng. **87**, 1716–1720 (2010) doi:10.1016/j.mee.2009.09.012

11.187 Y.C. Kang, S.L.I. Chan: Tensile properties of nanometric Al203 particulate-reinforced aluminum matrix composites, Mater. Chem. Phys. **85**, 438–443 (2004) doi:10.1016/j.matchemphys.2004.02.002

11.188 J. Jiao, X. Sun, T.J. Pinnavaia: Mesostructured silica for the reinforcement and toughening of rubbery and glassy epoxy polymers, Polymer **50**, 983–989 (2009) doi:10.1016/j.polymer.2008.12.042

11.189 J.E. Millstone, D.F.J. Kavulak, C.H. Woo, T.W. Holcombe, E.J. Westling, A.L. Briseno, M.F. Toney, J.M.J. Frechet: Synthesis, properties, and electronic applications of size-controlled poly(3-hexylthiophene) nanoparticles, Langmuir **26**, 13056–13061 (2010) doi:10.1021/la1022938

11.190 Y.S. Ye, X.L. Xie, J. Rick, F.C. Chang, B.J. Hwang: Improved anode materials for lithium-ion batteries comprise non-covalently bonded graphene and silicon nanoparticles, J. Power Sources **247**, 991–998 (2014) doi:10.1016/j.jpowsour.2013.08.048

11.191 H.B. Gu, B. Jin, D.K. Jun, Z. Han: Improved electrochemical performance of LiCoP04 nanoparticles for lithium ion batteries, J. Nanosci. Nanotechnol. **7**, 4037–4040 (2007) doi:10.1166/jnn.2007.078

11.192 K.T. Lee, Y.S. Jung, S.M. Oh: Synthesis of tin-encapsulated spherical hollow carbon for anode material in lithium secondary batteries, J. Am. Chem. Soc. **125**, 5652–5653 (2003) doi:10.1021/ja0345524

11.193 A. Kumar, P.K. Vemula, P.M. Ajayan, G. John: Silver-nanoparticle-embedded antimicrobial paints based on vegetable oil, Nat. Mater. **7**, 236–241 (2008) doi:10.1038/nmat2099

11.194 J.S. Kim, E. Kuk, K.N. Yu, J.H. Kim, S.J. Park, H.J. Lee, S.H. Kim, Y.K. Park, Y.H. Park, C.Y. Hwang, Y.K. Kim, Y.S. Lee, D.H. Jeong, M.H. Cho: Antimicrobial effects of silver nanoparticles, Nanomed.-Nanotechnol. Biol. Med. **3**, 95–101 (2007) doi:10.1016/j.nano.2006.12.001

11.195 S. Baruah, M.N. Khan, J. Dutta: Perspectives and applications of nanotechnology in water treatment, Environ. Chem. Lett. **14**, 1–14 (2016) doi:10.1007/s10311-015-0542-2

12. Graphene

Aravind Vijayaraghavan, Maria Iliut

Graphene is the two-dimensional allotrope of carbon, consisting of a hexagonal arrangement of carbon atoms on a single plane. This chapter begins with a brief history of graphene, followed by a discussion of different methods to produce graphene. The atomic and electronic structure of graphene is then described. The intrinsic properties of graphene can be tailored by nanofabrication, chemistry, electromagnetic fields, etc. Various applications of graphene have been proposed in electronic, optoelectronic and mechanical products. In addition, graphene has emerged as a candidate in chemical, biochemical and biological applications. Derivatives of graphene such as graphene oxide or graphane are also of interest both in terms of fundamental properties and applications.

12.1 **Methods of Production** 365
12.1.1 Micromechanical Exfoliation of Graphite (the Scotch Tape Technique) 365
12.1.2 Chemical Vapor Deposition 366
12.1.3 Decomposition of Carbides 366
12.1.4 Exfoliation in a Solvent 367
12.1.5 Synthetic Organic Chemistry Route 367
12.1.6 Unzipping of Carbon Nanotubes 368
12.1.7 Derivatives of Graphene 368
12.2 **Properties** 369
12.2.1 Structure and Physical Properties 369
12.2.2 Mechanical Properties 369
12.2.3 Electronic Properties 370
12.2.4 Optical Properties 371
12.2.5 Thermal and Thermoelectric Properties 371
12.2.6 Chemical Properties 372
12.2.7 Properties of Graphene Derivatives 372
12.3 **Characterization** 372
12.3.1 Optical Microscopy 372
12.3.2 Transmission Electron Microscopy 372
12.3.3 Scanning Probe Techniques 373
12.3.4 Angle-Resolved Photoemission Spectroscopy (ARPES) 374
12.3.5 Raman Spectroscopy 374
12.3.6 Electrical Characterization 375
12.4 **Applications** 376
12.4.1 Structural and Electrical Composites 377
12.4.2 Transparent Conductive Films 377
12.4.3 Sensors 379
12.4.4 Electronic Applications 379
12.4.5 Spintronics 382
12.4.6 Photonics and Optoelectronics 382
12.4.7 Filters and Barriers 382
12.4.8 Supercapacitors 384
12.4.9 Biomedical Applications 384
12.5 **Conclusions and Outlook** 385
References 386

As a three-dimensional (3-D) material, carbon exists as three predominant allotropes: diamond, graphite and amorphous carbon (carbon black). Graphite has a layered structure, where the sp^2 hybridized carbon atoms are arranged in a hexagonal lattice in each place, while the planes themselves are most commonly AB (Bernal) stacked and held together by Van der Waals forces. A single layer of graphite is known as graphene, which is a two-dimensional (2-D) allotrope of carbon. Carbon nanotubes (CNTs) form its one-dimensional (1-D) counterpart, while fullerenes are the zero-dimensional (0-D) allotropes. These various forms of carbon are summarized in Fig. 12.1.

One might argue that because even a single atom has a finite thickness, graphene is not a 2-D material; this is however an incorrect analysis. Graphene is regarded as a 2-D material for two main reasons, one electronic and one structural. Firstly, the electrons residing in a plane of graphene are confined to move in two dimensions and are forbidden from out-of-plane displacement. The behavior of electrons in graphene is therefore described by two-dimensional physics. Sec-

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_12

ondly, a sheet of graphene can be extended by adding carbon atoms in two dimensions to produce a larger sheet of graphene, which essentially remains the same material. However, the addition of carbon atoms out of the plane of the graphene sheet produces materials that no longer shares the properties of a single atomic layer graphene material, and is therefore no longer graphene.

Despite serving as the fundamental building block for these carbon allotropes, graphene remained a concept until 2004, even though fullerenes were discovered and described in 1985 [12.1] while carbon nanotubes were synthesized and their atomic structure elucidated in 1991 [12.2]. The term graphene was coined in 1987 [12.3], to describe one of the two alternating layers in graphite intercalation compounds (GIC), the other layer being the intercalating agent. It was postulated that independent, freestanding graphene would not become a physical reality since it would voluntarily transform into a more stable allotrope in an attempt to minimize its surface energy. Supported monolayers of carbon, however, had been previously synthesized and described, including epitaxial graphene, which has been known since the 1970s. Chemical derivatives of graphite such as graphite oxide can be traced back to the 1950s and have been known to exist as single layers (graphene oxide).

Graphene occurs naturally as a constituent of bulk graphite, from which one can obtain freestanding graphene. This approach led to the successful exfoliation of a single sheet of carbon atoms by *A. Geim* and *K. Novoselov* at the University of Manchester in 2004 [12.4] using a technique referred to as micromechanical cleavage, or colloquially, the Scotch tape method. This discovery and the subsequent investigations into the properties of graphene earned them the Nobel Prize in Physics in 2010. The existence of graphene, however, does not contradict the physics that predicted that it could not exist. It was discovered that graphene is not truly flat; there exist atomic-scale ripples in the carbon sheet, which accommodates the excess surface energy, thereby stabilizing the structure of graphene.

A surge of research into the structure and properties of graphene ensued. The number of peer-reviewed publications on the topic of graphene has risen exponentially since the seminal publication in 2004, surpassing the number of papers on carbon nanotubes in 2013, and reaching over 50 000 papers in 2015 (Fig. 12.2). It is also interesting to note the emergence of other 2-D materials, such as molybdenum disulphide (MoS_2) in recent years, fueled by the achievements of graphene.

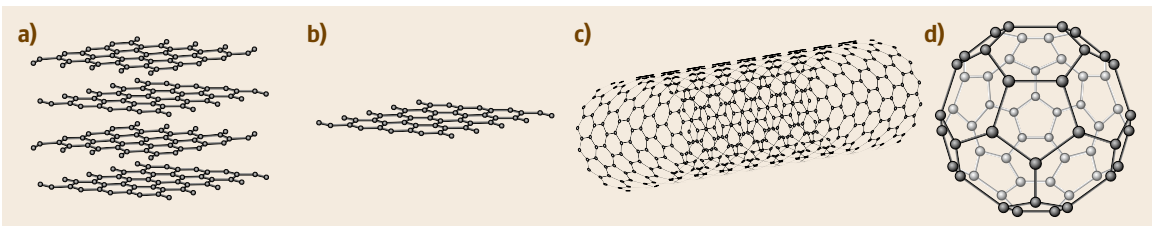


Fig. 12.1a–d Allotropes of sp^2 carbon: **(a)** graphite (3-D), **(b)** graphene (2-D), **(c)** carbon nanotube (1-D), **(d)** fullerene (0-D)

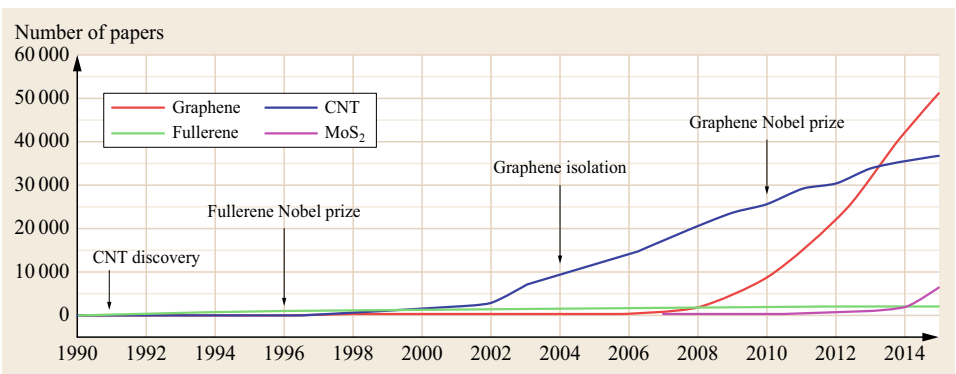


Fig. 12.2 Timeline of nanocarbon allotropes and the number of publications per year on each topic

12.1 Methods of Production

Historically, graphene supported on substrates such as metals and SiC was synthesized first; however, such graphene was not liberated from the substrate support to form a truly 2-D structure. Freestanding graphene was not a reality until it was produced by a micromechanical cleavage method in 2004, and this will be discussed first, followed by the epitaxial synthesis as well as chemical vapor deposition on metallic substrates. Graphene can also be exfoliated from graphite by chemical or mechanical exfoliation in a fluid, followed by a purification step to extract the monolayers. Recently, islands of nanographene have also been synthesized by chemical routes, and this bottom-up approach will be discussed. Finally, a separate section will be devoted to the production of derivatives of graphene.

12.1.1 Micromechanical Exfoliation of Graphite (the Scotch Tape Technique)

Graphite is known to cleave preferentially along the interlayer direction where layers are held together by weak Van der Waals forces, rather than across the strong covalent bonds that bind the atoms within a layer. The most common procedure to accomplish this is using adhesive tape [12.4]. Attaching a thick graphite flake to adhesive tape on both its exposed faces, and then peeling the two pieces of tape apart, results in two thinner flakes of graphite stuck to each of the adhesive tape pieces. Repeating this process sufficient number of times results in some monolayers of carbon atoms, i. e., graphene, adhered to the adhesive tape. This must then be transferred to a suitable weakly coupling substrate or be suspended across supports. Different variations of this method exist. In the resultant sample, the flakes of graphene, bilayer graphene and few-layer graphene must be distinguished from among a sea of flakes of various thicknesses. Fortunately, when deposited on certain thicknesses of the SiO_2 surface layer of a Si wafer, for instance 300 nm, the interference contrast generated by graphene flakes on the surface makes them relatively easy to spot and identify in an optical microscope. A trained eye can, in fact, distinguish between graphene, bilayer graphene, and thicker flakes. Figure 12.3 shows the appearance of micromechanically exfoliated graphene flakes of different thicknesses in an optical microscope [12.5] and an atomic force microscope (AFM) [12.6].

It should be noted that the *best quality* of graphene currently available is undisputedly that produced by micromechanical exfoliation, where the graphene quality is defined in terms of crystalline domain size, number of defects, electronic charge carrier mobility, etc. The choice of initial graphite material having large grain size, using freshly cleaved graphite for further exfoliation, the cleanliness and quality of the adhesive tape and SiO_2 substrate are all variables that significantly affect the quality of the final flakes obtained. Flakes of graphene hundreds of micrometers across are routinely produced in laboratories all over the world by this method, predominantly for fundamental research purposes.

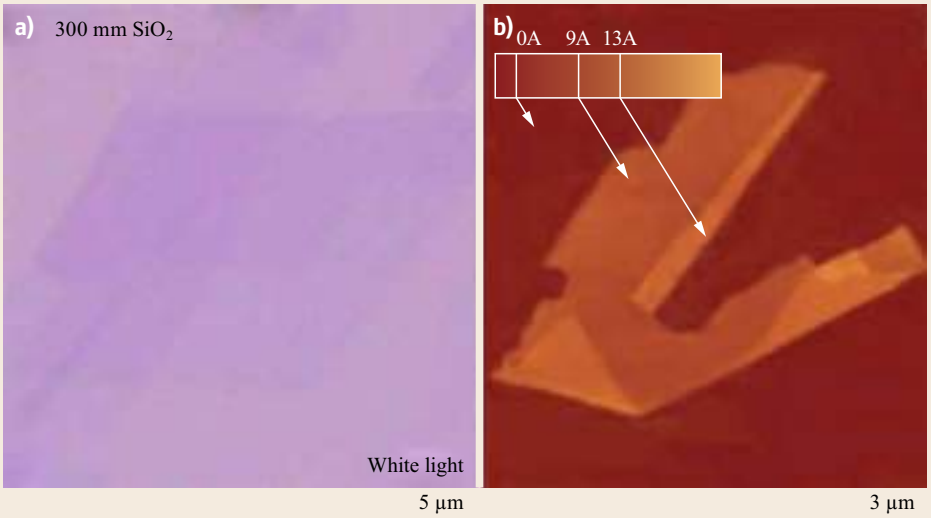


Fig. 12.3a,b Monolayer graphene flake as seen in **(a)** an optical microscope and **(b)** an atomic force microscope. **(a)** Reprinted from [12.5], with the permission of AIP Publishing. **(b)** Reprinted from [12.6]. Copyright (2005) National Academy of Sciences USA

12.1.2 Chemical Vapor Deposition

Chemical vapor deposition (CVD) to produce graphene involves the thermal decomposition of gaseous hydrocarbon sources followed by dissolution and recrystallization of the cracked carbon on the surface of metallic substrates. The solubility of carbon in various metals, such as Rh, Ru, Ir and Re has been measured [12.7] along with the observation that excess carbon dissolved in such metals at high temperatures can segregate as graphite on the surface upon cooling.

At present, the two predominant metallic substrates for CVD graphene growth are Ni and Cu. Large (cm^2)-size films of monolayer and few-layer graphene have been grown on Ni [12.8] with monolayer regions as large as 20 µm in size. Cu does even better, and predominantly monolayer graphene covering many cm^2 are grown in various laboratories using methane CVD [12.9]. The low solubility of C in Cu appears to make the process self-limiting, and at most 5% of the surface is covered by small islands of bilayer graphene. Most importantly, methods have been developed to detach these films from the metallic substrate and transfer them intact onto a dielectric substrate (Fig. 12.4) where they can be lithographically patterned and processed for electronic or optical applications. Truly large-area (m^2) production of graphene is achieved by either continuous or batch CVD and transfer processes.

12.1.3 Decomposition of Carbides

The second substrate-supported route for graphene production involves the thermal decomposition of surface layers of carbides such as SiC. α-SiC is the most commonly encountered polymorph and, until recently, the primary focus of epitaxial graphene growth. α-SiC has a hexagonal crystal structure, similar to Wurtzite. The (0001) (Si) and (000-1) (C) faces of 6H-(α-SiC) [12.11–13] and 4C-(α-SiC) [12.14] are suitable for the growth of epitaxial graphene.

On the Si surface of the hexagonal SiC graphitization occurs at 1350 °C. On this face, the C atoms are in epitaxy with the SiC underneath after graphitization. The surface is passivated by the first C layer, the interface extends to two C layers, and subsequent C layers are decoupled from the substrate and exhibit the properties of graphene. Graphitization occurs at 1150 °C on the C face of the hexagonal SiC and the first C layer already exhibits graphene properties. However, this C layer is no longer epitaxial with the underlying SiC, so the long-range order of the SiC substrate no longer imposes itself upon the C layer. Therefore, it has not been possible to accomplish both long-range ordering as well as decoupling from the surface simultaneously using 6H-(α-SiC).

It was believed that due to its cubic structure, β-SiC would be unsuitable for graphene growth. However, recently scientists have succeeded in growing graphene in the Si-rich (100) surface of β-SiC by a series of annealing cycles with temperature increasing from 1200−1550 K [12.15]. The strong lattice mismatch between graphene and underlying SiC results in very weak coupling similar to the (000-1) C face of α-SiC. However, the graphene growth on β-SiC is was guided along the [110] crystallographic direction despite the lattice mismatch, raising hopes that both substrate-graphene decoupling as well as substrate-guided large domain size might be simultaneously achievable after further process optimization and characterization of the graphene on β-SiC.

SiC decomposition can be used to grow graphene nanoribbons [12.16]. Furthermore, there was significant research into the growth and characterization of graphene on other metal carbides [12.17–19] such as TiC, TaC and HfC, as early as the 1980s. Back then, it was referred to as *monolayer graphite*. As with SiC,

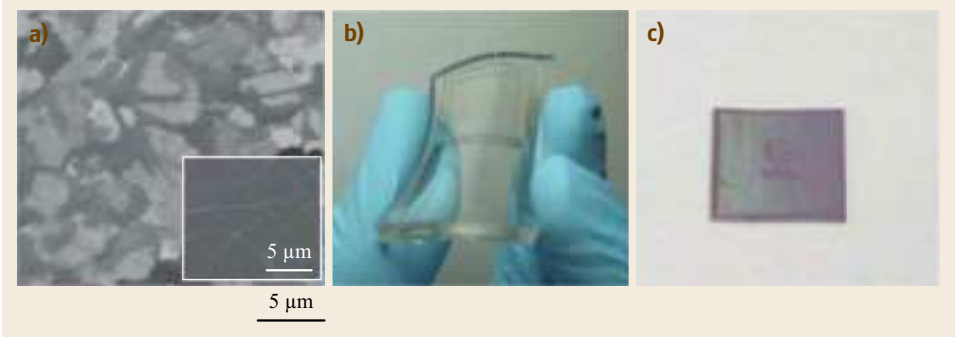


Fig. 12.4a–c (a) Scanning electron microscope (SEM) images of as-grown graphene films on thin (300 nm) nickel layers and thick (1 mm) Ni foils (inset). (b) Graphene films on the polydimethylsiloxane (PDMS) substrates are transparent and flexible. (c) The film transferred onto a SiO_2 substrate (after [12.10])

graphene nanoribbons as narrow as 1.3 nm with well-defined edge structures have been grown for instance on TiC (755) surfaces [12.20].

12.1.4 Exfoliation in a Solvent

The exfoliation of graphene from graphite involves overcoming the interlayer van der Waals bonds. Ultrasonication and high-shear mixing have been used to individualize graphene layers from graphite. This process can be facilitated if the interlayer attraction can be compromised by intercalates. The resultant flakes of graphene in a solvent can be stabilized to prevent aggregation, and separated into fractions that are enriched in particular flake thicknesses.

Graphene Oxide Route

Graphite can be oxidized to graphite oxide in various ways. In the *modified Staudenmaier method* [12.21], a mixture of 97% sulphuric acid and fuming nitric acid is cooled down to 5 °C in an ice bath, and graphite in flake or powder form is added, followed by repeated additions of potassium perchlorate every hour over a period of three days. The resulting solution, sometimes referred to as graphitic acid, is filtered and washed until the pH of the filtrate reaches five or more. The *Brodie method* [12.21] is identical, except that only nitric acid is used, and the potassium perchlorate is added every hour for three hrs. In the *Hummers method* [12.22], graphite is oxidized in a mixture of concentrated sulphuric acid, sodium nitrate and potassium permanganate at 45 °C for two hours. In the *electrochemical method* [12.21], a graphite sheet electrode is anodically polarized in perchloric acid with a platinum wire as the counter electrode. When dried, the above methods result in a powder consisting of graphite oxide flakes.

Graphite oxide is dispersed in a solvent by ultrasonication or high-shear mixing, resulting in graphene oxide (GO) dispersions (Fig. 12.5a). The phenol, carbonyl and epoxy groups resulting from the oxidation of graphite ensured colloidal stability in polar solvents [12.24]. Polymers, surfactants, DNA, etc. can be used to provide additional stabilization of the GO flakes in the colloidal suspensions in various solvents.

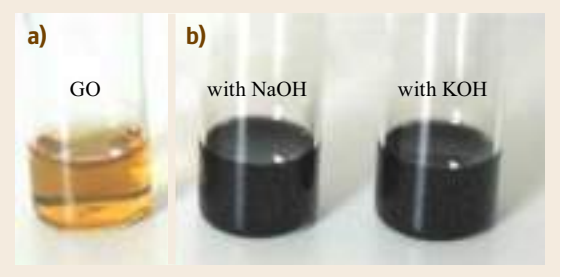


Fig. 12.5a,b Dispersion of (a) GO and (b) rGO reduced under alkaline conditions. Reprinted from [12.23], with permission

Exfoliated GO can be subsequently reduced to yield reduced (r) GO (Fig. 12.5b). It should not be referred to as graphene at this stage due to the incomplete nature of the reduction process. One route involves reduction in water with hydrazine hydrate [12.25, 26] or dimethylhydrazine [12.27]. A reduced GO suspension can also be obtained by heating the exfoliated GO suspension under strongly alkaline conditions [12.23]. Alternately, the GO flakes can be deposited on a substrate and reduced by hydrazine vapors or hydrogen plasma [12.28]. All these methods result in the formation of unsaturated and conjugated carbon atoms, which results in electrical conductivity and Raman signatures in between that of GO and pristine graphene.

Exfoliation Without Intercalation

In an effort to avoid disruption to the desirable structure and properties of graphene, efforts have been undertaken to directly exfoliate and disperse graphite in a solvent by ultrasonication and high-shear mixing. Certain solvents such as *N*-methylpyrolidone, *N*,*N*-dimethylamide (DMA), γ-buryrolacetone and 1,3-dimethyl-2-imidazolidinone are ideally suited to dissolve graphene in significant concentrations [12.29]. The procedure has been adopted successfully for using water as solvent in the presence of sodium dodecylbenzene sulfonate or sodium cholate as a stabilizing surfactant [12.30]. The predominant drawback of this process lies in the fact that the sonication breaks up the graphene into particularly small fragments, with the monolayer flake having lateral dimensions of, on average, 100 nm. This can be overcome to some extent by the use of high-shear mixing [12.31]. The aqueous graphene dispersion can now be processed by density gradient ultracentrifugation (DGU) using iodixanol as a density medium to yield fractions enriched in particular graphene thicknesses (Fig. 12.6) [12.32].

12.1.5 Synthetic Organic Chemistry Route

If the reduction of bulk graphite into graphene is viewed as a top-down approach, then the chemical synthesis of graphene from smaller aromatic hydrocarbons will constitute the bottom-up approach. If graphene is regarded as a polycyclic aromatic hydrocarbon (PAH), one the largest of these synthesized involves 222 carbon atoms or 37 benzene units in a hexagonal structure 3 nm in diameter [12.33] from an oligophenylene

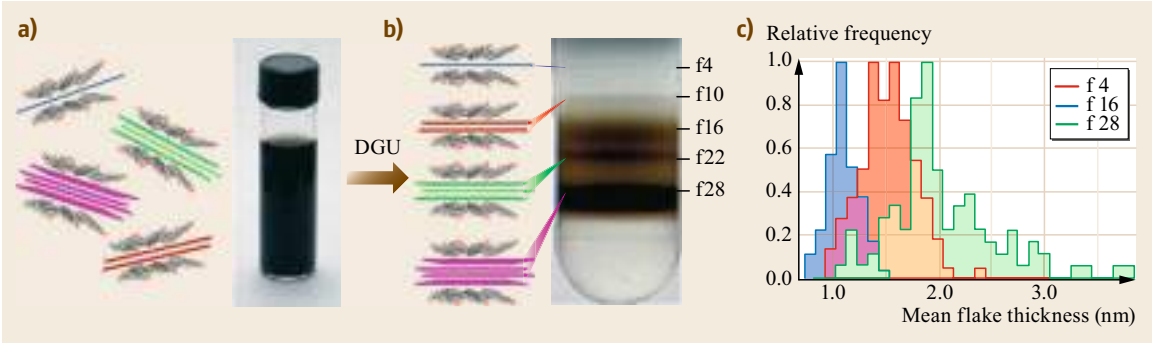


Fig. 12.6 **(a)** Schematic illustration of sodium cholate encapsulation of graphene sheets and a photograph of an unsorted aqueous graphene suspension with a graphene loading of ≈ 0.1 mg/ml. **(b)** Photograph of a centrifuge tube following DGU marked with the main bands of monodisperse graphene. **(c)** Mean flake thickness histogram of sorted graphene taken from the locations marked in **(b)**. Reprinted with permission from [12.32]. Copyright (2009) American Chemical Society

precursor, which was planarized by oxidative cyclohydrogenation. These structures have also shown a high tendency to self-assemble on surfaces [12.34] and could potentially act as precursors to larger synthetic graphene.

Very narrow ribbons of graphene with widths of tens of nanometers and with well-defined edge structure and orientation are of particular importance in electronic applications, since such nanoribbons of graphene are one of the means to engineer an electronic bandgap in otherwise gapless graphene. Graphene nanoribbons, 8–12 nm in length and 2–3 nm width, have also been synthesized by surface-assisted coupling of molecular precursors into linear polyphenylenes and their subsequent cyclohydrogenation (Fig. 12.7a) [12.35].

12.1.6 Unzipping of Carbon Nanotubes

CNTs have been described as rolled up graphene sheets, and now graphene nanoribbons have been made from unraveling CNTs. Oxidized nanoribbons were obtained by suspending CNTs in concentrated sulphuric acid followed by treatment with 500 wt% $KMnO_4$ for 1 hr at 22 °C and 1 hr at 55–70 °C. The process described as unzipping of a CNT could occur as a linear longitudinal cut or in a spiraling manner depending on the chirality of the CNT (Fig. 12.7b) [12.36]. CNTs have also been converted to graphene nanoribbons by controlled plasma-etching of CNTs that are partially embedded in a polymer film (Fig. 12.7c) [12.37].

12.1.7 Derivatives of Graphene

Graphane, a fully saturated hydrocarbon derived from graphene, with a formula CH, was predicted to be stable based on first-principle total-energy calculations [12.38]. Experimentally, it was later shown that graphene can be hydrogenated and converted to graphane using a low-pressure (0.1 mbar) hydrogen-argon mixture (10% H_2) with direct current (DC plasma for 2 hrs [12.39]. The hydrogenation is stable, but reversible and the graphane can be reconverted to

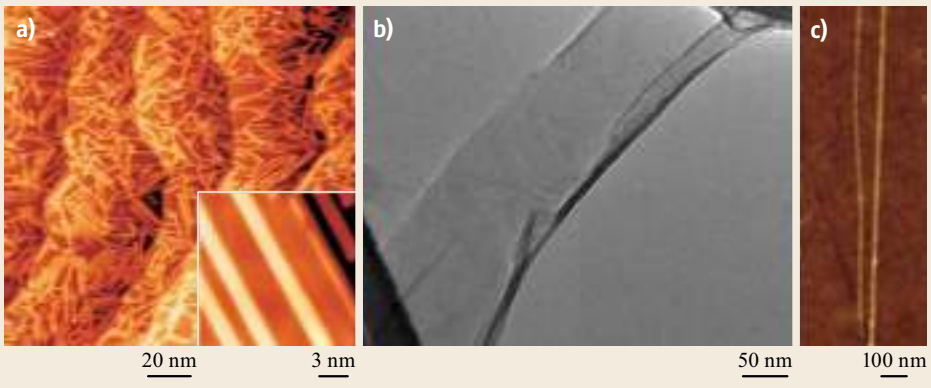


Fig. 12.7a–c Graphene nanoribbons formed by various means: **(a)** synthesized from polyphenylene precursors (after [12.35]), **(b)** unzipping of carbon nanotubes (after [12.36]) and **(c)** etching of carbon nanotubes embedded in a polymer (after [12.37])

graphene by annealing at 450 °C in Ar atmosphere for 24 hrs. The reconverted graphene, however, contains remnant defects just as vacancies and oxygenated or hydrogenated carbon atoms. In the case of substrate-supported graphene, only one side is hydrogenated, while both sides can be hydrogenated in the case of suspended graphene.

Graphene can also be fluorinated with XeF_2. When one side is exposed, F coverage saturates at 25% (C_4F), after which fluorination of both sides results in perfluorographene [12.40] and fluorographene [12.41], which are the nonstoichiometric and stoichiometric variations respectively. Hydrazine treatment has been shown to reverse the fluorination.

12.2 Properties

While its very existence as a freestanding 2-D material is a feather in graphene's cap, it is the properties of graphene that make it the truly exceptional material that has stoked feverish research in this field. The individual properties of pristine graphene will be discussed first, while the properties of graphene derivatives such as graphane and fluorographene will be discussed in the final section.

12.2.1 Structure and Physical Properties

Graphene shares most of its structure and physical properties with graphite, its parent material. The carbon atoms are arranged in a 2-D hexagonal lattice (Fig. 12.8), which can also be constructed as two interpenetrating triangular sublattices, which takes particular significance in bilayer and other multilayer graphenes. The carbon atoms are sp^2 hybridized, and the in-plane carbon–carbon bond length is $a = 1.42$ Å. The remaining p orbital is oriented perpendicular to the plane of carbon atoms and delocalized to form the π (valence) and π^* (conduction) electronic bands. The interplane spacing is 3.45 Å.

2-D structures like graphene had been postulated to be intrinsically unstable, and according to the Mermin–Wagner theorem [12.43], long-wavelength fluctuations destroy the long-range order of 2-D crystals. The puzzling stability of suspended 2-D graphene sheets has been attributed to intrinsic microscopic undulations in which the surface normal varies by several degrees and the out-of-plane deformation reaches 1 nm [12.44, 45]. Similar corrugation has also been reported in graphene supported on SiO_2 substrates [12.46]. Scrolling has been occasionally observed at the edges of graphene flakes, both suspended [12.47] and substrate supported [12.48], and this appears to rely on the fabrication method. Scrolling occurs when graphene is subjected to liquid-phase processing during microfabrication, while solid-phase or gas-phase processing appears to avoid this [12.42] (Fig. 12.9a). Graphene deposited on atomically flat terraces of cleaved mica surfaces are flat down to the atomic scale [12.49].

12.2.2 Mechanical Properties

Direct measurements of the elastic properties of graphene have been conducted by nanoindentation of suspended graphene layers in an AFM [12.51, 52]. For monolayer graphene, the force-displacement characteristics yield second- and third-order elastic stiffness

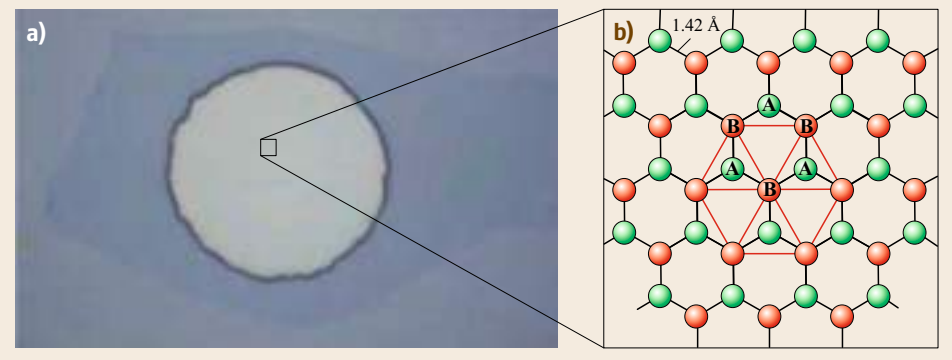


Fig. 12.8 **(a)** Optical micrograph of a graphene flake, partly suspended over a hole in a silicon substrate. Reprinted with permission from [12.42]. Copyright (2008) American Chemical Society. **(b)** Schematic of the atomic structure of graphene, showing the two triangular sublattices in green and orange as well as the position of the atoms in the second layer of bilayer graphene (A) with respect to atoms in the first layer (B)

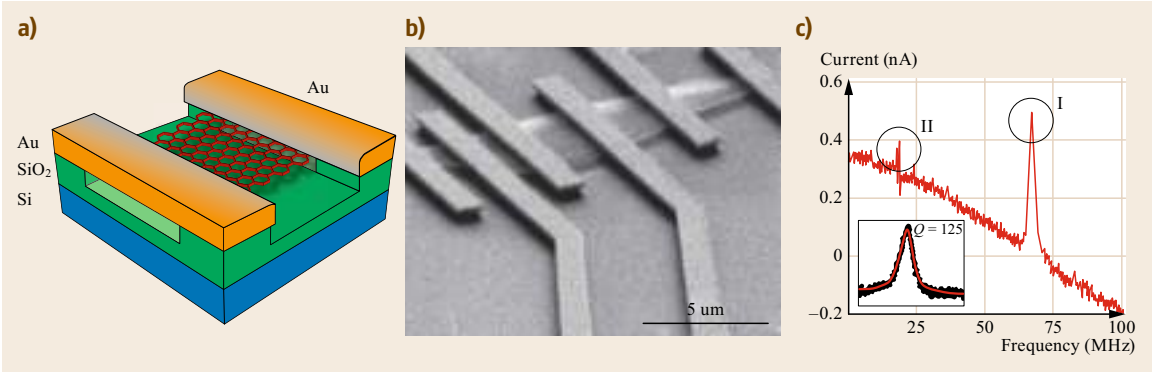


Fig. 12.9 **(a)** Schematic of a graphene resonator, with electrostatic actuation and electrical readout. **(b)** SEM image of such a resonator. **(c)** The graphene resonance (I) at 65 MHz. Resonances of metal beams (II) are also visible below 25 MHz. *Inset*: the graphene resonance at low driving power, and Lorentzian fit (*red line*) with $Q = 125$ (after [12.50])

of 340 Nm^{-1} and 690 Nm^{-1} respectively. The breaking strength is 42 Nm^{-1}, which represents the intrinsic strength of a defect-free sheet. This corresponds to a Young's modulus of $E = 1.0$ TPa, third-order elastic stiffness of 2.0 TPa and an intrinsic strength of 130 GPa. The figures mean that graphene is the strongest material ever measured.

Nonlinear finite elasticity theory for graphene resonators for both electrostatic and electrodynamic cases has been developed and agrees well with experiments on graphene resonators [12.53]. In general, natural resonance frequencies in the GHz regime are expected for graphene resonators, due to the combination of their extreme thinness and extraordinary stiffness. Experimentally, the mechanical vibrations in electrostatically actuated graphene resonators have been imaged by a special modification of atomic-force microscopy [12.54]. Resonance frequencies in the 10 s of MHz have been recorded in graphene resonators (Fig. 12.9), with quality factors of 4000 at room temperature [12.55, 56] and 10 000 at 5 K [12.50].

12.2.3 Electronic Properties

Monolayer, bilayer and trilayer graphenes are electronically distinct materials. Beyond three layers, graphene's electronic properties tend towards that of bulk graphite.

Monolayer Graphene

The electronic structure of graphene was first described in 1946 [12.28] as a theoretical building block to describe graphite. The valence and conduction bands of graphene are conical valleys that touch at the high-symmetry K and K′ points of the Brillouin zone. Near these points, the energy varies linearly with the magnitude of momentum, i. e., it follows a linear dispersion relation (Fig. 12.10a). In neutral graphene, this point of intersection coincides with the charge neutrality point, and is referred to as the Dirac point.

In 3-D electron systems, the electrons' behavior can be described by the Schrödinger equation. In graphene, on the other hand, electrons behave as relativistic particles, and are described by the Dirac equation [12.58, 59]. The interaction of electrons with the periodic potential of the graphene hexagonal lattice results in quasiparticles, which can be viewed as electrons devoid of their rest mass m_0 and therefore called massless Dirac fermions. The linear energy dispersion means that the speed of electrons in graphene is a constant, independent of momentum, as in the case of the speed of photons. The velocity of electrons in graphene is $\approx 10^6$ m/s, about 300 times slower than the speed of light (photons).

The electronic states near the Dirac point are composed of states belonging to the two graphene sublattices, and as a result the quasiparticles possess pseudospins, similar to the electron's spin [12.60]. These Dirac fermions are said to be chiral. Another relativistic phenomenon associated with these quasiparticles is the Klein paradox [12.60] wherein they tunnel through a potential barrier of any height and width with a transmission probability of 1 or without a reflected component. As a result, electrons in graphene can propagate over (relatively) vast distances on the order of microns through the graphene lattice, even in the presence of lattice defects or other external perturbing potentials.

Bilayer Graphene

In bilayer graphene, the AB (Bernal) stacking between the two graphene renders the two carbon atoms inequivalent and results in two graphene sublattices. As a result the unit-cell of bilayer graphene contains four atoms, and two additional bands in the π and π* states. If the inversion symmetry between the two layers is

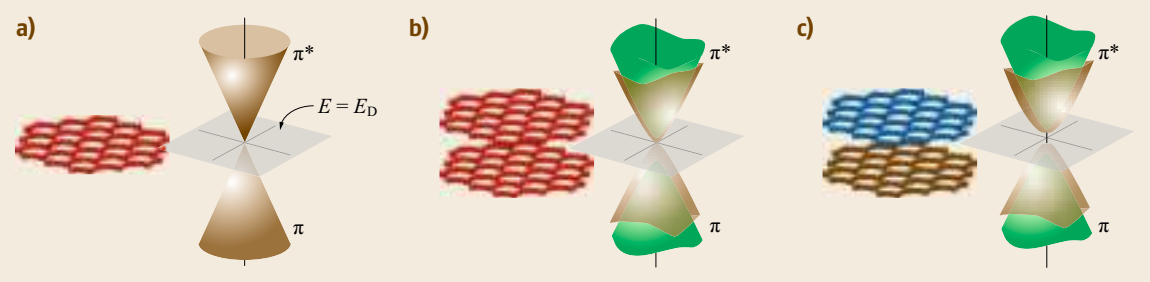


Fig. 12.10a–c Electronic structure of **(a)** monolayer, **(b)** symmetric bilayer, and **(c)** asymmetric bilayer graphene (after [12.57])

broken, then an energy gap between the low-energy valence and conduction bands forms at the Dirac point (Fig. 12.10b,c)

The electronic gap in bilayer graphene can be controlled by applying an external transverse electric field, such as by a gate bias, making it the only known semiconductor material with a tunable energy gap. Using a tight-binding model, the value of the gap was extracted as a function of electron density, and it was shown that it can be tuned to values larger than 0.2 eV.

12.2.4 Optical Properties

The successful exfoliation on monolayer graphene owes itself to recognizing the optical properties of graphene as much as the exfoliation procedure [12.5]. On a SiO_2 surface of a Si wafer, the contrast of a graphene flake depends on the SiO_2 thickness as well as the wavelength of light. Fig. 12.11 summarizes the expected contrast as a function of SiO_2 thickness as well as wavelength of monochromatic illumination. The observed contrast allows the determination of the number of graphene layers.

The absorbance of light by monolayer and bilayer graphene is 2.3 and 4.6% respectively in the visible spectrum (450–750 nm), and this linearity extends up to five layers. The optical transparency of noninteracting graphene is solely determined by the fine structure constant of quantum electrodynamics ($\alpha = e^2/\hbar c = 1/137$), which describes the coupling between light and relativistic electrons [12.61].

12.2.5 Thermal and Thermoelectric Properties

The thermal conductivity of suspended monolayer graphene is 5300 W/mK with a phonon mean free path of 775 nm near room temperature [12.62]. Interestingly, this value is higher than the bulk graphite limit of $K \approx 2000$ W/mK. The room-temperature thermal conductivity decreases from ≈ 2800 to ≈ 1300 W/mK as the number of graphene layers in few-layer graphene

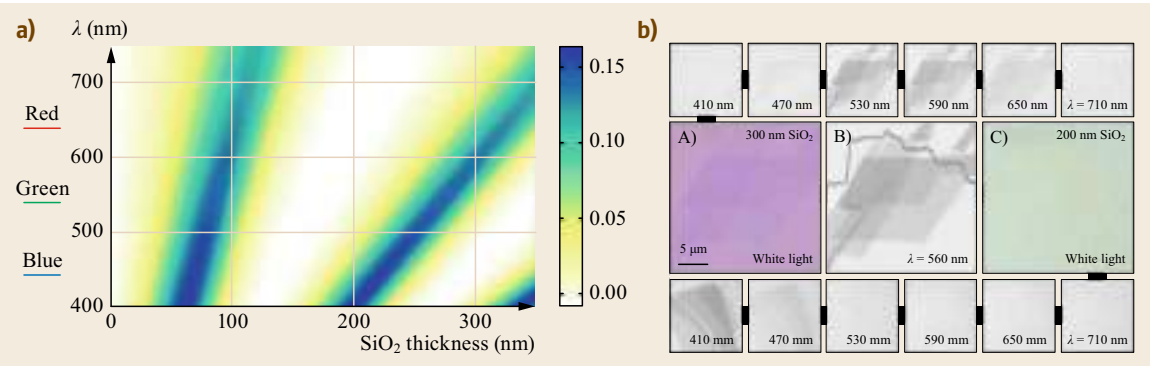


Fig. 12.11 **(a)** Color plot of the contrast as a function of wavelength and SiO_2 thickness. The color scale on the right shows the expected contrast. **(b)** Graphene crystallites on 300 nm SiO_2 imaged with white light (A), green light (B) and another graphene sample on 200 nm SiO_2 imaged with white light (C). Single-layer graphene is clearly visible on the left image (A), but even three layers are indiscernible on the right (C). Image sizes are $25 \times 25\,\mu m^2$. Top and bottom panels show the same flakes as in (A) and (C) respectively, but illuminated through various narrow bandpass filters with a bandwidth of ≈ 10 nm. Reprinted from [12.5], with the permission of AIP Publishing

(FLG) increases from two to four [12.63], corresponding to the evolution from 2-D graphene to bulk graphite.

Thermoelectric power (TEP) is the voltage developed across a sample when a constant temperature gradient is applied. TEP of graphene is $80\,\mu V/K$ at 300 K [12.64]. Similar to the quantum Hall effect in electronic transport, quantized TEP have also been observed in graphene at high magnetic fields. The TEP can be tuned in graphene, even to negative values, under the application of a gate bias or chemical potential [12.65]. Very large TEP values have been predicted for graphene nanoribbons, for instance, $4\,mV/K$ for a 1.6 nm-wide ribbon [12.66, 67]. A giant thermoelectric coefficient of $30\,mV/K$ has been observed in a nanostructure consisting of metallic electrodes periodically patterned over graphene, which is deposited on a SiO_2 substrate [12.68].

12.2.6 Chemical Properties

The chemistry of graphene is dominated entirely by its surface, since every carbon atom is a surface atom twice over. For graphene nanoribbons, the edges play a role in determining their reactivity.

It has been shown that for electron-transfer chemistries, single graphene sheets are almost 10 times more reactive than bilayer or multilayer graphene according to the relative intensity of the disorder (*D*) peak in the Raman spectrum examined before and after chemical reaction [12.69, 70]. Graphene can be readily functionalized through diazonium or nitrene [12.71] reactions, which can introduce reactive species covalently linked to graphene. These groups then serve as templates for further chemistry and grafting of functional groups, for instance through an azide linker. Chemical functionalization of graphene be monitored through its effect on the conductivity of graphene, and serve as a means of controlling the electrical transport properties of graphene [12.72]. Furthermore, p-doped and n-doped regions in graphene can be generated by suitably functionalizing them, for instance with diazonium salts and polyethyleneimine respectively [12.73]. Such chemical modification can also be performed on different parts of a single sheet to form p-n junctions in graphene [12.74].

12.2.7 Properties of Graphene Derivatives

Graphane has been theoretically predicted to take one of two configurations, a *chair* conformer with the hydrogen atoms alternating on both sides of the plane for the two graphene sublattices, and a *boat* conformer with the hydrogen atoms alternating in pairs [12.38]. The *chair* and *boat* conformers have a direct electronic bandgap of 3.5 eV and 3.7 eV respectively. Graphane is a completely insulating material; its resistivity changes by two orders of magnitude with decreasing temperature from 300 K to 4 K and carrier mobility decreases to $\approx 10\,cm^2V^{-1}s^{-1}$ at liquid-helium temperatures for typical carrier concentrations of $10^{12}\,cm^{-2}$ [12.39].

Similarly, fluorographene [12.41] is a high-quality insulator with a large optical bandgap of > 3 eV and room-temperature resistivity of $> 10^{12}\,\Omega$. The Young's modulus of fluorographene is 0.3 TPa, which is about 30% the stiffness of graphene. Similarly, fluorination reduces graphene's intrinsic breaking strength by 2.5 times. However, fluorographene is able to sustain the same ultimate strain of 15% as graphene. Fluorographene is also strongly hydrophobic, and can be considered the 2-D equivalent of Teflon.

12.3 Characterization

Each of the properties discussed in the previous section has to be measured and correlated using multiple characterization techniques. This is essential, since just one measurement, for instance electronic transport, might not be able to sufficiently distinguish between an electronic bandgap and a mobility gap. Similarly, mechanical properties have to be confirmed by a combination of tensile testing, electromechanical resonance and Raman spectroscopy.

12.3.1 Optical Microscopy

Based on the optical properties of graphene discussed in an earlier section, and the fact that green light is most comfortable for the eyes, optimal SiO_2 thicknesses of 90 and 280 nm can be recommended [12.5]. Similarly, it has been shown that graphene can be observed on 50 nm Si_3N_4 using blue light and on 90 nm poly(methyl methacrylate) (PMMA) using white light. Rayleigh scattering can identify the number of graphene layers as well as probe their dielectric constant [12.75].

12.3.2 Transmission Electron Microscopy

High-resolution transmission electron microscopy (TEM) can resolve individual carbon atoms as well as adatoms, defects and other anomalies in graphene (Fig. 12.12a) [12.47]. If folds occur in the transferred

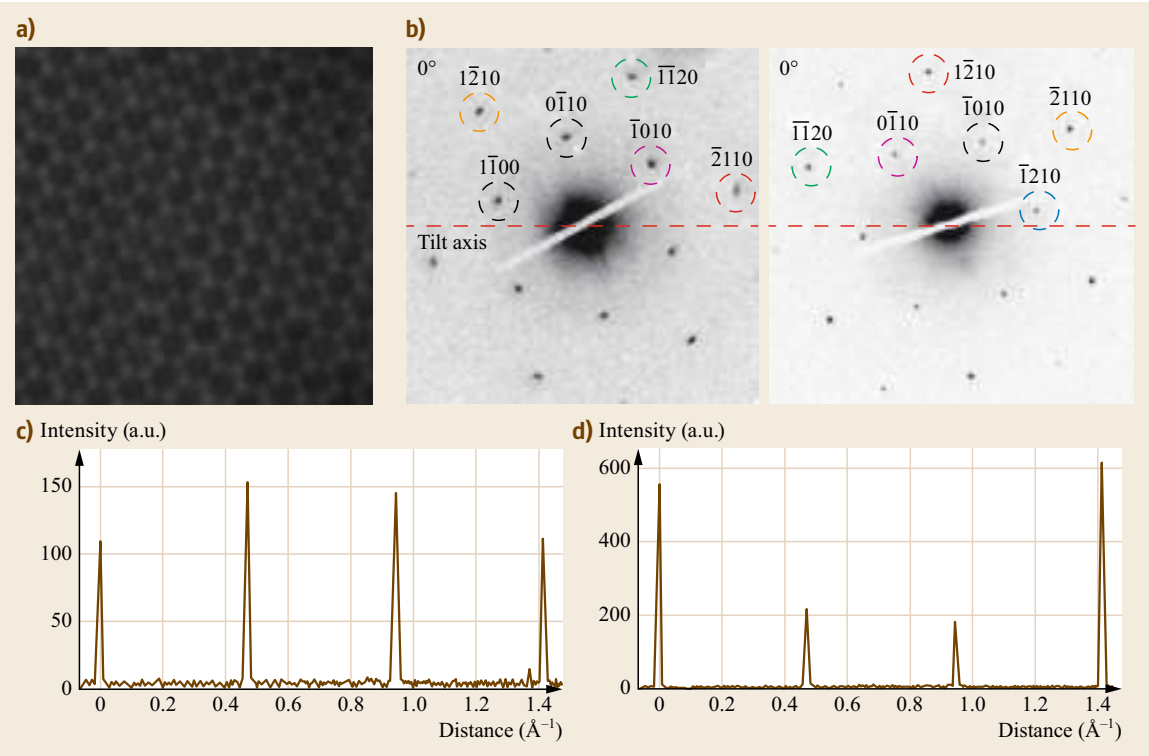


Fig. 12.12 **(a)** Atomic resolution transmission electron micrograph of graphene. **(b)** Nanobeam electron-diffraction patterns of monolayer and bilayer graphene. Relative intensities of 1-100 and 1-210 peaks in **(c)** monolayer and **(d)** bilayer graphene. Reprinted from [12.76] with permission from Elsevier

graphene flake, then observation of the folded edge can yield information about the number of layers in the graphene flake [12.45]; a monolayer fold-edge turns up as a single dark like while a bilayer fold-edge appears as two dark lines and so on. In addition, nanobeam electron diffraction (Fig. 12.12b) can also be used to quantify the layering in graphene [12.45]. Monolayer graphene can be distinguished from higher layered graphenes by the anomalous intensity ratio of the diffraction peaks; it's 0-110 peaks being more intense than the 1-210 peaks (Fig. 12.12c). The total (integrated) intensity of the 0-110 and 1-210 peaks of monolayer graphene varies weakly with tilt angle and no minima in intensity is observed. In contrast, the total intensity of bilayer graphene diffraction peaks varies strongly with tilt angle, including minima at certain angles where some peaks vanish [12.77]. However, while the total intensity in monolayer graphene only decreases slightly, significant peak broadening is observed with increasing tilt angle. This effect is most pronounced in monolayers, and decreases with increasing thickness of the graphene flake. This is attributed to nanoscale corrugations in 2-D graphene, with the surface normal deviating on average by $\pm 5°$ in monolayers and $\pm 2°$ in bilayers. This nanoscale corrugation extending into the third dimension is the reason for the existence of 2-D graphene considering the theoretical prediction that perfect 2-D atomic crystals cannot exist.

12.3.3 Scanning Probe Techniques

Atomic force microscopy (AFM) in tapping mode is commonly used to measure the thickness of graphene flakes on substrates; however, the correlation between measured thickness and actual thickness as well as number of layers is challenging [12.78]. Electrostatic interactions between the tip and the graphene, adsorbed moisture, and incorrect choice of AFM parameters such as free amplitude values can all influence the final measured thickness of a graphene flake. In addition, AFM is used to measure the flatness of graphene on various substrates. AFM can also be used in nanoindentation mode to probe the stiffness of suspended graphene [12.51].

Scanning tunneling microscopy (STM) can image graphene with atomic resolution, and the correlation of the graphene hexagonal lattice to the direction of the edge of an exfoliated flake reveals the orientation of

the edge as being either armchair or zigzag [12.79]. Scanning tunneling spectroscopy (STS) can be used to probe the atomically resolved local electronic structure of graphene [12.80, 81].

12.3.4 Angle-Resolved Photoemission Spectroscopy (ARPES)

ARPES is a direct experimental technique that has measured the electronic density of states in graphene with both energy and momentum information. The shape of the π and π* bands near E_F at the K point from ARPES reveals the transition from 2-D to bulk character for one to four layers of graphene [12.57], as well as the electronic band structures of bilayer graphene [12.57]. Fermi velocities and effective masses of the electrons can also be measured. ARPES also allows studies of electron–electron, electron–phonon and electron–plasmon interactions, and indicates that all three must be considered on an equal footing to understand the quasiparticle dynamics in graphene [12.82].

12.3.5 Raman Spectroscopy

The three significant Raman spectral features in graphene are the G peak at $\approx 1580\,\mathrm{cm}^{-1}$, the D peak at $\approx 1350\,\mathrm{cm}^{-1}$ and the 2D peak at $\approx 2700\,\mathrm{cm}^{-1}$, as seen in Fig. 12.13 [12.83]. The G peak is due to the E_{2g} mode, i. e., the in-plane vibrations of the carbon atoms. The D peak and 2D peak are strongly dispersive with excitation energy due to the Kohn anomaly at the K point, while the G peak is not.

The 2D peak is the second-order of the zone-boundary phonons and therefore does not require defects. For monolayer and few-layer graphene, the 2D peak serves as a fingerprint for identification. In general, the 2D peak of graphite has four components, $2D_{1B}$, $2D_{1A}$, $2D_{2A}$, and $2D_{2B}$. Monolayer graphene has a single sharp 2D peak, dominated by the $2D_{1A}$ component. In bilayer graphene, the $2D_{1A}$ and $2D_{2A}$ peaks have higher relative intensity compared to the other two, and the 2D peak appears upshifted and broader compared to monolayer graphene. It is also important to note here that non-AB stacked graphene, such as multilayer CVD graphene, also shows a single 2D peak [12.84]. However, this can be distinguished from monolayer graphene by its full-width at half-maximum (FWHM) of $50\,\mathrm{cm}^{-1}$, which is twice that of monolayer graphene. A similar observation can also be made with the D peak [12.83]. The D peak of monolayer graphene is a single sharp peak while in bulk graphite it can be resolved into two peaks, D_1 and D_2. The D peak is observed in defective graphene, and prominently at the edges of graphene flakes. A D′ peak at $\approx 1620\,\mathrm{cm}^{-1}$ is also observed in defective graphene.

The Raman spectrum of graphene also responds to doping, i. e., changes in the Fermi surface of graphene [12.85, 86]. The G peak upshifts for both hole and electron doping. The position of the 2D peak however decreases monotonically with increasing electron concentration or decreasing hole concentration, and the 2D peak can be used to distinguish between electron and hole doping. Taken together, a plot of 2D versus G peak positions can be used to distinguish between electron and hole doping in graphene, as seen in Fig. 12.14a [12.87].

Raman peaks in graphene can also be shifted by biaxial strain, induced for example by interaction with the

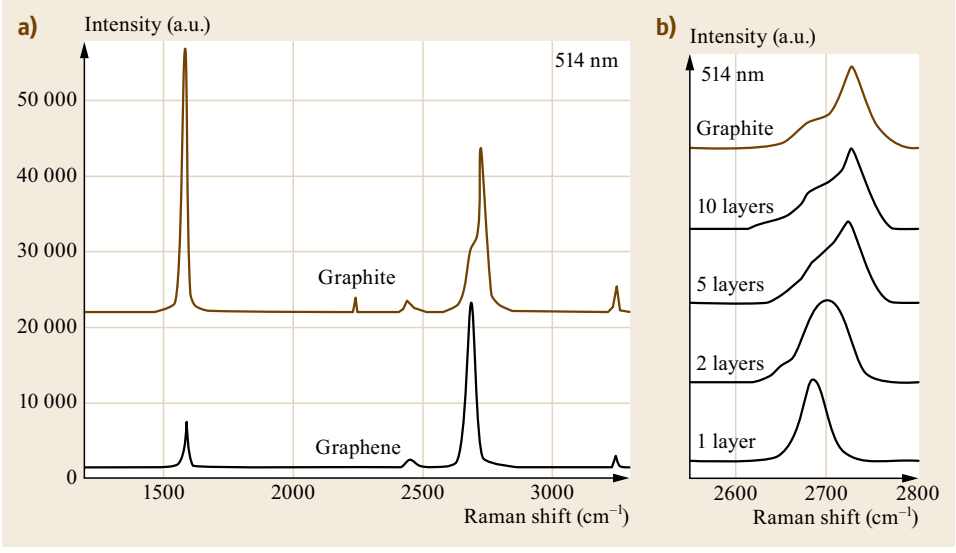


Fig. 12.13 **(a)** Comparison between Raman spectra of graphene and graphite. **(b)** Evolution of 2D peak shape with number of layers of AB stacked graphene. **(c)** Comparison of D peak between graphene and graphite (after [12.83])

substrate [12.88, 89]. The correlation between normalized shift of 2D and G peak positions, i. e., $\Delta 2D/2D_0$ and $\Delta G/G_0$, where $2D_0$ and G_0 are the peak positions for undoped graphene, indicates whether the shift arises from doping or strain [12.90]. When $\Delta 2D/2D_0$ versus $\Delta G/G_0$ is plotted from Raman spectra obtained at a number of points on a graphene sample, a linear fit with slope close to 1.58 ± 0.18 indicates that strain plays a predominant role, while a smaller slope indicates increasing influence of doping. Highly doped samples yield a slope of < 1. In another approach, if the plot of FWHM versus position of the G peak is monotonically decreasing, it is related to doping, while if it is monotonically increasing, it is caused by strain or disorder (Fig. 12.14b) [12.86]. Under uniaxial strain, the G peak splits into two bands, G^+ and G^-, analogous to the effect of curvature on the G peak of carbon nanotubes [12.91].

12.3.6 Electrical Characterization

The sheet resistivity ρ of graphene varies with applied gate-voltage, exhibiting a sharp peak of several kΩ and decaying to $\approx 100\,\Omega$ at high V_g (Fig. 12.15). The conductivity $\sigma = 1/\rho$ increases linearly with V_g on either side of the resistivity peak (conductivity valley). The Hall effect coefficient R_H reverses sign at the same V_G as the resistivity peak. This resembles an ambipolar semiconducting field-effect transistor, except that there is no zero-conductance region since there is no bandgap. This tunability of conductivity is achieved by transforming the graphene by electric-field doping from completely electron or completely hole conductor, passing through a mixed state where both electrons and holes contribute equally. This behavior holds true for monolayer graphene as well as undoped bilayer graphene. Also significant is the minimum conductivity of graphene at the charge neutrality point (σ_{min}), which has been shown to be about $4e^2/h$. This σ_{min} is not related to the physics of the Dirac point singularity, but instead related to the charge-density inhomogeneities (electron–hole puddles) induced by the substrate or charged impurities [12.92, 93]. Invariably, the position of the charge-neutrality point (ρ peak) is shifted to large positive V_G, leaving the ungated graphene as a hole-metal. This large shift is attributed to unintentional doping of the graphene by adsorbed species such as water. The peak position can also be shifted by intentional doping or removal of dopants by thermal or current-induced annealing [12.94] but such charged impurities do not affect σ_{min}. The mobility and minimum conductivity also decrease as a result of defects, which can be induced in a controlled fashion to study this effect, for instance by ion irradiation [12.95] or exposure to atomic hydrogen [12.96].

The carrier mobility in graphene on SiO_2 substrates is $\approx 10\,000\,cm^2/Vs$, which is independent of the absolute temperature, indicating that it is limited by scattering defects. Substrate-induced charge puddles are significantly reduced in suspended graphene and low-temperature mobility approaching $200\,000\,cm^2/Vs$ has been reported [12.97, 98]. In such devices, the con-

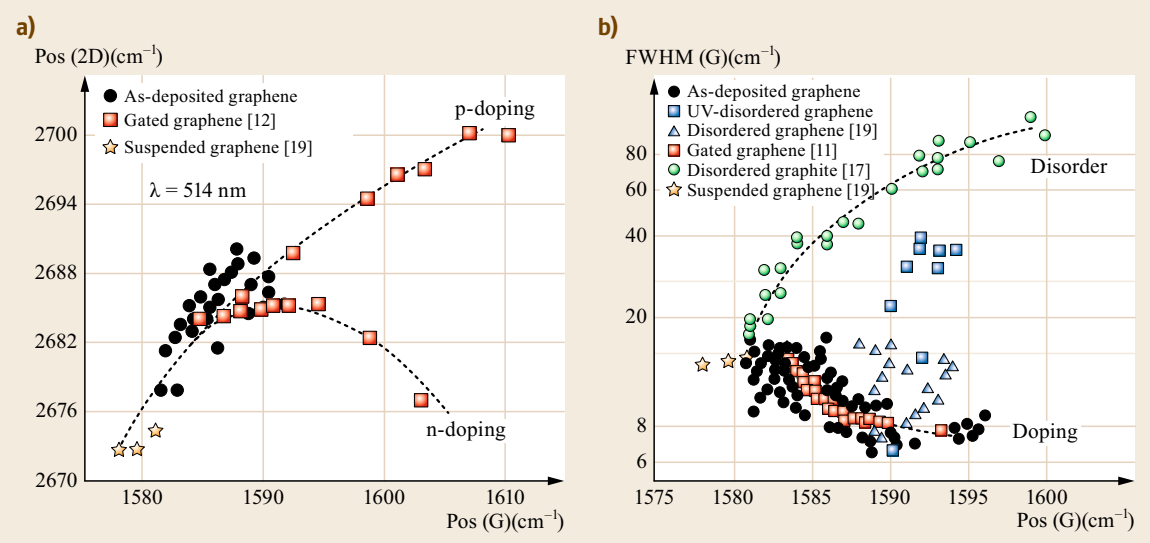


Fig. 12.14 **(a)** Pos(G) as a function of Pos(2D) for as-deposited graphene and gated graphene. **(b)** Pos(G) as a function of FWHM(G) for as-deposited graphene, compared with disordered graphene, disordered graphite and gated graphene. The *dotted lines* are only a guide for the eyes (after [12.87])

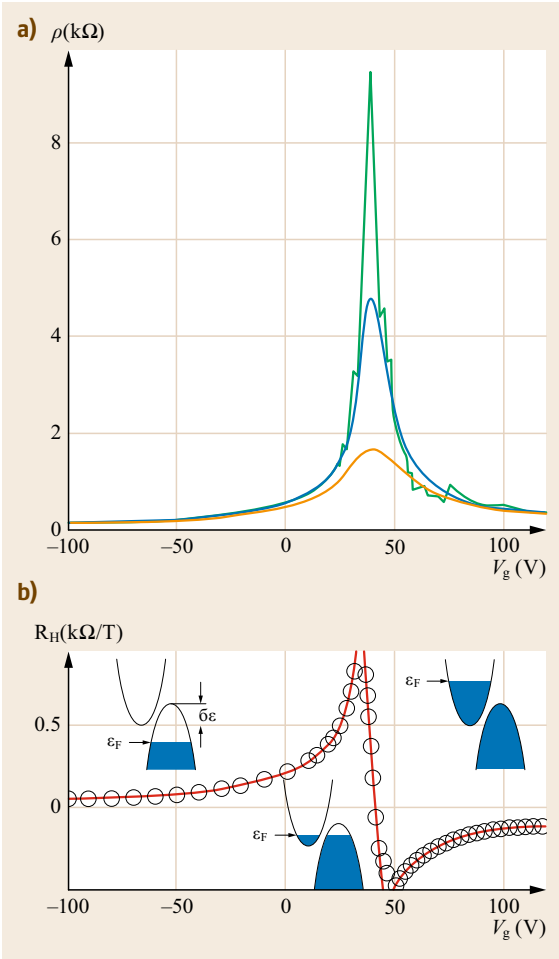


Fig. 12.15a,b Field effect in graphene. **(a)** Typical dependences of graphene's resistivity ρ on gate voltage for different temperatures ($T = 5$, 70 and 300 K for top to bottom curves respectively). **(b)** Hall coefficient R_H versus V_g for the same film at $T = 5$ K (after [12.4])

ductivity of suspended graphene at the Dirac point is strongly temperature dependent and approaches ballistic values at liquid helium temperatures. Due to the linear dispersion relation in graphene, intrinsic electron scattering by acoustic phonons is independent of carrier density and only contributed 30 Ω to the room-temperature resistivity of graphene. This would yield an intrinsic mobility of $2 \times 10^5\,\mathrm{cm}^2/\mathrm{Vs}$ at a carrier density of $1 \times 10^2\,/\mathrm{cm}^2$. However, a strong temperature dependence of mobility is observed in substrate-supported graphene devices, suggestive of extrinsic scattering, which limits the mobility to about $4 \times 10^4\,\mathrm{cm}^2/\mathrm{Vs}$ [12.99].

Graphene also exhibits pronounced Shubnikov–de Haas (ShdH) oscillations in both longitudinal resistivity ρ_{xx} and Hall resistivity ρ_{xy}. The oscillations depend only on the perpendicular component of the magnetic field $B\cos\Theta$, where Θ is the angle between the magnetic field and the graphene. The ShdH oscillations' frequency B_F depends linearly on V_G, indicating that the Fermi energies ε_F of holes and electrons are proportional to their concentration n. This is different from the 3-D dependence ε_F proportional to $n^{2/3}$, proving the 2-D nature of charge carriers in graphene.

The quantum Hall effect (QHE) has been observed in graphene even at room temperature, since the electrons suffer little scattering due to their relativistic nature and have a large cyclotron gap that exceeds the thermal energy $k_B T$ by a factor of 10. Fractional (F) QHE was reported in suspended graphene devices when a plateau at filling factor $\nu = 1/3$ was observed above a magnetic field as low as 2T [12.100, 101]. An insulating state was also observed at magnetic fields $B > 5$ T and filling factors $\nu < 0.15$, due to symmetry breaking of the zeroth Landau level by electron–electron interaction.

Under applied magnetic field, a plateau at zero Hall conductivity $\sigma_{xy} = 0$ occurs in biased bilayer graphene, as a result of the gap opened between the valence and conduction bands. Plateaus at $\sigma_{xy} = 4Ne^2/h$ occur as expected, including at $N = 0$. This is the standard integer QHE that is expected for an ambipolar semiconductor with an energy gap larger than the cyclotron energy. This plateau is absent in monolayer and unbiased bilayer graphene, which show an anomalous double-step of $8e^2/h$ at $n = 0$, indicative of the metallic state at the charge neutrality point. A huge peak in the longitudinal resistivity ρ_{xx} at $n = 0$ was also observed, exceeding 150 kΩ at 4 K compared to 6 kΩ for the unbiased case. From ShdH measurements, the cyclotron mass, m_c in biased bilayer graphene is found to be an asymmetric function of the carrier density, which is a clear signature of a bandgap. Measurements with dual-gated bilayer graphene have also confirmed these results [12.102].

12.4 Applications

While the field of experimental graphene research is still in its early years, graphene's potential for applications is already evident. Graphene is currently under evaluation for a wide range of applications, which will benefit from its unique combination of excellent mechanical, electronic, optical and chemical

properties. Commercial viability of most applications is severely limited by limited options available for large-scale production of graphene of appropriate quality and cost. Figure 12.16 summarizes the various methods of graphene production, the corresponding quality, scalability and cost, and potential applications for which each production method is appropriate.

12.4.1 Structural and Electrical Composites

The remarkable mechanical properties of graphene coupled with high electrical and thermal conductivity suggests that they would be ideal in composite materials for structural as well as electronic applications. GO and polystyrene composites show a low percolation threshold of ≈ 0.1 volume percent for room-temperature electrical conductivity and $\approx 0.1\,\mathrm{S/m}$ at 1 vol.%, which is sufficient for many electrical applications [12.27]. Such composites in thin-film form can show ambipolar field-effect behavior with electron and hole mobilities of 0.2 and $0.7\,\mathrm{cm^2/Vs}$ respectively [12.105]. GO-silica composites have been fabricated by spin coating GO-impregnated silica sols to yield transparent conducting composite films [12.106]. Layered GO-polymer composites have been fabricated by layer-by-layer (LBL) assembly, taking advantage of electrostatic interactions between anionic GO sheets and cationic polymer [12.56]. GO-P3HT/P3OT composites have been demonstrated as a superior acceptor layer in organic photovoltaics [12.107] while TiO_2-graphene composite has been demonstrated as a high-performance photocatalyst [12.108]. Composites of functionalized graphene derived from GO with polymers such as PMMA, poly(acrylonitrile) and poly(acrylicacid) show improvement of Young's modulus, glass transition temperature, ultimate strength and thermal degradation [12.109]. Composites of graphene with metal nanoparticles (Au, Pt, and Pd) have also been produced and the graphene-Pt composite has been demonstrated as a catalyst in direct-methanol fuel cells [12.110].

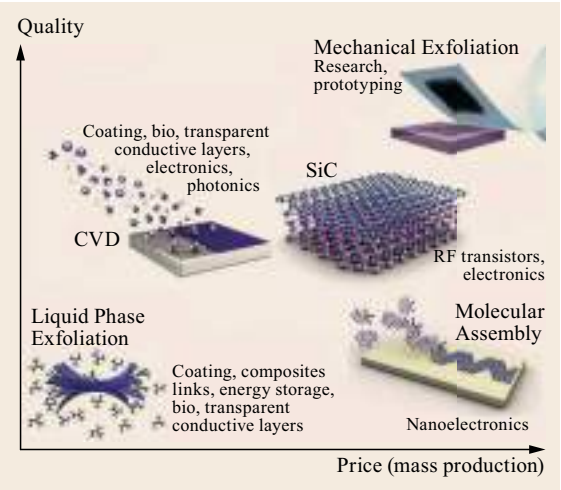


Fig. 12.16 Schematic representation of the various methods of graphene production as a function of the resulting quality and cost, and potential applications for which each production method is appropriate (after [12.103])

12.4.2 Transparent Conductive Films

Thin films have been deposited from solution-phase graphene with controllable thickness from a single monolayer to several layers over large areas (Fig. 12.17) [12.104]. The thickness of the film determines its conductivity, transparency and mechanical integrity. Electronic properties can be tuned from graphene-like ambipolar field-effect behavior for the thinnest films to graphite-like semimetal conductiv-

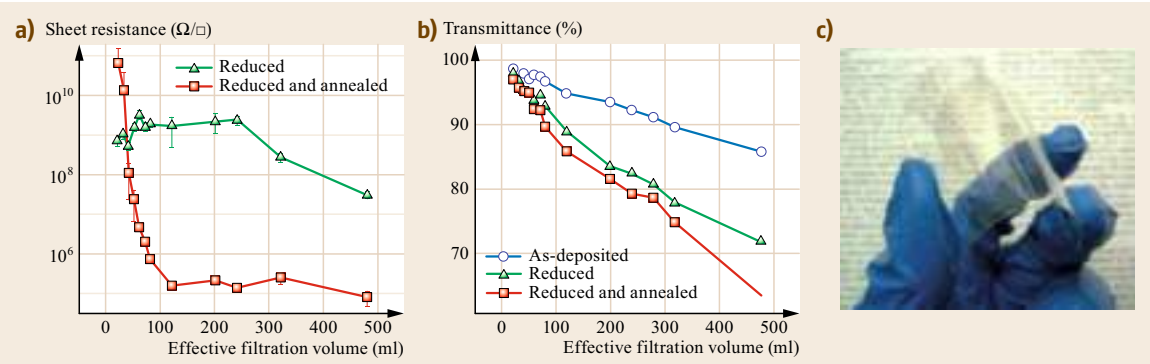


Fig. 12.17a–c Electrical and optical properties of rGO thin films. Sheet resistance **(a)** and transmittance **(b)** at $\lambda = 550\,\mathrm{nm}$ as a function of filtration volume for rGO thin films. The sheet resistance for as-deposited GO thin films is out of scale. **(c)** Photograph of GO thin films on filtration on a plastic substrate (after [12.104])

ity for thicker films. Flow-based deposition, vacuum filtration and layer-by-layer assembly are among the techniques used to deposit graphene thin films from solution.

Graphene conducting thin films are targeted as replacements for indium-tin oxide (ITO), which possesses a combination of 100 Ω per square sheet resistance at 90% transparency. Solution-processed rGO films [12.111] can achieve a sheet resistance of 1 KΩ per square at 90% transparency and have been demonstrated as transparent conductive anodes for inorganic, organic and dye-sensitized photovoltaic cells, but the short-circuit current and fill factor are still lower than that of control devices with ITO [12.112, 113]. However, they do have advantages such as tunable wettability and high chemical and thermal stability and flexibility.

Pristine graphene films grown by CVD on Ni or Cu substrates and transferred by a wet process can achieve ≈ 280 Ω per square at ≈ 80% transparency. The films show high electron mobility of 3700 $cm^2/V.s$ and exhibit the half-integer quantum-Hall effect typical of high-quality graphene, which cannot be accomplished with reduced or functionalized GO. Most importantly, the graphene films can recover their original resistance after being stretched by up to 6% and the resistance shows little variation up to a bending radius of 2.3 mm, corresponding to a tensile strain of 6.5%. This gives graphene films a significant advantage over ITO for applications in flexible electronics [12.10, 115].

A roll-to-roll dry transfer process has been developed for up to 35 inch monolayer graphene grown on Cu has yielded films with 97.4% transparency and 125 Ω per square sheet resistance (Fig. 12.18). The conductivity of graphene can be further improved by doping, for example p-doping with HNO_3, and resistance as low as 30 Ω per square can be achieved at 90% transparency, making graphene a strong potential candidate for ITO replacement, particularly in flexible electronics [12.114].

Solution-processed graphene as well as CVD graphene have been demonstrated as electrodes for organic

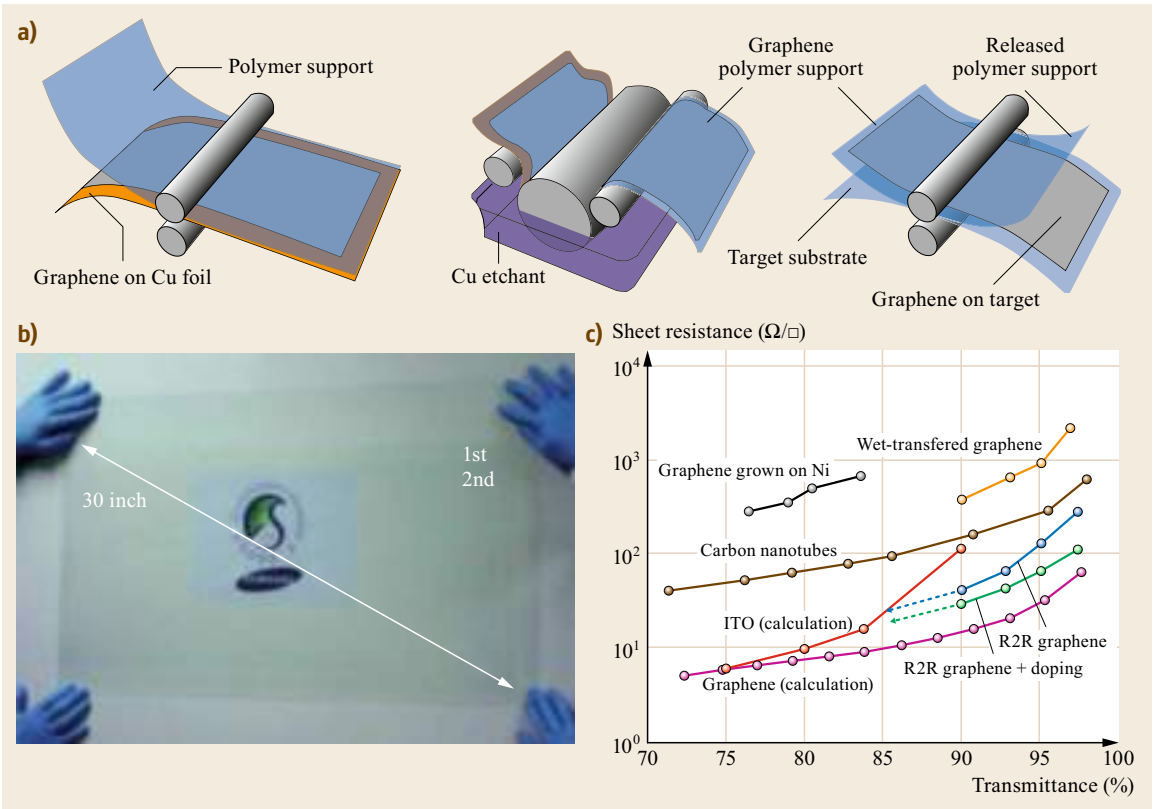


Fig. 12.18 **(a)** Schematic of the roll-based production of graphene films grown on a copper foil. The process includes adhesion of polymer supports, copper etching (rinsing) and dry transfer-printing on a target substrate. **(b)** A transparent ultralarge-area graphene film transferred on a 35-inch polyethylene terephthalate (PET) sheet. **(c)** Comparison of sheet resistance for graphene. The *dashed arrows* indicate the expected sheet resistances at lower transmittance (after [12.114])

light-emitting diodes [12.116]. This, combined with a graphene-based field-effect transistor and a graphene transparent conducting electrode allows for the fabrication of fully flexible and transparent touchscreen displays.

12.4.3 Sensors

The ultimate detection limit of any chemical sensor is the ability to detect individual molecules of a desired species (Fig. 12.19). The exceptionally low-noise electronic characteristics of graphene allows for such single-molecule sensitivity [12.117]. Adsorbed gas molecules change the local carrier concentration one electron at a time, which can be detected as stepwise changes in device resistance. Depending on the nature of the adsorbed species, the graphene can be either electron doped (NH_3) or hole doped (H_2O, NO_2). Graphene sensors are expected to be more sensitive than conventional semiconductor sensors.

rGO-based electronic device sensors have shown parts-per-billion sensitivity to chemical warfare agents and explosives [12.118] and parts-per-million sensitivity to gases such as NO_2 and NH_3 [12.119]. rGO-based surface acoustic wave sensors have shown parts-per-million sensitivity to H_2 and CO [12.120]. Electrolyte-gated graphene devices have been demonstrated as pH and biomolecule sensors with picomolar sensitivity [12.121].

Molecular structural mechanics modeling has shown that the principle frequency vibrations of monolayer graphene resonators are sensitive to added mass of $\approx 10^{-6}$ fg [12.122]. Mass sensitivity of ≈ 2 zg has been experimentally demonstrated using graphene resonators with a high quality factor of 14 000 at low temperature [12.50].

12.4.4 Electronic Applications

Field-Effect Transistors

In order for graphene to behave as a semiconductor for use in field-effect transistors (FETs), a bandgap needs to be generated in graphene, for example by exploiting lateral confinement of charge carriers in a graphene nanoribbon [12.123] (Fig. 12.20). A gap is observed for narrow ribbons, which scales inversely with ribbon width. Gaps in excess of 100 meV were observed for widths less than 20 nm.

It is also possible to fabricate quantum dot (QD) devices entirely out of graphene [12.124–126]. Such a device consists of a graphene island connected to the source and drain via two narrow graphene constrictions and three fully tunable graphene lateral gates, as shown in Fig. 12.21a. Graphene QDs show conventional single-electron transistor characteristics, with periodic Coulomb blockade peaks (Fig. 12.21c). This can be extended to a double-QD system (Fig. 12.21b) where the coupling of the dots to the leads and between the dots is tuned by graphene in-plane gates [12.127]. This structure has been proposed for the realization of spin qubits from graphene QDs.

Various approaches have been proposed to increase the performance of graphene devices, in particular to reduce impurity scattering and enhance mobility.

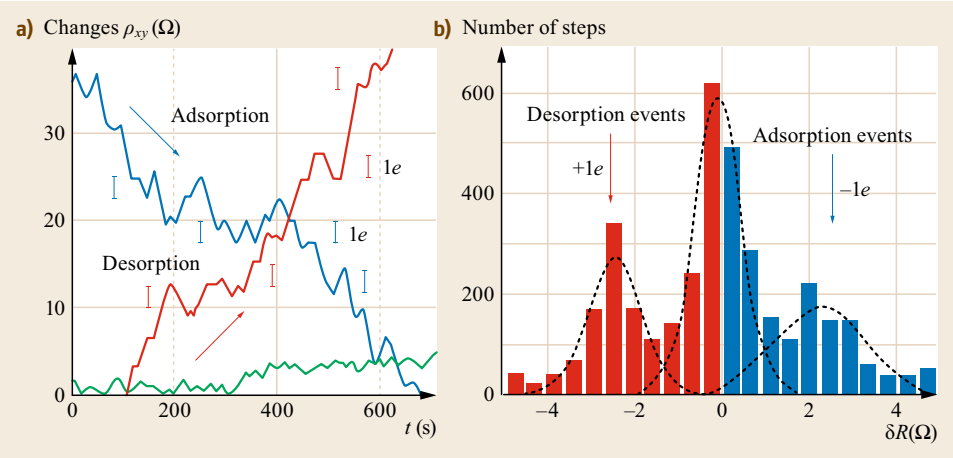


Fig. 12.19 **(a)** Changes in Hall resistivity near the neutrality point during adsorption of strongly diluted NO_2 (*blue curve*) and its desorption in vacuum at 50 °C (*red curve*). The green curve is a reference – the same device thoroughly annealed and then exposed to pure He. The *grid lines* correspond to changes in ρ_{xy} caused by adding one electron charge, $e(\delta R \approx 2.5\,\Omega)$. **(b)** Statistical distribution of step heights, δR, in this sensor (after [12.117])

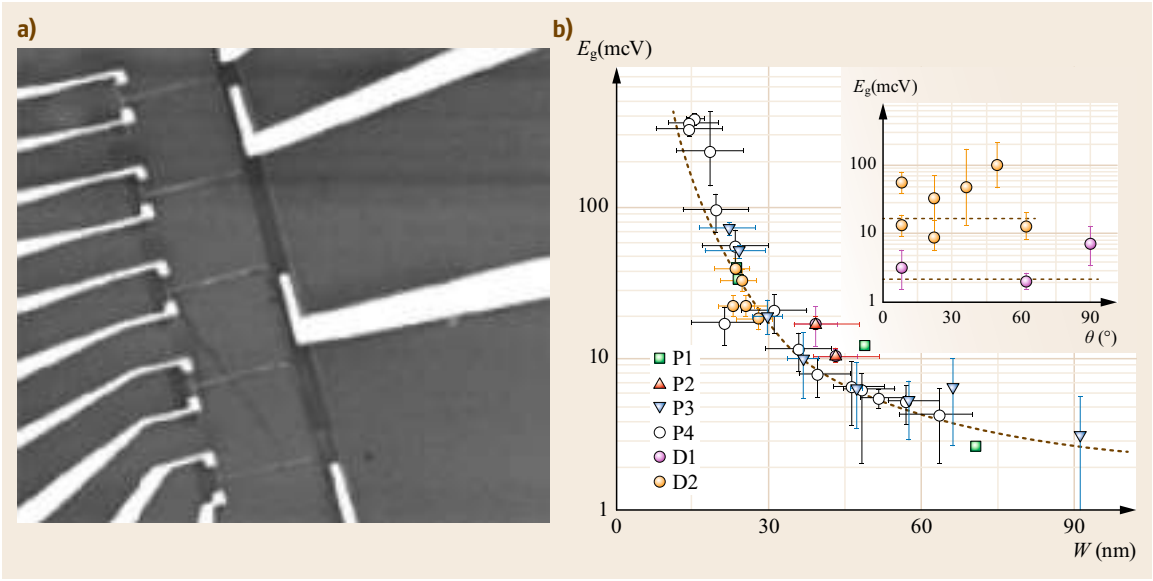


Fig. 12.20 **(a)** SEM image of a set of graphene nanoribbon devices of varying width. **(b)** Dependence of energy gap on ribbon width and orientation (*inset*). *Dashed line* shows the predicted empirical scaling. Reprinted with permission from [12.123]. Copyright (2007) by the American Physical Society

Ultrahigh current density-induced removal of adsorbents, photoresist or e-beam resist residue can be used to clean the graphene in situ during transport measurements [12.94]. A parylene-coated SiO_2 substrate used as dielectric stack for back-gating yields a stable charge-neutrality point and low hysteresis [12.128] and yields mobilities of up to $10\,000\,cm^2/Vs$. Similar results have been achieved by utilizing an organic polymer buffer between graphene and conventional top-gate dielectrics [12.129]. Large-scale integration of graphene devices can be achieved either by conventional top-down fabrication in the case of epitaxial graphene on SiC [12.67] or CVD graphene grown on metallic substrates and transferred to dielectric substrates. In the case of solution-phase graphene, bottom-up approaches such as dielectrophoretic trapping can achieve similar ultralarge-scale integration densities [12.130].

In an effort to realize commercial viability of graphene electronics, rGO has been explored as an alternative to pristine monolayer graphene. However, measurements in individual monolayer rGO flakes have yielded conductivities ranging between 0.05 and $2\,S/cm$ and field-effect mobilities of $2–200\,cm^2/Vs$ at room temperature. Conductivity decreases by up to three orders of magnitude when measured down to 4 K, following a $T^{-1/3}$ dependence, suggesting a variable range-hopping conduction between regions of highly reduced (nearly pristine) graphene islands separated by defective or poorly reduced regions.

High-Frequency Transistors

Charge carriers in graphene have very high mobility, which is not significantly affected by doping, leading to ballistic transport on a submicron scale even at room temperature. This is a much sought-after characteristic for the next generation of electronic materials, since it would allow for very fast switching times, perhaps less than 10^{-13} seconds. The low on-off ratio (< 100) in gapless graphene is not a fundamental limitation for high-frequency operation. Transistors fabricated on epitaxial graphene on SiC have achieved cutoff frequency of 100 GHz (Fig. 12.22) [12.131] and those fabricated on conventional Si/SiO_2 substrates have reached a cutoff frequency of 26 GHz [12.132].

Heterostructures and Tunneling Transistors

Graphene devices on standard SiO_2 substrates are highly disordered, exhibiting characteristics that are far inferior to the expected intrinsic properties of graphene. There is a need, therefore, for dielectrics that allow a substrate-supported geometry while retaining the intrinsic graphene quality. Hexagonal boron nitride, or hBN, offers an atomically smooth surface that is relatively free of dangling bonds and charge traps [12.133]. It also has a lattice constant similar to that of graphite, and has large optical phonon modes and a large electrical bandgap. A carrier mobility of $60\,000\,cm^2/V/s$ is obtained, which is three times larger than that on SiO_2.

An obstacle to the use of graphene as an alternative to silicon electronics has been the absence of

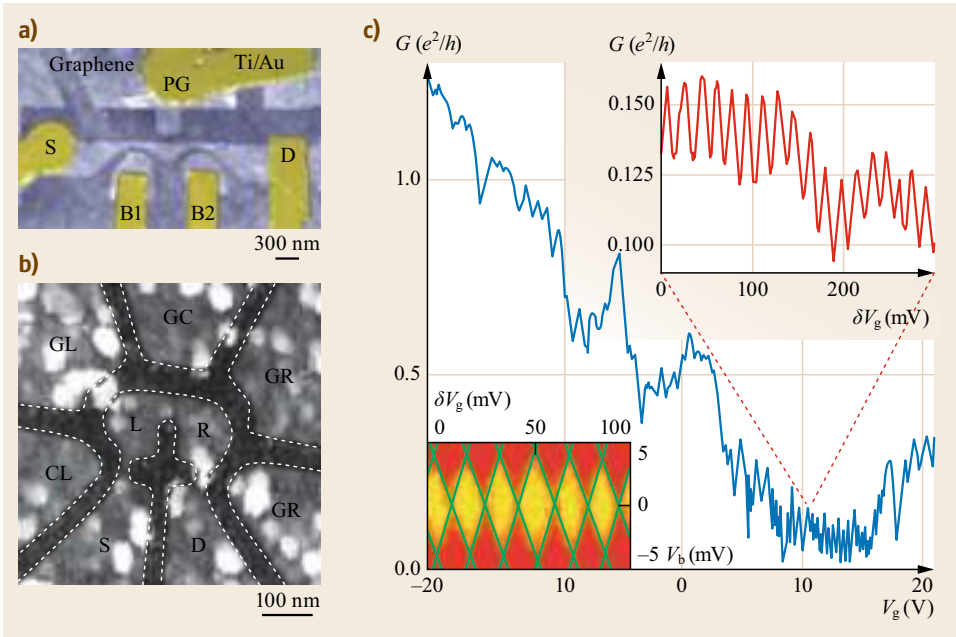


Fig. 12.21 **(a)** SEM of an all-graphene single quantum dot device. Reprinted with permission from [12.126]. Copyright (2008) American Chemical Society. **(b)** SEM of an all-graphene double quantum dot device. Reprinted from [12.127], with the permission from AIP Publishing. **(c)** Conductance of a graphene single QD device over a wide range of gate voltages at $T = 4\,\mathrm{K}$. *Upper inset*: Zoom in to low-conductance region showing Coulomb blockade oscillations. *Lower inset*: Coulomb diamonds showing differential conductance as a function of gate voltage and drains-source bias

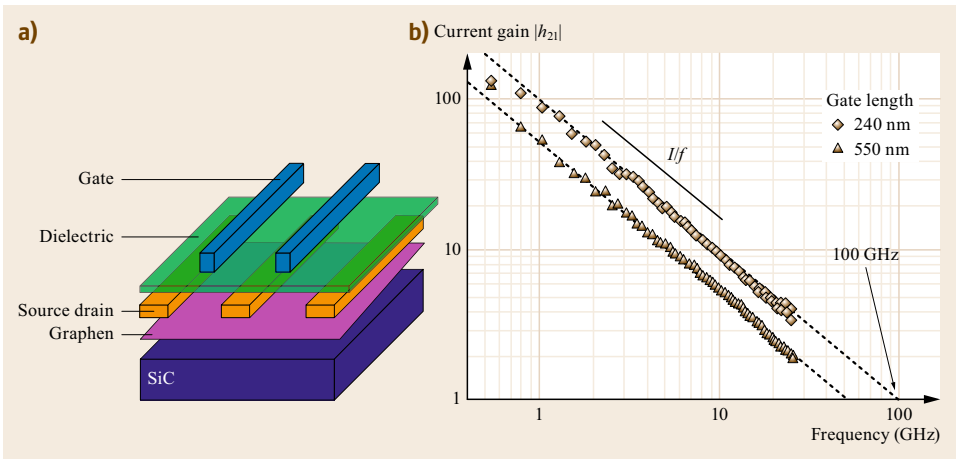


Fig. 12.22 **(a)** Schematic cross-sectional view of a top-gated graphene FET. **(b)** Measured small-signal current gain $|h_{21}|$ as a function of frequency f for a 240 nm gate (*diamonds*) and a 550 nm gate (*triangles*) graphene FET at $VD = 2.5\,\mathrm{V}$. Cutoff frequencies, f_T, were 53 and 100 GHz for the 550 nm and 240 nm devices respectively

an energy gap between its conduction and valence bands, which makes it difficult to achieve low-power dissipation in the off state, as we have encountered previously. This can be overcome in a bipolar field-effect transistor that exploits the low density of states in graphene and its one-atomic-layer thickness. These devices consist of graphene heterostructures with atomically thin boron nitride or molybdenum disulfide acting as a vertical transport barrier (Fig. 12.23a) [12.134]. The flow of charge carriers between the top and bottom graphene electrode occurs by tunneling across the semiconducting transport barrier, which can be

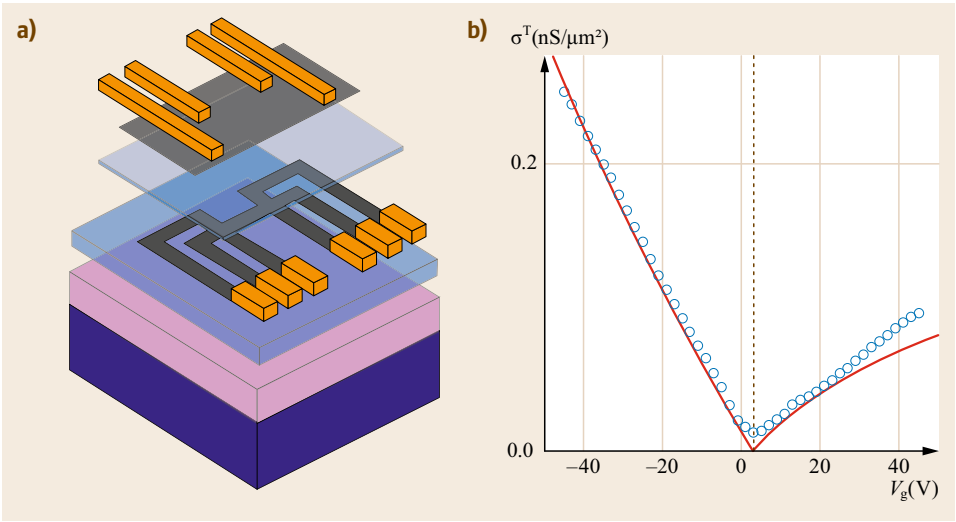


Fig. 12.23 **(a)** Schematic of a 2-D heterostructure device comprised of graphene (*black*) and hBN (*blue*) layers, on top if a SiO_2 surface (*purple*), and contacted by gold (*yellow*) electrodes. **(b)** Tunneling current versus gate voltage behavior (transfer characteristics of such a device) (after [12.134, 135])

modulated by a gate (Fig. 12.23b) [12.135]. Such devices exhibit room-temperature switching ratios of ≈ 50 for hBN and $\approx 10\,000$ for MoS_2 tunnel barriers.

As we have seen before, 2-D materials are also naturally transparent and often demonstrate a significant Young's modulus and breaking strength. 2-D materials-based heterostructures may be attractive solutions for transparent and flexible electronics. Tunneling transistors are only a few atomic layers thick and should not experience any bending-induced effects due to their vertical geometry. Therefore, vertical heterostructures have also been produced on a flexible substrates such as a polyethylene terephthalate (PET) film [12.136].

12.4.5 Spintronics

When graphene devices are fabricated with ferromagnetic electrodes, such as the soft magnetic NiFe or Co, it is possible to inject spin-polarized current into the graphene. A thin Al_2O_3 or MgO tunnel barrier is used at the ferromagnet-graphene interface. High-quality graphene enjoys ballistic transport with spin relaxation lengths between 1.5 μm and 2 μm even at room temperature, which is only weakly dependent on charge density [12.137]. The switching fields of the electrodes can be controlled by in-plane shape anisotropy. Graphene spin-valves have been constructed using either a local [12.138] or nonlocal [12.137, 139] geometry, and a clear bipolar spin signal has been observed, which is indicative of the relative magnetization directions of the electrodes (Fig. 12.24) and magnetoresistance of up to 12% has been reported in local geometry. An alternative configuration, comprised of a graphene monolayer sandwiched between two ferromagnetic layers, involves spin-polarized conduction perpendicular to the plane of the graphene in a CPP (conduction perpendicular to plan) geometry [12.140].

12.4.6 Photonics and Optoelectronics

Graphene enjoys broadband absorption of light, and therefore does not suffer from a *long-wavelength limit* of semiconductors, which become transparent at incident energies smaller than their bandgap. As a result, graphene-based photodetectors have been demonstrated to work over a much broader wavelength range. The photoresponse does not degrade for optical intensity modulations up to 40 GHz, with ultrafast response due to the high carrier mobility in graphene [12.141]. Graphene can also be used as the nonlinear saturable absorber in mode locking for ultrafast laser systems, which turns a continuous wave into a train of ultrafast optical pulses [12.142]. Graphene dispersions have been used as wideband optical limiters covering the visible and infrared regions of the spectrum [12.143].

12.4.7 Filters and Barriers

Graphene oxide sheets can be assembled into a paper-like material (Fig. 12.25a) under a directional flow, for instance by filtration of the aqueous colloidal GO through a membrane filter followed by air drying and peeling from the filter or dissolution of the fil-

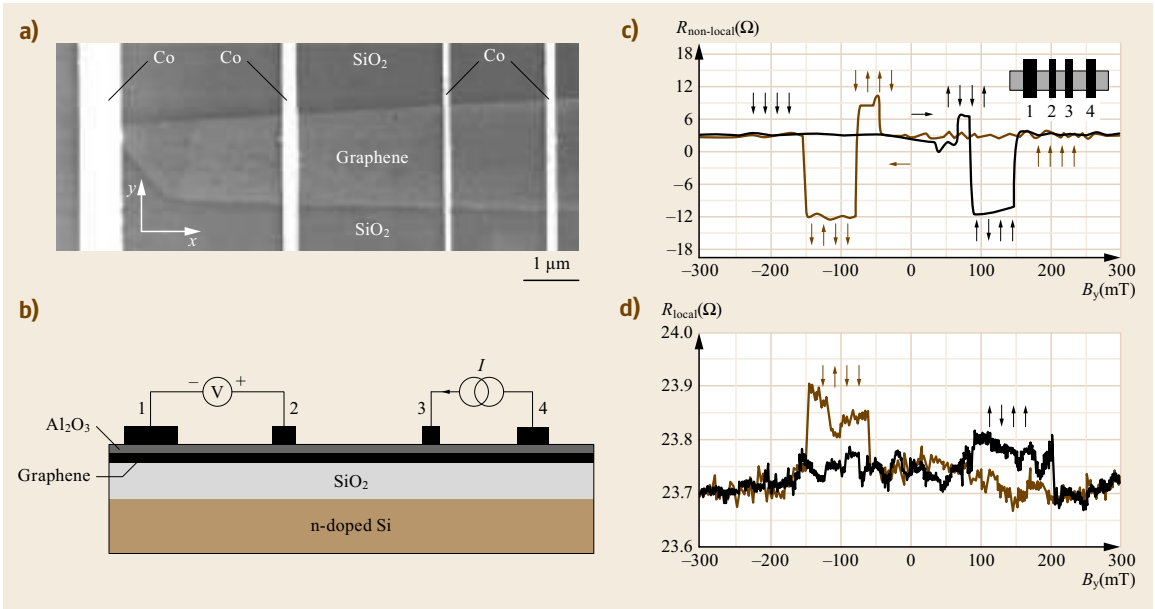


Fig. 12.24a–d Spin transport in a four-terminal spin valve device. **(a)** Scanning electron micrograph of a four-terminal single-layer graphene spin valve. **(b)** The nonlocal spin valve geometry. A current I is injected from electrode 3 through the Al_2O_3 barrier into graphene and is extracted at contact 4. The voltage difference is measured between contacts 2 and 1. **(c)** Nonlocal spin valve signal at 4.2 K. The sweep directions of the magnetic field are indicated (*brown* or *black arrows*). **(d)** A two-terminal local spin valve signal (measured between contacts 2 and 3) of about 60 V is measured at 4.2 K (after [12.137])

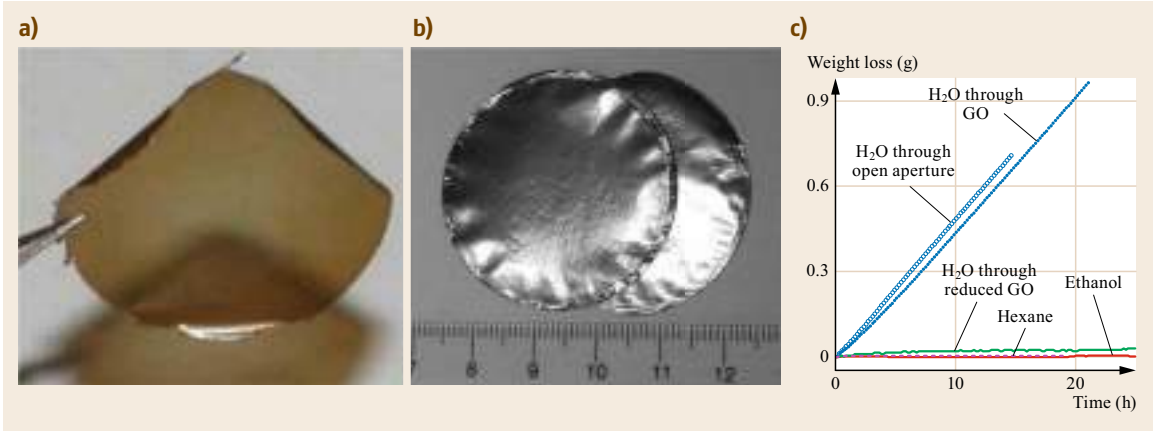


Fig. 12.25a–c Optical micrographs of **(a)** GO (after [12.144]) and **(b)** rGO films. Reprinted from [12.77] with permission. **(c)** Plot of water permeation across GO and rGO films (after [12.145])

ter [12.144]. The typical thickness of this paper ranges from submicron to tens of microns. The resulting GO paper can be cut and shaped with a knife or scissors similar to traditional paper. If the GO dispersion is first reduced, for example using hydrazine, to form a rGO colloid, followed by filtration, this results in a rGO paper (Fig. 12.25b) [12.77]. The as-prepared rGO paper displays a shiny metallic luster on both sides. These consist of well-packed layers through almost the entire cross-section of the paper.

Submicrometer-thick membranes of GO are completely impermeable to most liquids, vapors, and gases, including helium. However, for water, the evaporation rate through a GO paper is practically the same as in the absence of the GO film; that is, for the open aperture (Fig. 12.25c) [12.145]. A rGO membrane, in contrast, is

impermeable to water. Therefore, GO membranes can be used as barrier films in the design of filtration and separation materials and for selective removal of water, such as from alcohols, fuels, etc. GO membranes can also act as molecular sieves towards solutes dissolved in water, allowing fast permeation of solutes smaller than a sharp hydration radius cut-off, and acting as a barrier to solutes larger than this sharp cut-off [12.146].

It was observed that graphene films grown by chemical vapor deposition protect the surface of the metallic growth substrates of Cu and Cu/Ni alloys from air oxidation [12.147]. Unfortunately, it was found that while graphene indeed protects Cu from thermal oxidation reasonably well at high temperatures over relatively short timescales (minutes to hours), graphene greatly promotes galvanic corrosion of Cu at ambient temperature over long timescales (months to years) and can in fact be a much worse long-term anticorrosion barrier than no coating at all. Bare Cu forms a passivating native oxide, which limits further corrosion by blocking the half-reactions needed to continue oxidation [12.148]. A graphene coating on Cu slows diffusion of O_2/H_2O but assists long-term Cu corrosion in a number of ways: by maintaining a conductive pathway across the Cu surface and to selected regions of the Cu, where it maintains intimate contact; by causing nonuniform oxidation, beginning in locations of graphene defects, which leads to stress and cracking of the oxide; and providing an electrochemical driving force (graphitic materials are cathodic to many corrosion-prone metals). A solution-coated rGO membrane is one solution to accomplish large-area, multilayer and continuous barrier coatings, which also possess mechanical integrity, to avoid corrosion nucleation [12.149].

12.4.8 Supercapacitors

For capacitive energy storage, two types of systems can be used, either an electrical double-layer capacitor or a pseudocapacitor. Pure graphene electrodes are used for electrical double-layer capacitors [12.150], while polyaniline (PANI) is a material of choice for pseudocapacitors. In an electrical double-layer capacitor (EDLC), such as with graphene electrodes, the charge is stored electrostatically in a Helmholtz double layer at the electrode–electrolyte interface. There is no electrode–electrolyte charge transfer. In the case of pseudocapacitors, charge is stored electrochemically or Faradaically, where charge is reversibly transferred between the electrode and electrolyte through redox reactions. EDLCs possess good electrical conductivity and long cycle life, but relatively low specific capacitance, while pseudocapacitors have high specific capacitance but poor stability and slow charge–discharge rates. An emerging class of supercapacitors called hybrid supercapacitors utilize both ways of energy storage, and electrodes for such hybrid supercapacitors could be made of a graphene-PANI layered material [12.151].

12.4.9 Biomedical Applications

There have been multiple conflicting reports about the biocompatibility and antimicrobial activity of graphene oxide. Lately, it is perhaps becoming clearer that purified and well-prepared graphene oxide is not antibacterial. Bacteria have been shown to grow faster and to a higher optical density with GO added to the culture media or on a GO-coated surface than cultures without GO [12.152] indicating that GO is not a bactericidal or bacteriostatic material, but instead is a general growth enhancer that acts as a scaffold for cell-surface attachment and proliferation. Numerous papers have explored the biocompatibility and toxicology of GO with various animal and human cell types, and in general, GO has proven biocompatible and nontoxic at moderate dosages [12.153]. For example, GO is biocompatible with blood cells, so this will pave the way for further development of GO for targeted drug delivery and other biomedical applications. GO and rGO form excellent planar as well as 3-D scaffolds for a variety of cells, particularly promoting cell adhesion (Fig.12.26) [12.154, 155]. Graphene has the potential as a surface coating material for implants. Graphene and GO have also been shown to be biocompatible with stem cells and support stem cell attachment and proliferation [12.156]. Graphene could also serve as a scaffold for stem-cell differentiation and offer control over the differentiation pathway. Finally, due to its electrical conductivity, graphene can also form a conductive scaffold to electrically stimulate cells such as nerve cells [12.155].

While graphene is intrinsically nondispersible in water, various modifications and functionalizations can be undertaken to make it dispersible and stable in water and other aqueous physiological media. For applications in drug delivery, very small flakes of graphene known as nanographene or graphene quantum dots are used. For instance, a polyethyleneglycol (PEG) stabilized nanographene exhibited no apparent in vitro and in vivo toxicity [12.157]. Since these graphene quantum dots are intrinsically fluorescent, they could be used for bio-imaging and sensing platforms, and drug delivery systems.

Often, the graphene used for drug delivery is what is known as nanographene, with very small flake sizes of less than 100 nm, small enough to enter individual cells, and we will see an example of this later. Other approaches include graphene/polymer composite systems

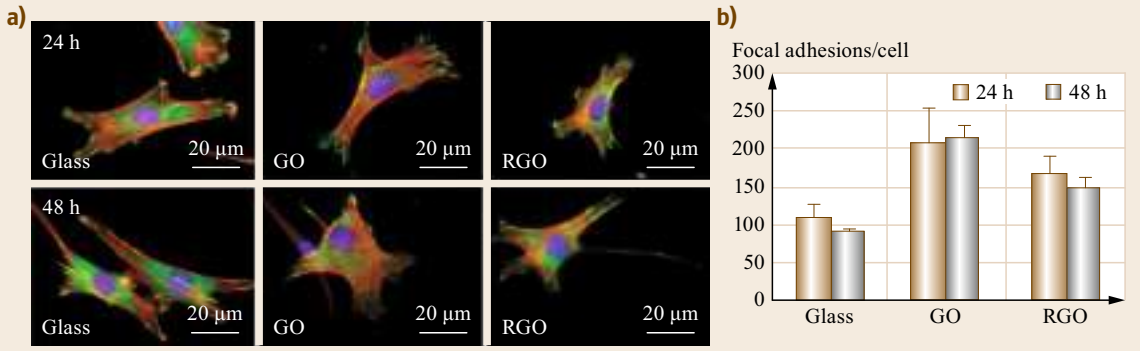


Fig. 12.26 **(a)** Cell adhesion patterns of NIH-3T3 fibroblasts on glass, GO and rGO. Reprinted with permission from [12.154]. Copyright (2010) American Chemical Society. **(b)** Quantitative analysis of cell adherence based on average focal adhesion number per cell

for stimulated drug delivery [12.158]. Graphene can also be used as a nontoxic nanovehicle for efficient gene transfection [12.159]. The PEGylated nanographene exhibits excellent tumor uptake and has strong optical absorbance in the near-infrared (NIR) region, which can be exploited for in vivo photothermal therapy, achieving ultra-efficient tumor ablation after intravenous administration and low-power NIR laser irradiation on the tumor [12.160]. Clinical use of various potent pharmacophoric molecules, many of which are aromatic, is often hampered by their poor solubility in water. Nanographene sheets exhibit a strong noncovalent association with hydrophobic aromatic drugs via physical adsorption and can be used as a vector to deliver such drugs [12.161].

With the development of its biomedical applications, the safety of GO, especially when it enters the body, is of concern to the scientific community and also the public. Towards this, the biodistribution of GO flakes of different sizes at different doses in healthy adult male mice by intravenous (IV) administration has been studied. After IV injection, small GO mainly accumulated in liver, while for large GO the lungs became the main accumulating organs [12.162]. On the other hand, PEG functionalized nanographene mainly accumulated in the reticuloendothelial system such as liver and spleen [12.163]. However, despite this accumulation, no obvious toxicity to mice was observed. Such investigations further strengthen the case for biomedical applications of graphene.

12.5 Conclusions and Outlook

The field of graphene research is still very much in its early stages, with a majority of researchers devoted to exploring the fundamental properties of graphene. Not only is graphene a new material, it is also quite literally a new dimension of possibilities. A lot of mystery still surrounds this exciting new material, and often one finds that theoretical predictions are well ahead of experimental data in an emerging field such as this. Exciting physics that only occurs in two dimensions and previously inaccessible in experiments is now being sought and new results are published regularly. Now that graphene has laid to rest the question of whether 2-D materials can exist in a stable self-supported form, researchers have started the hunt for other such 2-D materials. Boron nitride and molybdenum disulphide are the candidates that have attracted significant attention, particularly for their potential in combination with graphene.

Only recently, applications of the extraordinary properties of graphene have been experimentally demonstrated, and a still greater number of potential applications are still on the drawing board. In fact, only recently has the bulk production of graphene in sheet or solution form been achieved, but the economies of scale are still unfavorable for mass application. The commercial prospects of graphene remain unexplored, but it is hoped that in the near future, the substantial research and investment into graphene will be vindicated. But for a scientist motivated by the quest for new knowledge and understanding, graphene continues to reap rich rewards, as evidenced by the numerous results included here that were only published for the very first time during the duration of writing this chapter.

References

12.1 H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Curl, R.E. Smalley: C 60: buckminsterfullerene, Nature **318**, 162 (1985)

12.2 S. Iijima: Helical microtubules of graphitic carbon, Nature **354**(6348), 56–58 (1991)

12.3 A. Hamwi, S. Mouras, D. Djurado, J.C. Cousseins: Synthesis of first stage graphite intercalation compounds with fluorides, Rev. Chim. Miner. **24**(5), 572–582 (1987)

12.4 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov: Electric field effect in atomically thin carbon films, Science **306**(5696), 666–669 (2004)

12.5 P. Blake, E.W. Hill, A.H.C. Neto, K.S. Novoselov, D. Jiang, R. Yang, T.J. Booth, A.K. Geim: Making graphene visible, Appl. Phys. Lett. **91**(6), 63123–63124 (2007)

12.6 K.S. Novoselov, D. Jiang, F. Schedin, T.J. Booth, V.V. Khotkevich, S.V. Morozov, A.K. Geim: Two-dimensional atomic crystals, Proc. Natl. Acad. Sci. **102**(30), 10451–10453 (2005)

12.7 W.J. Arnoult, R.B. McLellan: The solubility of carbon in rhodium, ruthenium, iridium, and rhenium, Scr. Metall. **6**(10), 1013–1018 (1972)

12.8 A. Reina, X. Jia, J. Ho, D. Nezich, H. Son, V. Bulovic, M.S. Dresselhaus, J. Kong: Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition, Nano Lett. **9**(1), 30–35 (2009)

12.9 X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S.K. Banerjee, L. Colombo, R.S. Ruoff: Large area synthesis of high quality and uniform graphene films on copper foils, Science **324**(5932), 1312–1314 (2009)

12.10 K.S. Kim, Y. Zhao, H. Jang, S.Y. Lee, J.M. Kim, K.S. Kim, J.H. Ahn, P. Kim, J.Y. Choi, B.H. Hong: Large-scale pattern growth of graphene films for stretchable transparent electrodes, Nature **457**(7230), 706–710 (2009)

12.11 C. Virojanadara, M. Syväjarvi, R. Yakimova, L.I. Johansson, A.A. Zakharov, T. Balasubramanian: Homogeneous large-area graphene layer growth on 6H-SiC(0001), Phys. Rev. B **78**(24), 1–6 (2008)

12.12 E. Rollings, G.H. Gweon, S.Y. Zhou, B.S. Mun, J.L. McChesney, B.S. Hussain, A.V. Fedorov, P.N. First, W.A. de Heer, A. Lanzara: Synthesis and characterization of atomically thin graphite films on a silicon carbide substrate, J. Phys. Chem. Solids **67**(9/10), 2172–2177 (2006)

12.13 J. Penuelas, A. Ouerghi, D. Lucot, C. David, J. Gierak, H. Estrade-Szwarckopf, C. Andreazza-Vignolle: Surface morphology and characterization of thin graphene films on SiC vicinal substrate, Phys. Rev. B **79**(3), 33408 (2009)

12.14 J. Hass, R. Feng, T. Li, X. Li, Z. Zong, W.A. de Heer, P.N. First, E.H. Conrad, C.A. Jeffrey, C. Berger: Highly ordered graphene for two dimensional electronics, Appl. Phys. Lett. **89**(14), 143103–143106 (2006)

12.15 V.Y. Aristov, G. Urbanik, K. Kummer, D.V. Vyalikh, O.V. Molodtsova, A.B. Preobrajenski, A.A. Zakharov, C. Hess, T. Hänke, B. Büchner, I. Vobornik, J. Fujii, G. Panaccione, Y.A. Ossipyan, M. Knupfer, B. Victor, V. Olga: Graphene synthesis on cubic SiC/Si wafers. Perspectives for mass production of graphene-based electronic devices, Nano Lett. **10**(3), 992–995 (2010)

12.16 M. Sprinkle, M. Ruan, X. Wu, Y. Hu, M. Rubio-Roy, J. Hankinson, N.K. Madiomanana, C. Berger, W.A. de Heer: Directed self-organization of graphene nanoribbons on SiC, Nat. Nanotechnol. **5**, 727–731 (2010) doi:10.1038/nnano.2010.192

12.17 T. Aizawa, R. Souda, S. Otani, Y. Ishizawa, C. Oshima: Bond softening in monolayer graphite formed on transition-metal carbide surfaces, Phys. Rev. B **42**(18), 11469 (1990)

12.18 B.S. Itchkawitz, P.F. Lyman, G.W. Ownby, D.M. Zehner: Monolayer graphite on TaC(111): electronic band structure, Surf. Sci. **318**(3), 395–402 (1994)

12.19 T. Aizawa, R. Souda, S. Otani, Y. Ishizawa, C. Oshima: Anomalous bond of monolayer graphite on transition-metal carbide surfaces, Phys. Rev. Lett. **64**(7), 768 (1990)

12.20 T. Tanaka, A. Tajima, R. Moriizumi, M. Hosoda, R. Ohno, E. Rokuta, C. Oshima, S. Otani: Carbon nano-ribbons and their edge phonons, Solid State Commun. **123**(1/2), 33–36 (2002)

12.21 T. Nakajima, Y. Matsuo: Formation process and structure of graphite oxide, Carbon **32**(3), 469–475 (1994)

12.22 W.S. Hummers, R.E. Offeman: Preparation of graphitic oxide, J. Am. Chem. Soc. **80**(1937), 1939 (1957)

12.23 X. Fan, W. Peng, Y. Li, X. Li, S. Wang, G. Zhang, F. Zhang: Deoxygenation of exfoliated graphite oxide under alkaline conditions: A green route to graphene preparation, Adv. Mater. **20**(23), 4490–4493 (2008)

12.24 J.I. Paredes, S. Villar-Rodil, A. Martínez-Alonso, J.M.D. Tascón: Graphene oxide dispersions in organic solvents, Langmuir **24**(19), 10560–10564 (2008)

12.25 V.C. Tung, M.J. Allen, Y. Yang, R.B. Kaner: High-throughput solution processing of large-scale graphene, Nat. Nanotechnol. **4**(1), 25–29 (2009)

12.26 S. Stankovich, D.A. Dikin, R.D. Piner, K.A. Kohlhaas, A. Kleinhammes, Y. Jia, Y. Wu, S.T. Nguyen, R.S. Ruoff: Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide, Carbon **45**(7), 1558–1565 (2007)

12.27 S. Stankovich, D.A. Dikin, G.H.B. Dommett, K.M. Kohlhaas, E.J. Zimney, E.A. Stach, R.D. Piner, S.T. Nguyen, R.S. Ruoff: Graphene-based composite materials, Nature **442**(7100), 282–286 (2006)

12.28 P.R. Wallace: The band theory of graphite, Phys. Rev. **71**(9), 622 (1947)

12.29 Y. Hernandez, V. Nicolosi, M. Lotya, F. Blighe, Z. Sun, S. De, I.T. McGovern, B. Holland, M. Byrne,

Y. Gunko, J. Boland, P. Niraj, G. Duesberg, S. Krishnamurti, R. Goodhue, J. Hutchison, V. Scardaci, A.C. Ferrari, J.N. Coleman: High yield production of graphene by liquid phase exfoliation of graphite, Nat. Nanotechnol. **3**(9), 563–568 (2008)

12.30 M. Lotya, Y. Hernandez, P.J. King, R.J. Smith, V. Nicolosi, L.S. Karlsson, F.M. Blighe, S. De, W. Zhiming, I.T. McGovern, G.S. Duesberg, J.N. Coleman, Z. Wang: Liquid phase production of graphene by exfoliation of graphite in surfactant/water solutions, J. Am. Chem. Soc. **131**(10), 3611–3620 (2009)

12.31 K.R. Paton, E. Varrla, C. Backes, R.J. Smith, U. Khan, A. O'Neill, C. Boland, M. Lotya, O.M. Istrate, P. King, T. Higgins, S. Barwich, P. May, P. Puczkarski, I. Ahmed, M. Moebius, H. Pettersson, E. Long, J. Coelho, S.E. O'Brien, E.K. McGuire, B.M. Sanchez, G.S. Duesberg, N. McEvoy, T.J. Pennycook, C. Downing, A. Crossley, V. Nicolosi, J.N. Coleman: Scalable production of large quantities of defect-free few-layer graphene by shear exfoliation in liquids, Nat. Mater. **13**, 624–630 (2014)

12.32 A.A. Green, M.C. Hersam: Solution phase production of graphene with controlled thickness via density differentiation, Nano Lett. **9**(12), 4031–4036 (2009)

12.33 C.D. Simpson, J.D. Brand, A.J. Berresheim, L. Przybilla, H.J. Räder, K. Müllen: Synthesis of a giant 222 carbon graphite sheet, Chem. Eur. J. **8**(6), 1424–1429 (2002)

12.34 H.J. Räder, A. Rouhanipour, A.M. Talarico, V. Palermo, P. Samorì, K. Müllen: Processing of giant graphene molecules by soft-landing mass spectrometry, Nat. Mater. **5**(4), 276–280 (2006)

12.35 J. Cai, P. Ruffieux, R. Jaafar, M. Bieri, T. Braun, S. Blankenburg, M. Muoth, A.P. Seitsonen, M. Saleh, X. Feng, K. Müllen, R. Fasel, K. Mullen, K. Mu: Atomically precise bottom-up fabrication of graphene nanoribbons, Nature **466**(7305), 470–473 (2010)

12.36 D.V. Kosynkin, A.L. Higginbotham, A. Sinitskii, J.R. Lomeda, A. Dimiev, B.K. Price, J.M. Tour: Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons, Nature **458**(7240), 872–876 (2009)

12.37 L. Jiao, L. Zhang, X. Wang, G. Diankov, H. Dai: Narrow graphene nanoribbons from carbon nanotubes, Nature **458**(7240), 877–880 (2009)

12.38 J.O. Sofo, A.S. Chaudhari, G.D. Barber: Graphane: A two-dimensional hydrocarbon, Phys. Rev. B **75**(15), 153401 (2007)

12.39 D.C. Elias, R.R. Nair, T.M.G. Mohiuddin, S.V. Morozov, P. Blake, M.P. Halsall, A.C. Ferrari, D.W. Boukhvalov, M.I. Katsnelson, A.K. Geim, K.S. Novoselov: Control of graphene's properties by reversible hydrogenation: Evidence for graphane, Science **323**(5914), 610–613 (2009)

12.40 J.T. Robinson, J.S. Burgess, C.E. Junkermeier, S.C. Badescu, T.L. Reinecke, F.K. Perkins, M.K. Zalalutdniov, J.W. Baldwin, J.C. Culbertson, P.E. Sheehan, E.S. Snow: Properties of fluorinated graphene films, Nano Lett. **10**(8), 3001–3005 (2010)

12.41 R.R. Nair, W. Ren, R. Jalil, I. Riaz, V.G. Kravets, L. Britnell, P. Blake, F. Schedin, A.S. Mayorov, S. Yuan, M.I. Katsnelson, H.-M.M. Cheng, W. Strupinski, L.G. Bulusheva, A.V. Okotrub, I.V. Grigorieva, A.N. Grigorenko, K.S. Novoselov, A.K. Geim: Fluorographene: a two-dimensional counterpart of Teflon, Small **6**(24), 2877–2884 (2010)

12.42 T.J. Booth, P. Blake, R.R. Nair, D. Jiang, E.W. Hill, U. Bangert, A. Bleloch, M. Gass, K.S. Novoselov, M.I. Katsnelson, A.K. Geim: Macroscopic graphene membranes and their extraordinary stiffness, Nano Lett. **8**(8), 2442–2446 (2008)

12.43 N.D. Mermin: Crystalline order in two dimensions, Phys. Rev. **176**(1), 250 (1968)

12.44 A. Fasolino, J.H. Los, M.I. Katsnelson: Intrinsic ripples in graphene, Nat. Mater. **6**(11), 858–861 (2007)

12.45 J.C. Meyer, A.K. Geim, M.I. Katsnelson, K.S. Novoselov, T.J. Booth, S. Roth: The structure of suspended graphene sheets, Nature **446**(7131), 60–63 (2007)

12.46 V. Geringer, M. Liebmann, T. Echtermeyer, S. Runte, M. Schmidt, R. Rückamp, M.C. Lemme, M. Morgenstern: Intrinsic and extrinsic corrugation of monolayer graphene deposited on SiO_2, Phys. Rev. Lett. **102**(7), 76102 (2009)

12.47 M.H. Gass, U. Bangert, A.L. Bleloch, P. Wang, R.R. Nair, A.K. Geim: Free-standing graphene at atomic resolution, Nat. Nanotechnol. **3**(11), 676–681 (2008)

12.48 X. Xie, L. Ju, X. Feng, Y. Sun, R. Zhou, K. Liu, S. Fan, Q. Li, K. Jiang: Controlled fabrication of high-quality carbon nanoscrolls from monolayer graphene, Nano Lett. **9**(7), 2565–2570 (2009)

12.49 C.H. Lui, L. Liu, K.F. Mak, G.W. Flynn, T.F. Heinz: Ultraflat graphene, Nature **462**(7271), 339–341 (2009)

12.50 C. Chen, S. Rosenblatt, K.I. Bolotin, W. Kalb, P. Kim, I. Kymissis, H.L. Stormer, T.F. Heinz, J. Hone: Performance of monolayer graphene nanomechanical resonators with electrical readout, Nat. Nanotechnol. **4**(12), 861–867 (2009)

12.51 C. Lee, X. Wei, J.W. Kysar, J. Hone: Measurement of the elastic properties and intrinsic strength of monolayer graphene, Science **321**(18), 385–388 (2008)

12.52 M. Poot, H.S.J. van der Zant: Nanomechanical properties of few-layer graphene membranes, Appl. Phys. Lett. **92**(6), 63111–63113 (2008)

12.53 J. Atalaya, A. Isacsson, J.M. Kinaret: Continuum elastic modeling of graphene resonators, Nano Lett. **8**(12), 4196–4200 (2008)

12.54 D. Garcia-Sánchez, A. Van Der Zande, A.S. Paulo, B. Lassagne, P. McEuen, A. Bachtold: Imaging mechanical vibrations in suspended graphene sheets, Nano Lett. **8**(5), 1399–1403 (2008)

12.55 J.T. Robinson, M. Zalalutdinov, J.W. Baldwin, E.S. Snow, Z. Wei, P. Sheehan, B.H. Houston: Wafer-scale reduced graphene oxide films for nanomechanical devices, Nano Lett. **8**, 3441–3445

(2008)
12.56 N.I. Kovtyukhova, P.J. Ollivier, B.R. Martin, T.E. Mallouk, S.A. Chizhik, E.V. Buzaneva, A.D. Gorchinskiy: Layer-by-layer assembly of ultrathin composite films from micron-sized graphite oxide sheets and polycations, Chem. Mater. **11**(3), 771–778 (1999)
12.57 T. Ohta, A. Bostwick, J.L. McChesney, T. Seyller, K. Horn, E. Rotenberg: Interlayer interaction and electronic screening in multilayer graphene investigated with angle-resolved photoemission spectroscopy, Phys. Rev. Lett. **98**(20), 206802–206804 (2007)
12.58 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, M.I. Katsnelson, I.V. Grigorieva, S.V. Dubonos, A.A. Firsov: Two-dimensional gas of massless Dirac fermions in graphene, Nature **438**(7065), 197–200 (2005)
12.59 M.I. Katsnelson, K.S. Novoselov: Graphene: New bridge between condensed matter physics and quantum electrodynamics, Solid State Commun. **143**(1/2), 3–13 (2007)
12.60 M.I. Katsnelson, K.S. Novoselov, A.K. Geim: Chiral tunnelling and the Klein paradox in graphene, Nat. Phys. **2**(9), 620–625 (2006)
12.61 R.R. Nair, P. Blake, A.N. Grigorenko, K.S. Novoselov, T.J. Booth, T. Stauber, N.M.R. Peres, A.K. Geim: Fine structure constant defines visual transparency of graphene, Science **320**, 1308 (2008)
12.62 A.A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, C.N. Lau: Superior thermal conductivity of single-layer graphene, Nano Lett. **8**(3), 902–907 (2008)
12.63 S. Ghosh, W. Bao, D.L. Nika, S. Subrina, E.P. Pokatilov, C.N. Lau, A.A. Balandin: Dimensional crossover of thermal transport in few-layer graphene, Nat. Mater. **9**(7), 555–558 (2010)
12.64 Y.M. Zuev, W. Chang, P. Kim: Thermoelectric and magnetothermoelectric transport measurements of graphene, Phys. Rev. Lett. **102**, 96807 (2009)
12.65 P. Wei, W. Bao, Y. Pu, C.N. Lau, J. Shi: Anomalous thermoelectric transport of Dirac particles in graphene, Phys. Rev. Lett. **102**, 166808 (2009)
12.66 Y. Ouyang, J. Guo: A theoretical study on thermoelectric properties of graphene nanoribbons, Appl. Phys. Lett. **94**(26), 263107 (2009)
12.67 J. Kedzierski, H. Pei-Lan, P. Healey, P. Wyatt, C. Keast: Epitaxial graphene transistors on SiC substrates. In: *Device Res. Conf.* (2008) pp. 25–26
12.68 D. Dragoman, M. Dragoman: Giant thermoelectric effect in graphene, Appl. Phys. Lett. **91**(20), 203113–203116 (2007)
12.69 R. Sharma, J.H. Baik, C.J. Perera, M.S. Strano: Anomalously large reactivity of single graphene layers and edges toward electron transfer chemistries, Nano Lett. **10**(2), 398–405 (2010)
12.70 F.M. Koehler, A. Jacobsen, K. Ensslin, C. Stampfer, W.J. Stark: Selective chemical modification of graphene surfaces: Distinction between single- and bilayer graphene, Small **6**(10), 1125–1130 (2010)
12.71 H. He, C. Gao: General approach to individually dispersed, highly soluble, and conductive graphene nanosheets functionalized by nitrene chemistry, Chem. Mater. **22**(17), 5054–5064 (2010)
12.72 A. Sinitskii, A. Dimiev, D.A. Corley, A.A. Fursina, D.V. Kosynkin, J.M. Tour: Kinetics of diazonium functionalization of chemically converted graphene nanoribbons, ACS Nano **4**(4), 1949–1954 (2010)
12.73 P.A.D.B. Farmer, R. Golizadeh-Mojarad, V. Perebeinos, Y.-M. Lin, G.S. Tulevski, J.C. Tsang: Chemical doping and electron–hole conduction asymmetry in graphene devices, Nano Lett. **9**, 388–392 (2008)
12.74 D.B. Farmer, Y.-M.M. Lin, A. Afzali-Ardakani, P. Avouris: Behavior of a chemically doped graphene junction, Appl. Phys. Lett. **94**(21), 213103–213106 (2009)
12.75 C. Casiraghi, A. Hartschuh, E. Lidorikis, H. Qian, H. Harutyunyan, T. Gokus, K.S. Novoselov, A.C. Ferrari: Rayleigh imaging of graphene and graphene layers, Nano Lett. **7**(9), 2711–2717 (2007)
12.76 J.C. Meyer, A.K. Geim, M.I. Katsnelson, K.S. Novoselov, D. Obergfell, S. Roth, C. Girit, A. Zettl: On the roughness of single- and bi-layer graphene membranes, Solid State Commun. **143**, 101–109 (2007)
12.77 H. Chen, M.B. Müller, K.J. Gilmore, G.G. Wallace, D. Li: Mechanically strong, electrically conductive, and biocompatible graphene paper, Adv. Mater. **20**(18), 3557–3561 (2008)
12.78 P. Nemes-Incze, Z. Osváth, K. Kamarás, L.P. Biró: Anomalies in thickness measurements of graphene and few layer graphite crystals by tapping mode atomic force microscopy, Carbon **46**(11), 1435–1442 (2008)
12.79 S. Neubeck, Y.M. You, Z.H. Ni, P. Blake, Z.X. Shen, A.K. Geim, K.S. Novoselov: Direct determination of the crystallographic orientation of graphene edges by atomic resolution imaging, Appl. Phys. Lett. **97**(5), 53110–53113 (2010)
12.80 V.W. Brar, Y. Zhang, Y. Yayon, T. Ohta, J.L. McChesney, A. Bostwick, E. Rotenberg, K. Horn, M.F. Crommie: Scanning tunneling spectroscopy of inhomogeneous electronic structure in monolayer and bilayer graphene on SiC, Appl. Phys. Lett. **91**(12), 122102–122103 (2007)
12.81 Y. Zhang, V.W. Brar, F. Wang, C. Girit, Y. Yayon, M. Panlasigui, A. Zettl, M.F. Crommie: Giant phonon-induced conductance in scanning tunnelling spectroscopy of gate-tunable graphene, Nat. Phys. **4**(8), 627–630 (2008)
12.82 A. Bostwick, T. Ohta, T. Seyller, K. Horn, E. Rotenberg: Quasiparticle dynamics in graphene, Nat. Phys. **3**(1), 36–40 (2007)
12.83 A.C. Ferrari, J.C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K.S. Novoselov, S. Roth, A.K. Geim: Raman spectrum of graphene and graphene layers, Phys. Rev. Lett. **97**(18), 187401 (2006)
12.84 A.N. Obraztsov, A.V. Tyurnina, E.A. Obraztsova, A.A. Zolotukhin, B. Liu, K.-C. Chin, A.T.S. Wee: Ra-

man scattering characterization of CVD graphite films, Carbon **46**(6), 963–968 (2008)
12.85 A. Das, S. Pisana, B. Chakraborty, S. Piscanec, S.K. Saha, U.V. Waghmare, K.S. Novoselov, H.R. Krishnamurthy, A.K. Geim, A.C. Ferrari, A.K. Sood: Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor, Nat. Nanotechnol. **3**(4), 210–215 (2008)
12.86 C. Casiraghi, S. Pisana, K.S. Novoselov, A.K. Geim, A.C. Ferrari: Raman fingerprint of charged impurities in graphene, Appl. Phys. Lett. **91**(23), 233103–233108 (2007)
12.87 C. Casiraghi: Probing disorder and charged impurities in graphene by Raman spectroscopy, Phys. Status Solidi – Rapid Res. Lett. **3**(6), 175–177 (2009)
12.88 T. Yu, Z. Ni, C. Du, Y. You, Y. Wang, Z. Shen: Raman mapping investigation of graphene on transparent flexible substrate: The strain effect, J. Phys. Chem. C **112**(33), 12602–12605 (2008)
12.89 J.E. Proctor, E. Gregoryanz, K.S. Novoselov, M. Lotya, J.N. Coleman, M.P. Halsall: High-pressure Raman spectroscopy of graphene, Phys. Rev. B **80**(7), 73408 (2009)
12.90 M. Bruna, A. Vaira, A. Battiato, E. Vittone, S. Borini: Graphene strain tuning by control of the substrate surface chemistry, Appl. Phys. Lett. **97**(2), 21911–21913 (2010)
12.91 T. Mohiuddin, A. Lombardo, R. Nair, A. Bonetti, G. Savini, R. Jalil, N. Bonini, D. Basko, C. Galiotis, N. Marzari, K.S. Novoselov, A.K. Geim, A.C. Ferrari: Uniaxial strain in graphene by Raman spectroscopy: G peak splitting, Grüneisen parameters, and sample orientation, Phys. Rev. B **79**(20), 205433 (2009)
12.92 J.H. Chen, C. Jang, S. Adam, M.S. Fuhrer, E.D. Williams, M. Ishigami: Charged-impurity scattering in graphene, Nat. Phys. **4**(5), 377–381 (2008)
12.93 Y.W. Tan, Y. Zhang, K. Bolotin, Y. Zhao, S. Adam, E.H. Hwang, S. Das Sarma, H.L. Stormer, P. Kim: Measurement of Scattering rate and minimum conductivity in graphene, Phys. Rev. Lett. **99**(24), 246803–246804 (2007)
12.94 J. Moser, A. Barreiro, A. Bachtold: Current-induced cleaning of graphene, Appl. Phys. Lett. **91**(16), 163513 (2007)
12.95 J.-H. Chen, W.G. Cullen, C. Jang, M.S. Fuhrer, E.D. Williams: Defect scattering in graphene, Phys. Rev. Lett. **102**(23), 236805 (2009)
12.96 Z.H. Ni, L.A. Ponomarenko, R.R. Nair, R. Yang, S. Anissimova, I.V. Grigorieva, F. Schedin, P. Blake, Z.X. Shen, E.H. Hill, K.S. Novoselov, A.K. Geim: On resonant scatterers as a factor limiting carrier mobility in graphene, Nano Lett. **10**(10), 3868–3872 (2010)
12.97 K.I. Bolotin, K.J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, H.L. Stormer: Ultrahigh electron mobility in suspended graphene, Solid State Commun. **146**(9/10), 351–355 (2008)
12.98 X. Du, I. Skachko, E.Y. Andrei, A. Barker, E.Y. Andrei: Approaching ballistic transport in suspended graphene, Nat. Nanotechnol. **3**(8), 491–495 (2008)
12.99 J.-H. Chen, C. Jang, S. Xiao, M. Ishigami, M.S. Fuhrer: Intrinsic and extrinsic performance limits of graphene devices on SiO2, Nat. Nanotechnol. **3**(4), 206–209 (2008)
12.100 X. Du, I. Skachko, F. Duerr, A. Luican, E.Y. Andrei: Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene, Nature **462**(7270), 192–195 (2009)
12.101 K.I. Bolotin, F. Ghahari, M.D. Shulman, H.L. Stormer, P. Kim: Observation of the fractional quantum Hall effect in graphene, Nature **462**(7270), 196–199 (2009)
12.102 J.B. Oostinga, H.B. Heersche, X. Liu, A.F. Morpurgo, L.M.K. Vandersypen: Gate-induced insulating state in bilayer graphene devices, Nat. Mater. **7**(2), 151–157 (2008)
12.103 K.S. Novoselov, V.I. Fal, L. Colombo, P.R. Gellert, M.G. Schwab, K. Kim: A roadmap for graphene, Nature **490**, 192–200 (2012)
12.104 G. Eda, G. Fanchini, M. Chhowalla: Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material, Nat. Nanotechnol. **3**(5), 270–274 (2008)
12.105 G. Eda, M. Chhowalla: Graphene-based composite thin films for electronics, Nano Lett. **9**(2), 814–818 (2009)
12.106 S. Watcharotone, D.A. Dikin, S. Stankovich, R. Piner, I. Jung, G.H.B. Dommett, G. Evmenenko, S.-E. Wu, S.-F. Chen, C.-P. Liu, S.T. Nguyen, R.S. Ruoff: Graphene–silica composite thin films as transparent conductors, Nano Lett. **7**(7), 1888–1892 (2007)
12.107 Z. Liu, Q. Liu, Y. Huang, Y. Ma, S. Yin, X. Zhang, W. Sun, Y. Chen: Organic photovoltaic devices based on a novel acceptor material: Graphene, Adv. Mater. **20**(20), 3924–3930 (2008)
12.108 H. Zhang, X. Lv, Y. Li, Y. Wang, J. Li: P25-graphene composite as a high performance photocatalyst, ACS Nano **4**(1), 380–386 (2009)
12.109 T. Ramanathan, A.A. Abdala, S. Stankovich, D.A. Dikin, M.A. Herrera, R.D. Piner, D.H. Adamson, H.C. Schniepp, X. Chen, R.S. Ruoff, S.T. Nguyen, I.A. Aksay, R.K. Prud'Homme, L.C. Brinson: Functionalized graphene sheets for polymer nanocomposites, Nat. Nanotechnol. **3**(6), 327–331 (2008)
12.110 J.W.Z.C. Xu, X. Wang: Graphene-metal particle nanocomposites, J. Phys. Chem. C **112**, 19841–19845 (2008)
12.111 X. Li, G. Zhang, X. Bai, X. Sun, X. Wang, E. Wang, H. Dai: Highly conducting graphene sheets and Langmuir-Blodgett films, Nat. Nanotechnol. **3**(9), 538–542 (2008)
12.112 J. Wu, H.A. Becerril, Z. Bao, Z. Liu, Y. Chen, P. Peumans: Organic solar cells with solution-processed graphene transparent electrodes, Appl. Phys. Lett. **92**(26), 263302–263303 (2008)
12.113 X. Wang, L. Zhi, K. Müllen: Transparent, conductive graphene electrodes for dye-sensitized solar cells, Nano Lett. **8**(1), 323–327 (2008)

12.114 S. Bae, H. Kim, Y. Lee, X. Xu, J.-S. Park, Y. Zheng, J. Balakrishnan, T. Lei, H.K. Ri, Y. Il Song, Y.-J. Kim, K.S. Kim, B. Ozyilmaz, J.-H. Ahn, B.H. Hong, S. Iijima: Roll-to-roll production of 30-inch graphene films for transparent electrodes, Nat. Nanotechnol. **5**(8), 574–578 (2010)
12.115 X.S. Li, Y.W. Zhu, W.W. Cai, M. Borysiak, B.Y. Han, D. Chen, R.D. Piner, L. Colombo, R.S. Ruoff: Transfer of large-area graphene films for high-performance transparent conductive electrodes, Nano Lett. **9**(12), 4359–4363 (2009)
12.116 J. Wu, M. Agrawal, H.A. Becerril, Z. Bao, Z. Liu, Y. Chen, P. Peumans: Organic light-emitting diodes on solution-processed graphene transparent electrodes, ACS Nano **4**, 43–48 (2010)
12.117 F. Schedin, A.K. Geim, S.V. Morozov, E.W. Hill, P. Blake, M.I. Katsnelson, K.S. Novoselov: Detection of individual gas molecules adsorbed on graphene, Nat. Mater. **6**(9), 652–655 (2007)
12.118 J.T. Robinson, F.K. Perkins, E.S. Snow, Z. Wei, P.E. Sheehan: Reduced graphene oxide molecular sensors, Nano Lett. **8**(10), 3137–3140 (2008)
12.119 J.D. Fowler, M.J. Allen, V.C. Tung, Y. Yang, R.B. Kaner, B.H. Weiller: Practical chemical sensors from chemically derived graphene, ACS Nano **3**(2), 301–306 (2009)
12.120 R. Arsat, M. Breedon, M. Shafiei, P.G. Spizziri, S. Gilje, R.B. Kaner, K. Kalantar-zadeh, W. Wlodarski: Graphene-like nano-sheets for surface acoustic wave gas sensor applications, Chem. Phys. Lett. **467**(4–6), 344–347 (2009)
12.121 Y. Ohno, K. Maehashi, Y. Yamashiro, K. Matsumoto: Electrolyte-gated graphene field-effect transistors for detecting pH and protein adsorption, Nano Lett. **9**, 3318–3322 (2009)
12.122 A. Sakhaee-Pour, M.T. Ahmadian, A. Vafai: Applications of single-layered graphene sheets as mass sensors and atomistic dust detectors, Solid State Commun. **145**(4), 168–172 (2008)
12.123 M.Y. Han, B. Özyilmaz, Y. Zhang, P. Kim: Energy band-gap engineering of graphene nanoribbons, Phys. Rev. Lett. **98**(20), 206805 (2007)
12.124 L.A. Ponomarenko, F. Schedin, M.I. Katsnelson, R. Yang, E.W. Hill, K.S. Novoselov, A.K. Geim: Chaotic Dirac billiard in graphene quantum dots, Science **320**(5874), 356–358 (2008)
12.125 C. Stampfer, J. Guttinger, F. Molitor, D. Graf, T. Ihn, K. Ensslin: Tunable Coulomb blockade in nanostructured graphene, Appl. Phys. Lett. **92**(1), 12102–12103 (2008)
12.126 C. Stampfer, E. Schurtenberger, F. Molitor, J. Güttinger, T. Ihn, K. Ensslin: Tunable graphene single electron transistor, Nano Lett. **8**(8), 2378–2383 (2008)
12.127 F. Molitor, S. Droscher, J. Guttinger, A. Jacobsen, C. Stampfer, T. Ihn, K. Ensslin: Transport through graphene double dots, Appl. Phys. Lett. **94**(22), 222103–222107 (2009)
12.128 S.S. Sabri, P.L. Levesque, C.M. Aguirre, J. Guillemette, R. Martel, T. Szkopek: Graphene field effect transistors with parylene gate dielectric, Appl. Phys. Lett. **95**(24), 242103–242104 (2009)
12.129 D.B. Farmer, H.-Y. Chiu, Y.-M. Lin, K.A. Jenkins, F. Xia, P. Avouris: Utilization of a buffered dielectric to achieve high field-effect carrier mobility in graphene transistors, Nano Lett. **9**(12), 4474–4478 (2009)
12.130 A. Vijayaraghavan, C. Sciascia, S. Dehm, A. Lombardo, A. Bonetti, A.C. Ferrari, R. Krupke: Dielectrophoretic assembly of high-density arrays of individual graphene devices for rapid screening, ACS Nano **3**(7), 1729–1734 (2009)
12.131 Y.-M.Y. Lin, C. Dimitrakopoulos, K.A. Jenkins, D.B. Farmer, H.-Y.H.Y. Chiu, A. Grill, P. Avouris: 100-GHz transistors from wafer-scale epitaxial graphene, Science **327**(5966), 662 (2010)
12.132 Y.-M.Y. Lin, K.K.A. Jenkins, A. Valdes-Garcia, J.J.P. Small, D.B. Farmer, P. Avouris: Operation of graphene transistors at gigahertz frequencies, Nano Lett. **9**(1), 422–426 (2008)
12.133 C.R. Dean, A.F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K.L. Shepard, J. Hone: Boron nitride substrates for high-quality graphene electronics, Nat. Nanotechnol. **5**(10), 722–726 (2010)
12.134 S.J. Haigh, A. Gholinia, R. Jalil, S. Romani, L. Britnell, D.C. Elias, K.S. Novoselov, L.A. Ponomarenko, A.K. Geim, R. Gorbachev: Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices, Nat. Mater. **11**(9), 764–767 (2012)
12.135 L. Britnell, R.V. Gorbachev, R. Jalil, B.D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M.I. Katsnelson, L. Eaves, S.V. Morozov, N.M.R. Peres, J. Leist, A.K. Geim, K.S. Novoselov, L.A. Ponomarenko: Field-effect tunneling transistor based on vertical graphene heterostructures, Science **335**(6071), 947–950 (2012)
12.136 T. Georgiou, R. Jalil, B.D. Belle, L. Britnell, R.V. Gorbachev, S.V. Morozov, Y. Kim, A. Gholinia, S.J. Haigh, O. Makarovsky, L. Eaves, L.A. Ponomarenko, A.K. Geim, K.S. Novoselov, A. Mishchenko: Vertical field-effect transistor based on graphene-WS2 heterostructures for flexible and transparent electronics, Nat. Nanotechnol. **8**(2), 100–103 (2013)
12.137 N. Tombros, C. Jozsa, M. Popinciuc, H.T. Jonkman, B.J. van Wees: Electronic spin transport and spin precession in single graphene layers at room temperature, Nature **448**, 571–574 (2007)
12.138 E.W. Hill, A.K. Geim, K. Novoselov, F. Schedin, P. Blake: Graphene spin valve devices, Magn. IEEE Trans. **42**(10), 2694–2696 (2006)
12.139 S. Cho, Y.-F. Chen, M.S. Fuhrer: Gate-tunable graphene spin valve, Appl. Phys. Lett. **91**, 123103–123105 (2007)
12.140 T.M.G. Mohiuddin, E. Hill, D. Elias, A. Zhukov, K. Novoselov, A. Geim: Graphene in multilayered CPP spin valves, IEEE Trans. Magn. **44**, 2624–2627 (2008)
12.141 F. Xia, T. Mueller, Y.-M. Lin, A. Valdes-Garcia, P. Avouris: Ultrafast graphene photodetector, Nat. Nanotechnol. **4**, 839–843 (2009)

12.142 Q. Bao, H. Zhang, Y. Wang, Z. Ni, Y. Yan, Z.X. Shen, K.P. Loh, D.Y. Tang: Atomic-layer graphene as a saturable absorber for ultrafast pulsed lasers, Adv. Funct. Mater. **19**, 3077–3083 (2009)
12.143 J. Wang, Y. Hernandez, M. Lotya, J.N. Coleman, W.J. Blau: Broadband nonlinear optical response of graphene dispersions, Adv. Mater. **21**(23), 2430–2435 (2009)
12.144 D.A. Dikin, S. Stankovich, E.J. Zimney, R.D. Piner, G.H.B. Dommett, G. Evmenenko, S.T. Nguyen, R.S. Ruoff: Preparation and characterization of graphene oxide paper, Nature **448**(7152), 457–460 (2007)
12.145 R.R. Nair, H.A. Wu, P.N. Jayaram, I.V. Grigorieva, A.K. Geim: Unimpeded permeation of water through helium-leak-tight graphene-based membranes, Science **335**(6067), 442–444 (2012)
12.146 R.K. Joshi, P. Carbone, F.C. Wang, V.G. Kravets, Y. Su, I.V. Grigorieva, H.A. Wu, A.K. Geim, R.R. Nair: Precise and ultrafast molecular sieving through graphene oxide membranes, Science **343**(6043), 752–754 (2014)
12.147 S. Chen, L. Brown, M. Levendorf, W. Cai, S.Y. Ju, J. Edgeworth, X. Li, C.W. Magnuson, A. Velamakanni, R.D. Piner, J. Kang, J. Park, R.S. Ruoff: Oxidation resistance of graphene-coated Cu and Cu/Ni alloy, ACS Nano **5**(2), 1321–1327 (2011)
12.148 M. Schriver, W. Regan, W.J. Gannett, A.M. Zaniewski, M.F. Crommie, A. Zettl: Graphene as a long-term metal oxidation barrier: Worse than nothing, ACS Nano **7**(7), 5763–5768 (2013)
12.149 Y. Su, V.G. Kravets, S.L. Wong, J. Waters, A.K. Geim, R.R. Nair: Impermeable barrier films and protective coatings based on reduced graphene oxide, Nat. Commun. **5**, 4843 (2014)
12.150 Y. Wang, Z. Shi, Y. Huang, Y. Ma, C. Wang, M. Chen, Y. Chen: Supercapacitor devices based on graphene materials, J. Phys. Chem. C **113**(30), 13103–13107 (2009)
12.151 H.-P. Cong, X.-C. Ren, P. Wang, S.-H. Yu: Flexible graphene–polyaniline composite paper for high-performance supercapacitor, Energy Environ. Sci. **6**(4), 1185 (2013)
12.152 O.N. Ruiz, K.A.S. Fernando, B. Wang, N.A. Brown, P.G. Luo, N.D. McNamara, M. Vangsness, Y.-P.P. Sun, C.E. Bunker: Graphene oxide: a nonspecific enhancer of cellular growth, ACS Nano **5**(10), 8100–8107 (2011)
12.153 X. Zhang, J. Yin, C. Peng, W. Hu, Z. Zhu, W. Li, C. Fan, Q. Huang: Distribution and biocompatibility studies of graphene oxide in mice after intravenous administration, Carbon **49**(3), 986–995 (2011)
12.154 S.R. Ryoo, Y.K. Kim, M.H. Kim, D.H. Min: Behaviors of NIH-3T3 fibroblasts on graphene/carbon nanotubes: Proliferation, focal adhesion, and gene transfection studies, ACS Nano **4**(11), 6587–6598 (2010)
12.155 N. Li, Q. Zhang, S. Gao, Q. Song, R. Huang, L. Wang, L. Liu, J. Dai, M. Tang, G. Cheng: Three-dimensional graphene foam as a biocompatible and conductive scaffold for neural stem cells, Sci. Rep. **3**, 1604 (2013)
12.156 G.-Y. Chen, D.W.P. Pang, S.M. Hwang, H.-Y. Tuan, Y.-C. Hu: A graphene-based platform for induced pluripotent stem cells culture and differentiation, Biomaterials **33**(2), 418–427 (2012)
12.157 Y. Chong, Y. Ma, H. Shen, X. Tu, X. Zhou, J. Xu, J. Dai, S. Fan, Z. Zhang: The in vitro and in vivo toxicity of graphene quantum dots, Biomaterials **35**(19), 5041–5048 (2014)
12.158 C.L. Weaver, J.M. LaRosa, X. Luo, X.T. Cui: Electrically controlled drug delivery from graphene oxide nanocomposite films, ACS Nano **8**(2), 1834–1843 (2014)
12.159 L. Feng, S. Zhang, Z. Liu: Graphene based gene transfection, Nanoscale **3**(3), 1252–1257 (2011)
12.160 K. Yang, S. Zhang, G. Zhang, X. Sun, S.-T. Lee, Z. Liu: Graphene in mice: ultrahigh in vivo tumor uptake and efficient photothermal therapy, Nano Lett. **10**(9), 3318–3323 (2010)
12.161 S.K. Misra, P. Kondaiah, S. Bhattacharya, C.N.R. Rao: Graphene as a nanocarrier for tamoxifen induces apoptosis in transformed cancer cell lines of different origins, Small **8**(1), 131–143 (2012)
12.162 J.-H. Liu, S.-T. Yang, H. Wang, Y. Chang, A. Cao, Y. Liu: Effect of size and dose on the biodistribution of graphene oxide in mice, Nanomedicine **7**, 1801–1812 (2012)
12.163 K. Yang, J. Wan, S. Zhang, Y. Zhang, S.T. Lee, Z. Liu: In vivo pharmacokinetics, long-term biodistribution, and toxicology of pegylated graphene in mice, ACS Nano **5**(1), 516–522 (2011)

Part C MEMS/NEMS

13 MEMS/NEMS Devices and Applications
Philip X.-L. Feng, Cleveland, USA
Darrin J. Young, Salt Lake City, USA
Christian A. Zorman, Cleveland, USA

14 Single-Walled Carbon Nanotube Sensor Concepts
Cosmin I. Roman, Zurich, Switzerland
Thomas Helbling, Zurich, Switzerland
Miroslav Haluška, Zurich, Switzerland
Christofer Hierold, Zurich, Switzerland

15 Nanomechanical Cantilever Array Sensors
Hans Peter Lang, Basel, Switzerland
Martin Hegner, Dublin, Ireland
Christoph Gerber, Basel, Switzerland

16 Microfluidic Devices and Their Applications
Aditya Aryasomayajula, Hamilton, Canada
Pouriya Bayat, Toronto, Canada
Pouya Rezai, Toronto, Canada
P. Ravi Selvaganapathy, Hamilton, Canada

17 Microfluidic Micro/Nano Droplets
Gopakumar Kamalakshakurup, Irvine, USA
Derek Vallejo, Irvine, USA
Abraham Lee, Irvine, USA

18 Nanorobotics
Bradley J. Nelson, Zürich, Switzerland
Lixin Dong, East Lansing, USA

13. MEMS/NEMS Devices and Applications

Philip X.-L. Feng, Darrin J. Young, Christian A. Zorman

Microelectromechanical systems (MEMS) have played key roles in many important areas, for example, transportation, communication, automated manufacturing, environmental monitoring, healthcare, defense systems, and a wide range of consumer products. MEMS are inherently small, thus offering attractive characteristics such as reduced size, weight, and power dissipation and improved speed and precision compared to their macroscopic counterparts. Integrated circuits (IC) fabrication technology has been the primary enabling technology for MEMS besides a few special etching, bonding and assembly techniques. Microfabrication provides a powerful tool for batch processing and miniaturizing electromechanical devices and systems to a dimensional scale that is not accessible by conventional machining techniques. As IC fabrication technology continues to scale toward deep submicrometer and nanometer feature sizes, a variety of nanoelectromechanical systems (NEMS) have been rapidly emerging. Nanoscale mechanical devices and systems integrated with nanoelectronics will open a vast number of new exploratory research areas in science and engineering. NEMS will most likely serve as an enabling technology, merging engineering with the fundamental physics, quantum science, and life sciences in ways that are not currently feasible with microscale tools and technologies.

MEMS has been applied to a wide range of fields. Hundreds of microdevices have been developed for specific applications. It is thus difficult to provide an overview covering every aspect of the topic. In this chapter, key aspects of MEMS technology and applications are illustrated by selecting a few demonstrative device examples, such as pressure sensors, inertial sensors, optical and wireless communication devices. Microstructure examples with dimensions on the order of submicrometer are presented with fabrication technologies for emerging NEMS applications.

Although MEMS has experienced significant growth over the past decade, many challenges still remain. In broad terms, these challenges can be grouped into three general categories: (1) fabrication challenges; (2) packaging challenges; and (3) application challenges. Challenges in these areas will, in large measure, determine the commercial success of a particular MEMS device in both technical and economic terms. This chapter presents a brief discussion of some of these challenges as well as possible approaches to addressing them.

13.1 **MEMS Devices and Applications** 397
13.1.1 Pressure Sensor 397
13.1.2 Inertial Sensor 400
13.1.3 Optical MEMS 404
13.1.4 RF MEMS 408
13.2 **NEMS Devices and Applications** 415
13.2.1 Fabrication Techniques for NEMS 415
13.2.2 NEMS Measurement and Signal Transduction 419
13.2.3 NEMS Devices and Emerging Applications 419
13.3 **Challenges and Perspectives** 425
References 426

Microelectromechanical systems (MEMS) have enjoyed decades of research and development. Besides the traditional microfabricated sensors and actuators, the field covers microelectromechanical structures, components and devices integrated or microassembled with electronics on the same substrate or package, achieving operational systems with new enabling functions and high performance that are competitive with, or often superior to, those of their conventional counterparts. These devices and systems have played increasingly important roles in many important areas such as transportation, communication, positioning, navigation, timing, automated manufacturing, environmental monitoring, healthcare, defense systems and a wide range of

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_13

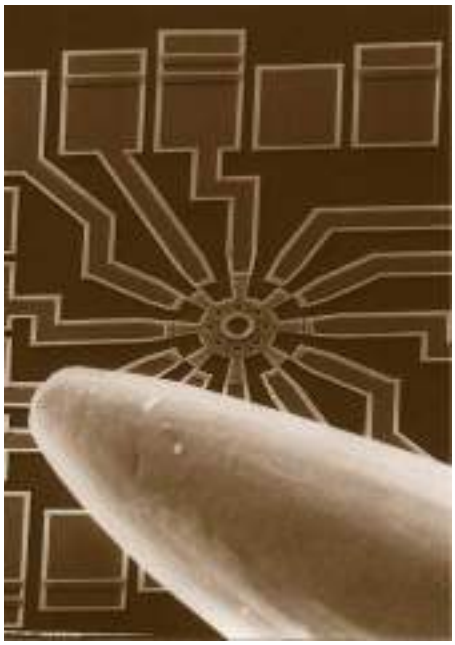


Fig. 13.1 SEM micrograph of a polysilicon microelectromechanical motor (after [13.1])

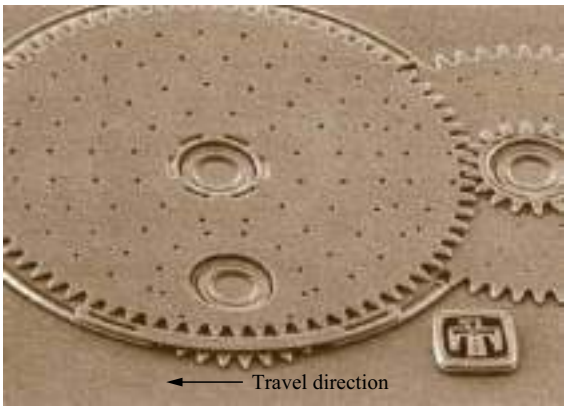


Fig. 13.2 SEM micrograph of polysilicon microgears (after [13.3])

consumer products. MEMS are inherently miniature, thus offering attractive characteristics such as reduced size and volume, low weight, low power dissipation and improved operating speed and precision compared to their macroscopic counterparts. The development of MEMS requires appropriate fabrication technologies that enable the definition of small geometries, precise dimension control, design flexibility, interfacing with microelectronics, repeatability, reliability, high yield and low cost. Relentless advances in integrated circuit (IC) fabrication technologies keep meeting all of the above criteria, and have been the primary enabling fabrication approaches for MEMS, besides some special lithography, etching, bonding and assembly techniques, as supplementary methods. Microfabrication provides a powerful tool for batch processing and miniaturization of electromechanical devices and systems into a dimensional scale, which is not accessible by conventional machining techniques. Most MEMS devices exhibit a length or width ranging from a few micrometers to several hundred micrometers with a thickness from less than a micrometer up to tens of micrometers, depending upon fabrication technique employed. Physical displacements of a sensor or an actuator are typically on the same order of magnitude as the thickness (or width), or smaller. Figure 13.1 shows a scanning electron microcopy (SEM) image of a microelectromechanical motor developed in the late 1980s [13.1]. Polycrystalline silicon (polysilicon or poly-Si) surface-micromachining technology was used to fabricate the micromotor achieving a diameter of 150 μm and a minimum vertical feature size on the order of a micrometer. A probe tip is also shown in the micrograph for a size comparison. This and similar devices [13.2] demonstrated what MEMS technology could accomplish in microscale machining at that time and served as a strong technology indicator for continued MEMS development. The field has expanded greatly over the last two decades along with rapid technology advances. Figure 13.2, for example, shows a photo of microgears fabricated in the mid-1990s using a five-level polysilicon surface-micromachining technology [13.3]. This device represents one of the most advanced surface-micromachining fabrication process developed in the early days. A wide range of sophisticated microelectromechanical devices and systems have been developed through applying such technologies since then.

Since the 1990s, in parallel to the rapid development of MEMS spurred by growing pervasive applications and advances in mainstream fabrication technologies in the semiconductor industry, physicists and engineers have also been actively exploring various means for patterning and fabricating suspended, mechanically free-to-move nanoscale structures, and engineering such nanostructures into nanoelectromechanical systems (NEMS). As IC fabrication technologies continue to scale toward deep submicron and nanometer feature sizes, this has been envisioned as a new reinforcing trend for more mature MEMS and has been rapidly emerging since the 2000s. Nanoscale electromechanical devices and systems integrated with nanoelectronics promise to open a vast number of new exploratory research areas in science and engineering. From a historical perspective, NEMS have been a fascinating frontier where mesoscopic physics meets advanced engineering, and where the *top-down* paradigm (e.g., nanolithography and nanomachining) meets the *bottom-up* nanofabrication approaches (e.g., growth, synthesis and

self-assembly processes). NEMS will most likely serve as an enabling technology merging engineering with the physical sciences, precision instruments, novel analytical tools for the life sciences and medicine, etc., in ways that are not currently feasible with the microscale tools and technologies.

This chapter will provide a general overview on MEMS and NEMS devices along with their mainstream and emerging applications. MEMS technology has been applied to a wide range of fields. Over hundreds of types of microdevices have been developed for specific applications. Thus, it is difficult to provide an overview covering every aspect of the field. It is the authors' intent to illustrate key aspects of MEMS technology and its impact to specific applications by selecting a few demonstrative device examples in this chapter. For a wide-ranging discussion of nearly all types of micromachined sensors and actuators, books by *Kovacs* [13.4] and *Senturia* [13.5] are recommended.

13.1 MEMS Devices and Applications

MEMS devices have played key roles in many areas of development. Microfabricated sensors, actuators and electronics for MEMS-IC integration are the most critical components required to implement a complete microsystem for a specific function. Microsensors and actuators can be fabricated by various micromachining processing technologies. In this section, a number of selected MEMS devices are presented to illustrate the basic device operating principles as well as to demonstrate key aspects of the microfabrication technology and application impact.

13.1.1 Pressure Sensor

Pressure sensors were one of the early devices developed by silicon (Si) micromachining technologies and have already become successful commercial products. MEMS pressure sensors and their variant devices have been widely used in various industrial and biomedical applications. These sensors can be based on piezoelectric, piezoresistive, capacitive and resonant sensing mechanisms. Si bulk and surface-micromachining techniques have been used for batch fabrication of pressure sensors, thus achieving size miniaturization and low cost. Two types of pressure sensors, piezoresistive and capacitive, are described here for an illustration purpose.

Piezoresistive Pressure Sensor

The piezoresistive effect in Si has been widely used for implementing pressure sensors. A pressure-induced strain changes the Si band structure, thus changing the resistivity of the crystal. The piezoresistive effect is typically crystal orientation-dependent and is also affected by doping and temperature. A practical piezoresistive pressure sensor can be implemented by fabricating four sensing resistors along the edges of a thin silicon diaphragm, which acts as a mechanical amplifier to increase the stress and strain at the sensor site. The four sensing elements are connected in a bridge configuration with push–pull signals to increase the responsivity and sensitivity. The measurable pressure range for such a sensor can be from 10^{-3} to 10^{6} Torr depending upon the design. An example of a piezoresistive pressure sensor is shown in Fig. 13.3. The device consists of a Si diaphragm suspended over a reference vacuum cavity to form an absolute pressure sensor. An external pressure applied over the diaphragm introduces a stress on the sensing resistors, thus resulting in a resistance value change corresponding to the pressure. The fabrication sequence is outlined as follows. The piezoresistors are typically first formed through a boron diffusion process followed by a high-temperature annealing step in order to achieve a resistance value on the order of a few kilo-ohms (kΩ). The wafer is then passivated with a silicon dioxide (SiO_2) layer and contact windows are opened for metallization. At this point, the wafer is patterned on the backside, followed by a timed Si wet etch to form the diaphragm, typically having a thickness around a few tens of micrometers. The diaphragm can have a length of several hundreds of micrometers. A second Si wafer is then bonded to the device wafer in vacuum to form a reference vacuum cavity, thus completing the fabrication process. The second wafer

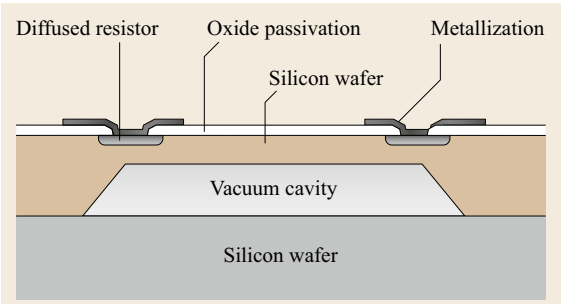


Fig. 13.3 Cross-sectional schematic of a piezoresistive pressure sensor

can also be further etched through to form an inlet port, implementing a gauge pressure sensor [13.6]. The piezoresistive sensors are simple to fabricate and can be readily interfaced with electronic systems. However, the resistors are temperature-dependent and consume DC power. Long-term characteristic drift and resistor thermal noise ultimately limit the sensor detection resolution.

Capacitive Pressure Sensor

Capacitive pressure sensors are attractive because they are virtually temperature-independent and consume zero DC power. The devices do not exhibit initial turn-on drift and are stable over time. Furthermore, CMOS (complementary metal-oxide-semiconductor) microelectronic circuits can be readily interfaced with the sensors to provide advanced signal conditioning and processing, thus improving overall system performance. An example of a capacitive pressure sensor is shown in Fig. 13.4. The device consists of an edge-clamped Si diaphragm suspended over a vacuum cavity. The diaphragm can be square or circular with a typical thickness of a few micrometers and a length or radius of a few hundred micrometers, respectively. The vacuum cavity typically has a depth of a few micrometers. The diaphragm and substrate form a pressure-dependent air-gap variable capacitor. An increased external pressure causes the diaphragm to deflect towards the substrate, thus resulting in an increase in the capacitance value. A simplified fabrication process can be outlined as follows. A Si wafer is first patterned and etched through to form the cavity. The wafer is then oxidized followed by bonding to a second Si wafer with a heavily-doped boron layer, which defines the diaphragm thickness, at the surface. The bonding process can be performed in vacuum to create the vacuum cavity. If the vacuum bonding is not performed at this stage, a low-pressure sealing process can be used to form the vacuum cavity after patterning the sensor diaphragm, provided that sealing channels are available. The Si substrate above the boron layer is then removed through a wet-etching process, followed by patterning to form the sensor diaphragm, which serves as the device top electrode. Contact pads are formed by metallization and patterning. This type of pressure sensor exhibits a nonlinear characteristic and a limited dynamic range. These phenomena, however, can be alleviated through applying an electrostatic force-balanced feedback architecture. A common practice is to introduce another electrode above the sensing diaphragm through wafer bonding [13.7], thus forming two capacitors in series with the diaphragm being the middle electrode. The capacitors are interfaced with electronic circuits, which convert the sensor capacitance value to an output voltage corresponding to the diaphragm position. This voltage is further processed to generate a feedback signal to the top electrode, thus introducing an electrostatic pull up force to maintain the deflectable diaphragm at its nominal position. This negative feedback loop would substantially minimize the device nonlinearity and also extend the sensor dynamic range.

A capacitive pressure sensor achieving an inherent linear characteristic response and a wide dynamic range can be implemented by employing a touch-mode architecture [13.8]. Figure 13.5 shows the cross-sectional view of a touch-mode pressure sensor. The device consists of an edge-clamped Si diaphragm suspended over a vacuum cavity. The diaphragm deflects under an increasing external pressure and touches the substrate, causing a linear increase in the sensor capacitance value beyond the touch-point pressure. Figure 13.6 shows a typical device characteristic curve. The touch-point pressure can be designed through engineering the sensor geometric parameters such as the diaphragm size, thickness, cavity depth, etc. for various application requirements. The device can be fabricated using a process flow similar to the flow outlined for the basic capacitive pressure sensor. Figure 13.7 presents a photo of a fabricated touch-mode sensor employing a circular diaphragm with a diameter of 800 μm and a thickness of 5 μm suspended over a 2.5 μm vacuum cavity. The device achieves a touch-point pressure of 8 psi and exhibits a linear capacitance range from 33 pF at 10 psi to 40 pF at 32 psi (absolute pressures).

The above processes use bulk Si materials for machining and are usually referred to as bulk microma-

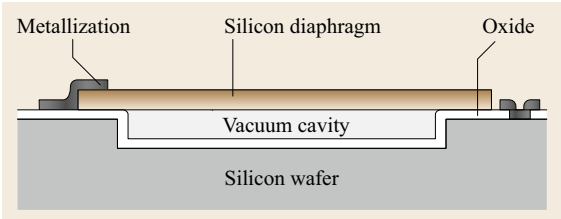


Fig. 13.4 Cross-sectional schematic of a capacitive pressure sensor

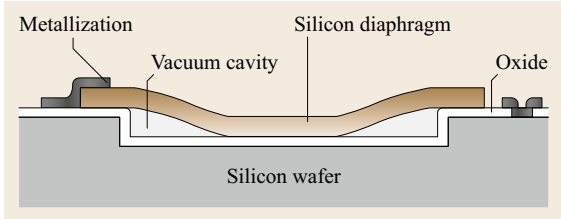


Fig. 13.5 Cross-sectional schematic of a touch-mode capacitive pressure sensor

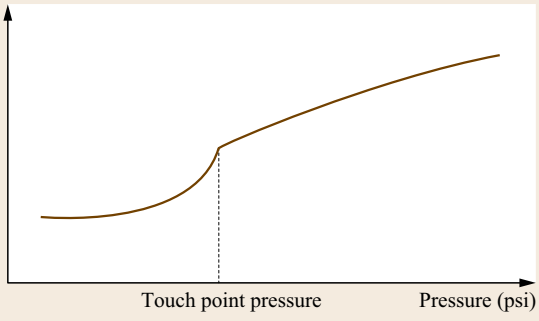


Fig. 13.6 Touch-mode capacitive pressure sensor characteristic response

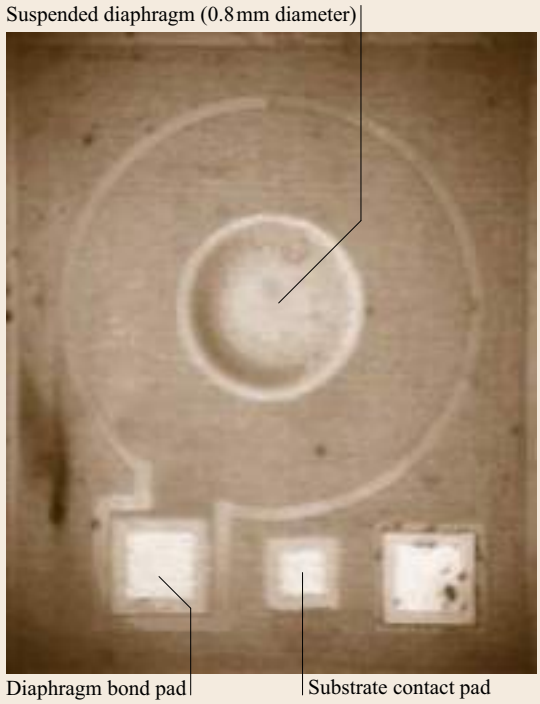


Fig. 13.7 Photo of a touch-mode capacitive pressure sensor (after [13.8])

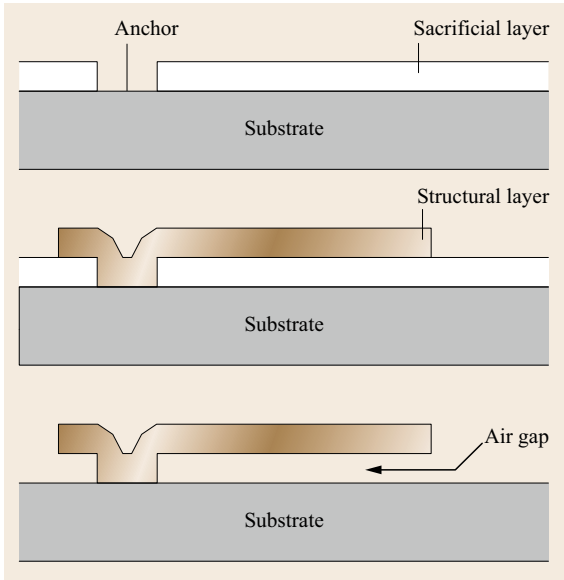


Fig. 13.8 Simplified fabrication sequence of surface-micromachining technology

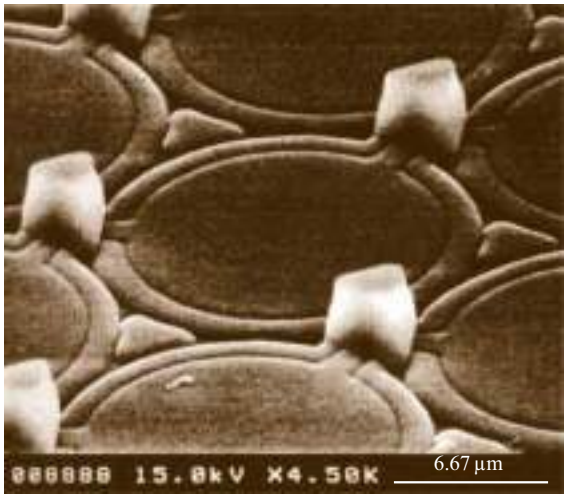


Fig. 13.9 SEM micrograph of polysilicon surface-micromachined capacitive pressure sensors (after [13.9])

chining. The same devices can also be fabricated using surface micromachining. Surface-micromachining technology is attractive for integrating MEMS sensors with on-chip electronic circuits. As a result, advanced signal processing capabilities such as data conversion, offset and noise cancellation, digital calibration, temperature compensation, etc. can be implemented adjacent to microsensors on the same substrate, providing a complete high-performance microsystem solution. The single chip approach also eliminates external wiring, which is critical for minimizing noise pickup and enhancing system performance. Surface micromachining, simply stated, is a method of fabricating MEMS through depositing, patterning and etching a sequence of thin films with thicknesses on the order of a micrometer. Figure 13.8 illustrates a typical surface-micromachining process flow [13.10]. The process starts by depositing a layer of sacrificial material such as SiO_2 over a wafer followed by anchor formation. A structural layer, typically a poly-Si film, is deposited and patterned. The underlying sacrificial layer

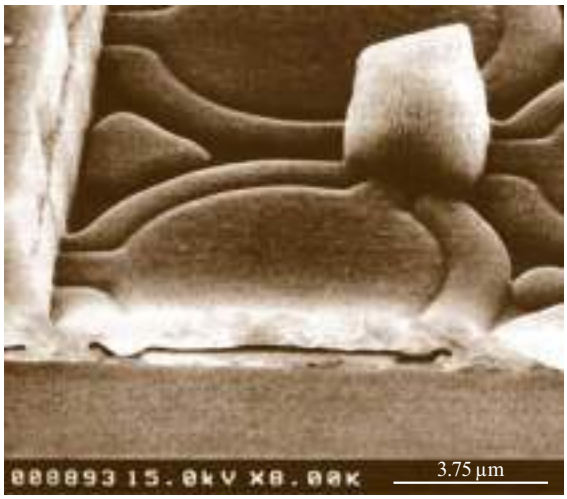


Fig. 13.10 SEM micrograph of a close-up view of a polysilicon surface-micromachined capacitive pressure sensor (after [13.9])

is then removed to freely release the suspended microstructure and to complete the fabrication sequence. The processing materials and steps are compatible with standard integrated circuit process, thus can be readily incorporated as an add-on module to an IC process [13.10–12]. A similar surface-micromachining technology has been developed to produce monolithic pressure sensor systems [13.13]. Figure 13.9 shows an SEM micrograph of an array of MEMS capacitive pressure sensors fabricated with BiCMOS electronics on the same substrate. Each sensor consists of a 0.8 µm thick circular poly-Si membrane with a diameter on the order of 20 µm suspended over a 0.3 µm deep vacuum cavity. The devices operate using the same principle as the sensor shown in Fig. 13.4. A close-up view of the sensor cross-section is shown in Fig. 13.10, which shows the suspended membrane and underneath air gap. These sensors have demonstrated operations in pressure ranges up to 400 bar with an accuracy of 1.5%.

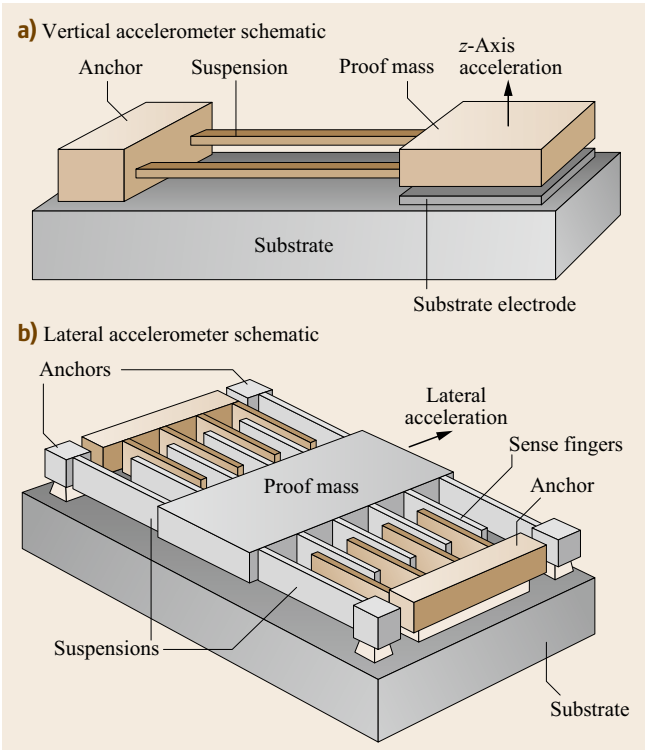


Fig. 13.11a,b Schematics of vertical (a) and lateral (b) accelerometers

13.1.2 Inertial Sensor

Micromachined inertial sensors consist of accelerometers and gyroscopes. These devices are one of the important representative types of Si MEMS sensors that have been successfully commercialized. MEMS accelerometers alone have the second largest sales volume after pressure sensors. Gyroscopes are expected to reach a comparable sales volume in the foreseeable future. Accelerometers have been used in a wide range of applications including automotive application for safety systems, active suspension and stability control, biomedical application for activity monitoring, and numerous consumer products such as head-mount displays, camcorders and three-dimensional mice. High-

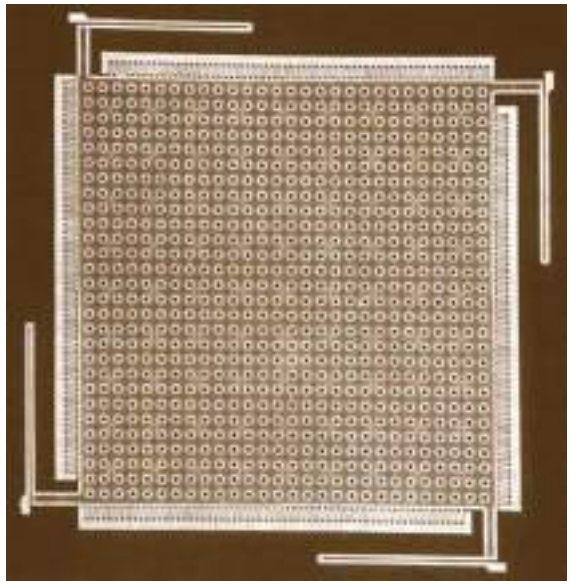


Fig. 13.12 SEM micrograph of a polysilicon surface-micromachined z-axis accelerometer (after [13.14])

sensitivity accelerometers are crucial for implementing self-contained navigation and guidance systems. A gyroscope is another type of inertial sensor that measures rate or angle of rotation. The devices can be used along with accelerometers to provide heading information in an inertial navigation system. Gyroscopes also are useful in applications such as automotive ride stabilization and rollover detection, camcorder stabilization and virtual reality. Inertial sensors fabricated by micromachining technology can achieve reduced size, weight and cost, all of which are critical for consumer applications. More importantly, these sensors can be integrated with microelectronic circuits to achieve a functional microsystem with high performance.

Accelerometer

An accelerometer generally consists of a proof mass suspended by compliant mechanical suspensions anchored to a fixed frame. An external acceleration displaces the support frame relative to the proof mass. The displacement can result in an internal stress change in the suspension, which can be detected by piezoresistive sensors as a measure of the external acceleration. The displacement can also be detected as a capacitance change in capacitive accelerometers. Capacitive sensors are attractive for various applications because they exhibit high sensitivity and low temperature-dependence, turn-on drift, power dissipation and noise. The sensors can also be readily integrated with CMOS electronics to perform advanced signal processing for high system-level performance. Capacitive accelerometers may be divided into two categories: vertical- and lateral-type sensors. Figure 13.11 shows sensor structures for the two versions. In a vertical device, the proof mass is suspended above the substrate electrode by a small gap typically on the order of a micrometer, forming a parallel-plate sense capacitance. The proof mass moves in the direction perpendicular to the substrate

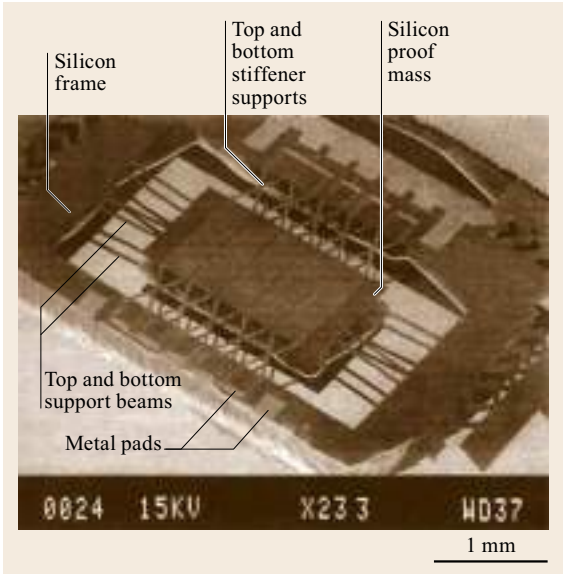


Fig. 13.13 SEM micrograph of a MEMS z-axis accelerometer fabricated using a combined-surface and bulk-micromachining technology (after [13.15])

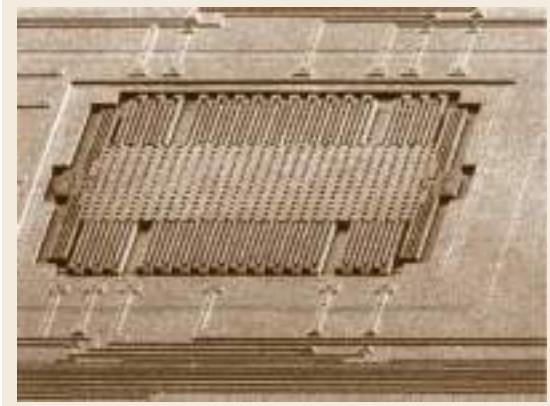


Fig. 13.14 SEM micrograph of a polysilicon surface-micromachined lateral accelerometer (Copyright Analog Devices Inc.)

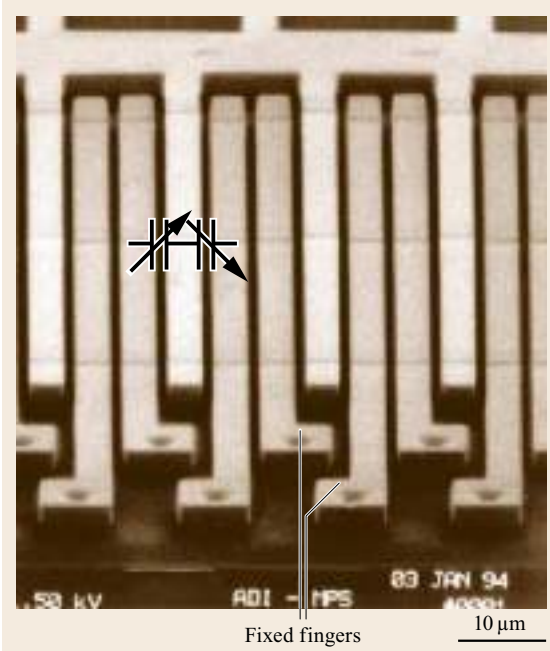


Fig. 13.15 SEM micrograph of a capacitive sensing finger structure

(z-axis) upon a vertical input acceleration, thus changing the gap and hence the capacitance value. The lateral accelerometer consists of a number of movable fingers attached to the proof mass, forming a sense capacitance with an array of fixed parallel fingers. The sensor proof mass moves in a plane parallel to the substrate when subjected to a lateral input acceleration, thus changing the overlap area of these fingers and hence the capacitance value. Figure 13.12 shows a top-view SEM image of a surface-micromachined poly-Si z-axis accelerometer [13.14]. The device consists of a 400 μm × 400 μm proof mass with a thickness of 2 μm suspended above the substrate electrode by four folded-beam suspensions with an air gap around 2 μm, thus achieving a sense capacitance of approximately 500 fF. The visible holes are used to ensure complete removal of the sacrificial oxide underneath the proof mass at the end of the fabrication process. The sensor can be interfaced with a microelectronic charge amplifier converting the capacitance value to an output voltage for further signal processing and analysis. Force-feedback architecture can be applied to stabilize the proof mass position. The combs around the periphery of the proof mass can exert an electrostatic levitation force on the proof mass to achieve the position control, thus improving the system frequency response and linearity performance [13.14].

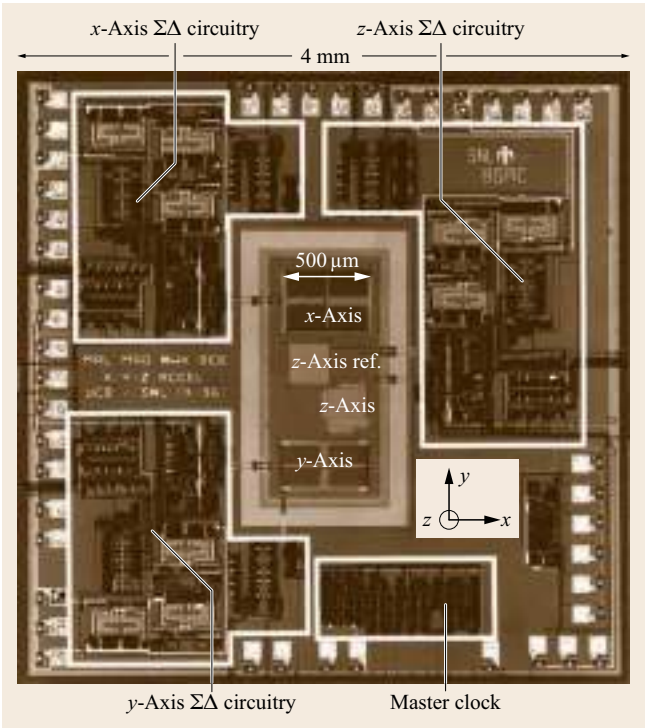


Fig. 13.16 Photo of a monolithic three-axis polysilicon surface-micromachined accelerometer with integrated interface and control electronics (after [13.16])

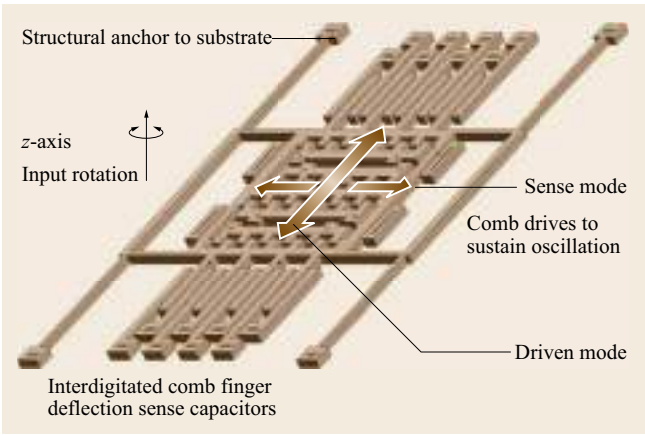


Fig. 13.17 Schematic of a vibratory rate gyroscope

Surface-micromachined accelerometers typically suffer from severe thermomechanical fluctuations, often referred to as Brownian motions [13.17], due to the small proof mass, thus can result in a high mechanical noise floor, which ultimately limits the sensor resolution. Vacuum packaging can be employed to minimize this adverse effect but with a penalty of increasing system complexity and cost. Accelerometers using large proof masses fabricated by bulk micromachining or a combination of surface- and bulk-micromachining techniques are attractive for circumventing this problem. Figure 13.13 shows an SEM micrograph of an all-Si z-axis accelerometer fabricated through a single silicon wafer by using combined-surface and bulk-micromachining processes to obtain a large proof mass with dimensions of approximately 2 mm × 1 mm × 450 μm [13.15]. The large mass suppresses the Brownian motion effect, achieving a high performance with a resolution on the order of several μg.

A surface-micromachined lateral accelerometer developed by Analog Devices Inc. is shown in Fig. 13.14. The sensor consists of a center proof mass supported by folded beam suspensions with arrays of attached movable fingers, forming a sense capacitance with the fixed parallel fingers. The device is fabricated using a 6 μm-thick poly-Si structural layer with a small air gap on the order of a micrometer to increase the sensor capacitance value, thus improving the device resolution. Figure 13.15 shows a close-up view of the finger structure for a typical lateral accelerometer. Each movable finger forms differential capacitances with two adjacent fixed fingers. This sensing capacitance configuration is attractive for interfacing with differential electronic detection circuits to suppress common-mode noise and other undesirable signal coupling. Monolithic accelerometers with a 3-axis sensing capability integrated

with on-chip electronic detection circuits have been developed using surfacing micromachining and CMOS microelectronics fabrication technologies [13.16]. Figure 13.16 shows a photo of one of these microsystem chips, which has an area of 4 mm × 4 mm. One vertical accelerometer and two lateral accelerometers are placed at the chip center with corresponding detection electronics along the periphery. A z-axis reference device, which is not movable, is used with the vertical sensor for electronic interfacing. The prototype system achieves a sensing resolution on the order of 1 mG with a 100 Hz bandwidth along each axis. The level of performance is adequate for automobile safety activation systems, vehicle stability and active suspension control, and various consumer products.

Gyroscope

Most micromachined gyroscopes employ vibrating mechanical elements to sense rotations. The sensors rely on energy transfer between two vibration modes of a structure caused by Coriolis acceleration. Figure 13.17 presents a schematic of a z-axis vibratory rate gyroscope. The device consists of an oscillating mass electrostatically driven into resonance along the drive-mode axis using comb fingers. An angular rotation along the vertical axis (z-axis) introduces a Coriolis acceleration, which results in a structure deflection along the sense-mode axis, shown in the figure. The deflection changes the differential sense capacitance value, which can be detected as a measure of input angular rotation. A z-axis vibratory rate gyroscope operating upon this principle is fabricated using surface-micromachining technology and integrated together with electronic detection circuits, as illustrated in Fig. 13.18 [13.18]. The micromachined sensor is fabricated using poly-Si structural material with a thickness around 2 μm and occupies an area of 1 mm × 1 mm. The sensor achieves a resolution of approximately $1^\circ s^{-1} Hz^{-1/2}$ under a vacuum pressure around 50 mTorr.

A dual-axis gyroscope based on a rotational disk at its resonance can be used to sense angular rotation along two lateral axes (x-axis and y-axis). Figure 13.19 shows a device schematic demonstrating the operating principle. A rotor disk supported by four mechanical suspensions can be driven into angular resonance along the z-axis. An input angular rotation along the x-axis will generate a Coriolis acceleration causing the disk to rotate along the y-axis and vice versa. This Coriolis acceleration-induced rotation will change the sensor capacitance values between the disk and different sensing electrodes underneath. The capacitance change can be detected and processed by electronic interface circuits. Angular rotations along the two lateral axes can be measured simultaneously using this device

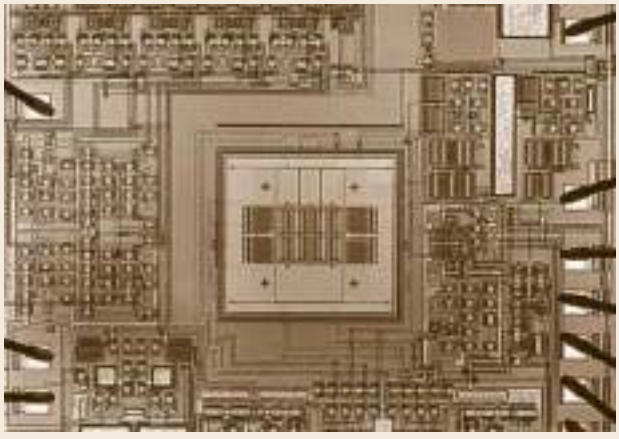


Fig. 13.18 Photo of a monolithic polysilicon surface-micromachined z-axis vibratory gyroscope with integrated interface and control electronics (after [13.18])

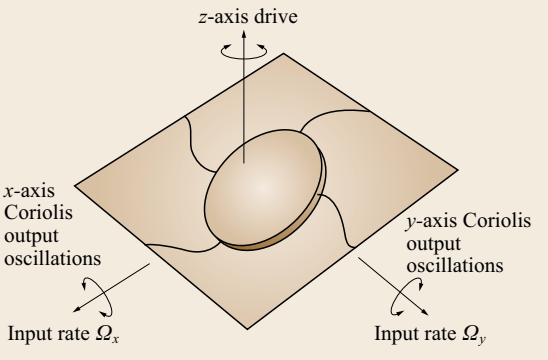


Fig. 13.19 Schematic of a dual-axis gyroscope

architecture. Figure 13.20 shows a photo of a dual-axis gyroscope fabricated using a 2 μm-thick poly-Si surface-micromachining technology [13.19]. As shown in the figure, curved electrostatic drive combs are positioned along the circumference of the rotor dick to drive it into resonance along the vertical axis. The gyroscope exhibits a low random walk of $1^\circ hr^{-1/2}$ under a vacuum pressure around 60 mTorr. With accelerometers and gyroscopes each capable of three-axis sensing, a micromachine-based inertial measurement system providing a six-degrees-of-freedom sensing capability can be obtained. Figure 13.21 presents a photo of such a system containing a dual-axis gyroscope, a z-axis gyroscope, and a three-axis accelerometer chip integrated with microelectronic circuitry. Due to the precision in device layout and fabrication, the system can measure angular rotation and acceleration without the need to align individual sensors.

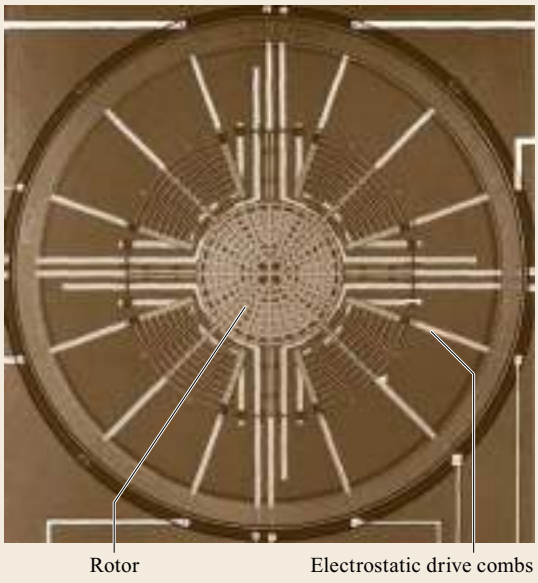


Fig. 13.20 Photo of a polysilicon surface-micromachined dual-axis gyroscope (after [13.19])

13.1.3 Optical MEMS

Surface micromachining has served as a key enabling technology to develop microelectromechanical optical devices for various applications ranging from sophisticated visual information displays and fiber-optic telecommunication to barcode reading. Most of the existing optical systems are implemented using conventional optical components, which suffer from bulky size, high cost, large power consumption, poor efficiency and reliability issues. MEMS technology is promising for producing miniaturized, reliable, inexpensive optical components to revolutionize conventional optical systems [13.20]. In this section of the chapter, a few selected MEMS optical devices will be presented to illustrate their impact in the fields of visual display, precision optical platform and data switching for optical communication.

Visual Display

An early MEMS device successfully used for various display applications is the Texas Instruments digital micromirror device (DMD). The DMD technology can achieve higher performance in terms of resolution, brightness, contrast ratio and convergence than the conventional cathode ray tube and is critical for digital high-definition television applications. A DMD consists of a large array of small mirrors with a typical area of $16\,\mu m \times 16\,\mu m$ as illustrated in Fig. 13.22. A probe tip is shown in the figure for a size comparison. Figure 13.23 shows an SEM micrograph of a close-up view of a DMD pixel array [13.21]. Each mirror is capable of rotating by $\pm 10^\circ$ corresponding to either the *on* or *off* position due to an electrostatic actuation force. Light reflected from any on-mirrors passes through a projection lens and creates images on a large screen. Light from the remaining off-mirrors is reflected away from the projection lens to an absorber. The proportion of time during each video frame that a mirror remains in the on-state determines shades of gray, from black for zero on-time to white for a hundred percent on-time. Color can be added by a color wheel or a three-DMD chip setup. The three DMD chips are used for projecting red, green and blue colors. Each DMD pixel consists of a mirror connected by a mirror support post to an underlying yoke. The yoke, in turn, is connected by torsion hinges to hinge support posts, as shown in Fig. 13.24 [13.22]. The support post and hinges are hidden under the mirror to avoid light diffraction and thus improve contrast ratio

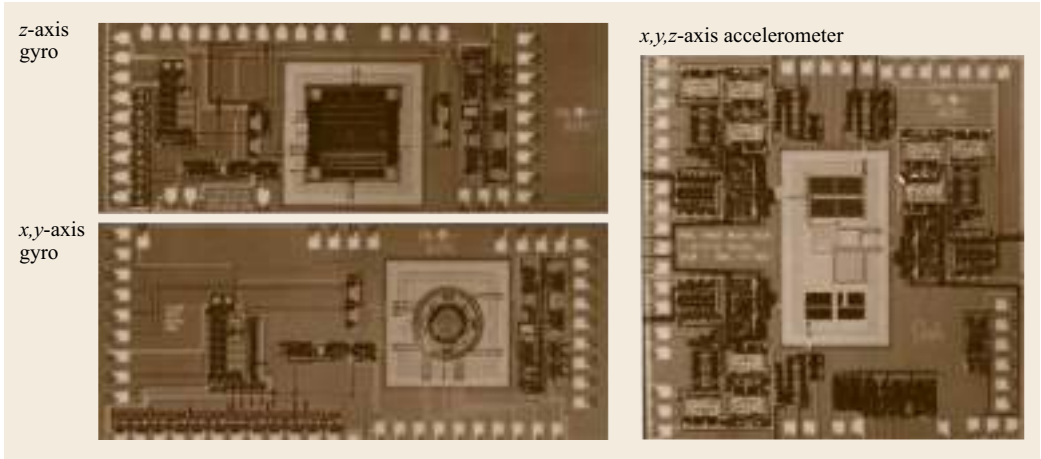


Fig. 13.21 Photo of a surface-micromachined inertial measurement system with a six-degrees sensing capability

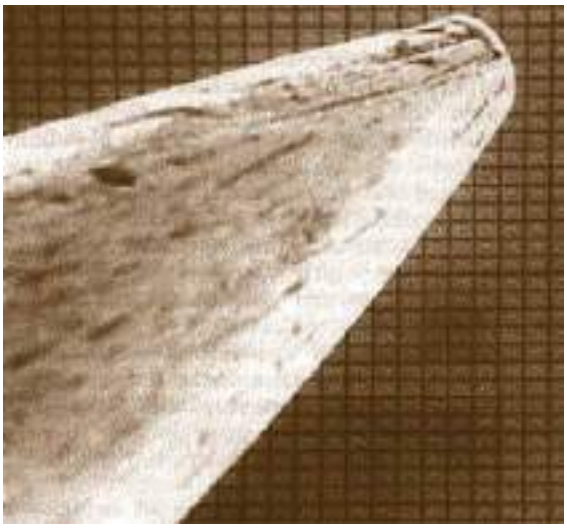


Fig. 13.22 Photo of a digital micromirror device (DMD) array (Copyright Texas Instruments)

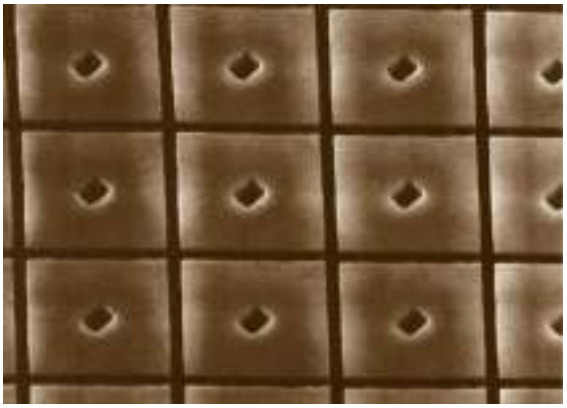


Fig. 13.23 SEM micrograph of a close-up view of a DMD pixel array (after [13.21])

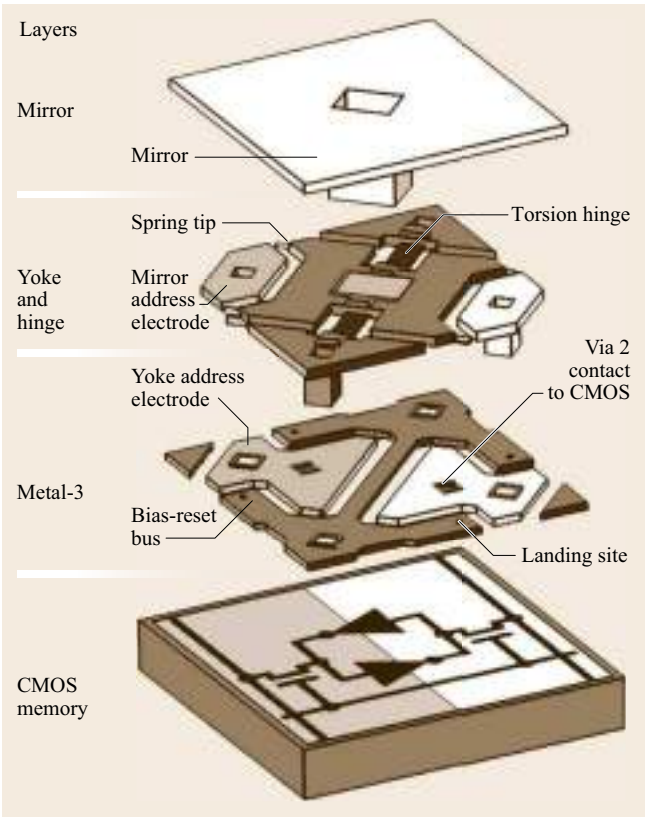


Fig. 13.24 Detailed structure layout of a DMD pixel (after [13.22])



Fig. 13.25 SEM micrograph of a DMD pixel after removing half of the mirror plates using ion milling (Copyright Texas Instruments)

and optical efficiency. There are two gaps on the order of a micrometer, one between the mirror and the underlying hinges and address electrodes, and a second between the coplanar address electrodes and hinges and an underlying metal layer from the CMOS static random access memory (SRAM) structure. The yoke is tilted over the second gap by an electrostatic actuation force, thus rotating the mirror plate. The SRAM determines which angle the mirror needs to be tilted by applying proper actuation voltages to the mirror and address electrodes. The DMD is fabricated using an aluminum-based surface-micromachining technology. Three layers of aluminum thin film are deposited and patterned to form the mirror and its suspension system. Polymer material is used as a sacrificial layer and is removed by a plasma etch at the end of the process to freely release the micromirror structure. The micromachining process is compatible with standard CMOS fabrication, allowing the DMD to be monolithically integrated with a mature CMOS address circuit technology, thus achieving high yield and low cost. Figure 13.25 shows an SEM micrograph of a fabricated DMD pixel revealing its cross section after an ion milling. A close-up view on the yoke and hinge support under the mirror is shown in Fig. 13.26.

Precision Optical Platform

The growing optical communication and measurement industry requires low cost, high-performance optoelectronic modules, such as laser-to-fiber couplers, scanners, and interferometers. A precision alignment and

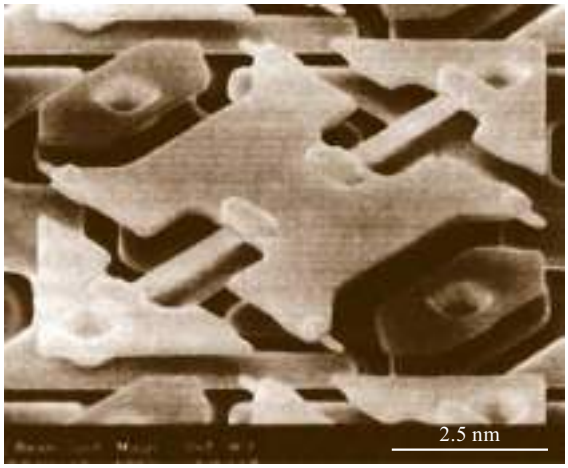


Fig. 13.26 SEM micrograph of a close-up view of a DMD yoke and hinges (after [13.22])

the ability to actuate optical components such as mirrors, gratings, and lenses with sufficient accuracy are critical for high-performance optical applications. Conventional hybrid optical integration approaches, such as the silicon optical bench, suffers from a limited alignment tolerance of $\pm 1\,\mu m$ and also lacks of component actuation capability [13.24, 25]. As a result, only simple optical systems can be constructed with no more than a few components, thus severely limiting the performance. Micromachining, however, provides a critical enabling technology, allowing movable optical components to be fabricated on a silicon substrate. Component movement with high precision can be achieved through electrostatic actuation. By combining micromachined movable optical components with lasers, lenses and fibers on the same substrate, an on-chip complex self-aligning optical system can be developed. Figure 13.27 a shows an SEM micrograph of a surface-micromachined, electrostatically actuated microreflector for laser-to-fiber coupling and external-cavity laser applications [13.23]. The device consists of a polysilicon mirror plate hinged to a support beam. The mirror and the support, in turn, are hinged to a vibromotor-actuator slider. The microhinge technology [13.26] allows the joints to rotate out of the substrate plane to achieve large aspect ratios. Common-mode actuation of the sliders results in a translational motion, while differential slider motion produces an out-of-plane mirror rotation. These motions permit the microreflector to redirect an optical beam in a desirable location. Each of the two slides is actuated with an integrated microvibromotor shown in Fig. 13.27b. The vibromotor consists of four electrostatic comb resonators with attached impact arms driving a slider through oblique impact. The two opposing impactors are used for each

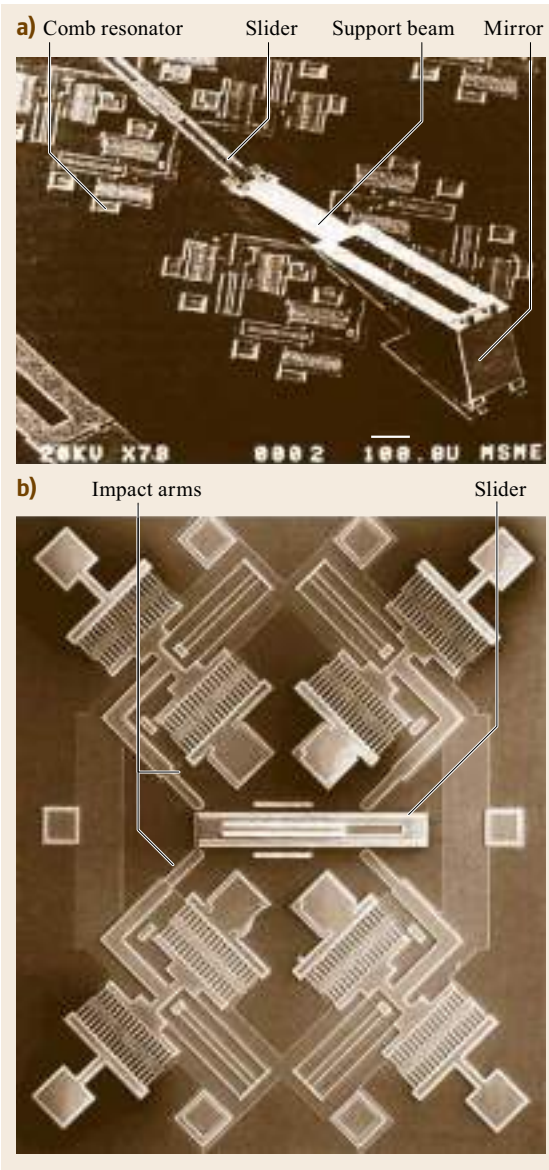


Fig. 13.27 **(a)** SEM micrograph of a surface-micromachined, electrostatically actuated microreflector; **(b)** SEM micrograph of a surface-micromachined vibromotor (after [13.23])

travel direction to balance the forces. The resonator is a capacitively driven mass anchored to the substrate through a folded beam flexure. The flexure compliance determines the resonant frequency and travel range of the resonator. When the comb structures are driven at their resonant frequency (around 8 KHz), the slider exhibits a maximum velocity of over 1 mm/s. Characterization of the vibromotor also shows that a slider step resolution of less than $0.3\,\mu m$ can be achieved [13.27],

making it attractive for precision alignment of various optical components. The prototype microreflector can obtain an angular travel range over 90° and a translational travel range of 60 μm. Using this device, beam steering, fiber coupling and optical scanning have been demonstrated.

Optical Data Switching

High-speed communication infrastructures are highly desirable for transferring and processing real-time voice and video information. Optical fiber communication technology has been identified as the critical backbone to support such systems. A high-performance optical data switching network, which routes various optical signals from sources to destinations, is one of the key building blocks for system implementation. At present, optical signal switching is performed using hybrid optical-electronic-optical (O-E-O) switches. These devices first convert incoming light from input fibers to electrical signals and then route the electrical signals to the proper output ports after signal analyses. At the output ports, the electrical signals are converted back to streams of photons or optical signals for further transmission over the fibers. The O-E-O switches are expensive to build, integrate and maintain. Furthermore, they consume substantial amount of power and introduce additional latency. It is therefore highly desirable to develop an all-optical switching network in which optical signals can be routed without intermediate conversion into electrical form, thus minimizing power dissipation and system delay. While a number of approaches are being considered for building all-optical switches, MEMS technology is attractive because it can provide arrays of tiny movable mirrors which can redirect incoming beams from input fibers to corresponding output fibers. As described in the previous sections, these micromirrors can be batch fabricated using Si micromachining technologies, thus achieving an integrated solution with the potential for low cost. A significant reduction in power dissipation is also expected.

Figure 13.28 shows an architecture of a two-dimensional micromirror array forming a switching matrix with rows of input fibers and columns of output fibers (or vice versa). An optical beam from an input fiber can be directed to an output fiber through activating the corresponding reflecting micromirror. Switches with eight inputs and eight outputs can be readily implemented using this technique, which can be further extended to a 64×64 matrix. The micromirrors are moved between two fixed stops by digital control, thus eliminating the need for precision motion control. Figure 13.29 presents an SEM micrograph of a simple 2×2 MEMS fiber optic switching network prototype for illustration purposes [13.28]. The network includes

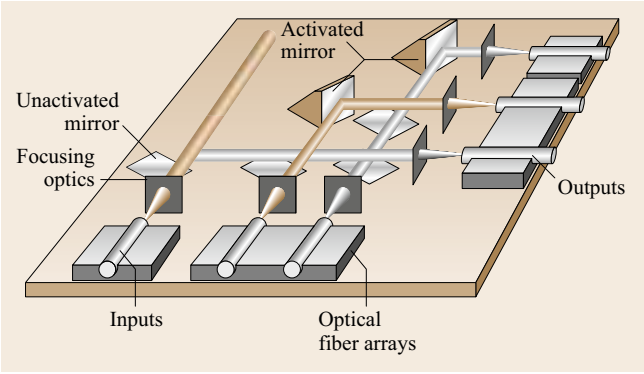


Fig. 13.28 Schematic of a two-dimensional micromirror-based fiber optic switching matrix

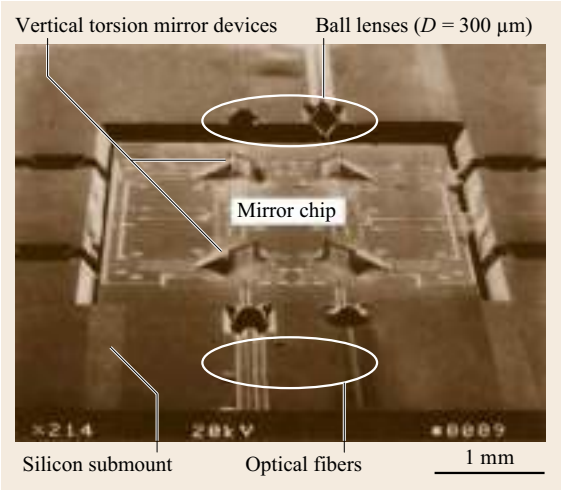


Fig. 13.29 SEM micrograph of a 2×2 MEMS fiber optic switching network (after [13.28])

a mirror chip passively integrated with a silicon submount, which contains optical fibers and ball lenses. The mirror chip consists of four surface-micromachined vertical torsion mirrors. The four mirrors are arranged such that in the *reflection* mode, the input beams are reflected by two 45° vertical torsion mirrors and coupled into the output fibers located on the same side of the chip. In the *transmission* mode, the vertical torsion mirrors are rotated out of the optical paths, thus allowing the input beams to be coupled into the opposing output fibers. Figure 13.30 shows an SEM micrograph of a polysilicon vertical torsion mirror. The device consists of a mirror plate attached to a vertical supporting frame by torsion beams and a vertical back electrode plate. The mirror plate is approximately 200 μm wide, 160 μm long, and 1.5 μm thick. The mirror surface is coated with a thin layer of gold to improve the optical reflectivity. The back plate is used to electrostatically

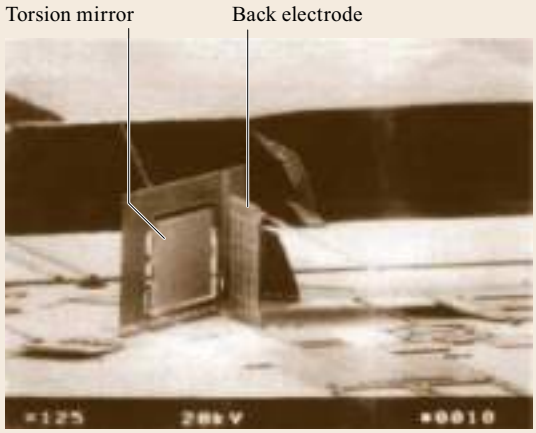


Fig. 13.30 SEM micrograph of a polysilicon surface-micromachined vertical torsion mirror (after [13.28])

actuate the mirror plate so that the mirror can be rotated out of the optical path in the *transmission* mode. Surface micromachining with microhinge technology is used to realize the overall structure. The back electrode plate is integrated with a scratch drive actuator array [13.29] for self-assembly. The self-assembly approach is critical when multiple vertical torsion mirrors are used to implement more advanced functions.

A more sophisticated optical switching network with a large scaling potential can be implemented using a three-dimensional (3-D) switching architecture as shown in Fig. 13.31. The network consists of arrays of two-axis mirrors to steer optical beams from input fibers to output fibers. A precision analog closed-loop mirror position control is required to accurately direct a beam along two angles so that one input fiber can be optically connected to any output fiber. The optical length depends little on which set of fibers are connected, thus achieving a more uniform switching characteristic, which is critical for implementing a large scale network. Two-axis mirrors are the crucial components for implementing the 3-D architecture. Figure 13.32 shows an SEM micrograph of a surface-micromachined two-axis beam-steering mirror positioned using a self-assembly technique [13.30]. The self-assembly is accomplished during the final release step of the mirror processing sequence. Mechanical energy is stored in a special high-stress layer during the deposition, which is put on top of the four assembly arms. Immediately after the assembly arms are released, the tensile stress in this layer causes the arms to bend up, pushing the mirror frame and lifting it above the silicon substrate. All mirrors used in the switching network can be fabricated simultaneously without any human intervention or external power supply.



Fig. 13.31 Schematic of a three-dimensional micromirror-based fiber optic switching matrix

13.1.4 RF MEMS

The increasing demand for wireless communication applications, such as mobile and cordless telephony, wireless data networks, two-way paging and global positioning system, motivates a growing interest in building miniaturized wireless transceivers with multistan-

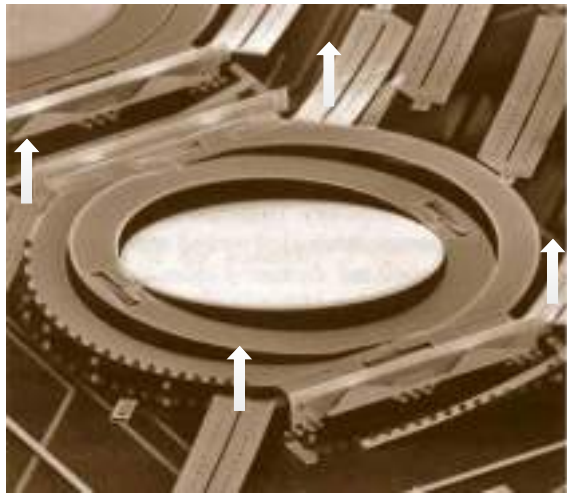


Fig. 13.32 SEM micrograph of a surface-micromachined two-axis beam-steering micromirror positioned using a self-assembly technique (after [13.30])

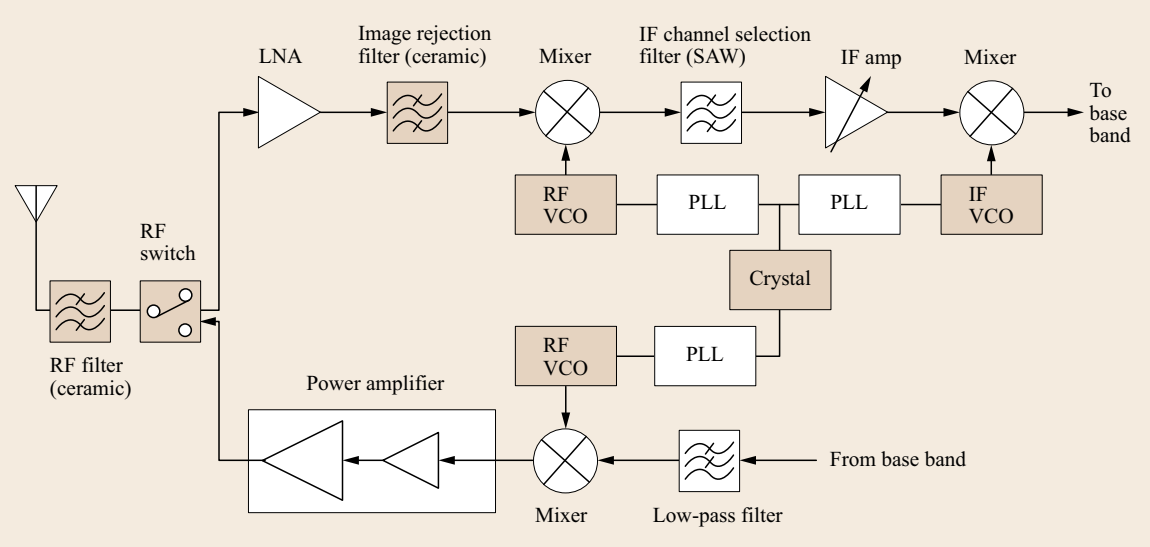


Fig. 13.33 Schematic of a superheterodyne radio architecture; *SAW*: surface acoustic wave, *VCO*: voltage controlled oscillator, *LNA*: low-noise amplifier, *PLL*: phase-locked loop

dard capabilities. Such transceivers will greatly enhance the convenience and accessibility of various wireless services independent of geographic location. Miniaturizing current single-standard transceivers, through a high level of integration, is a critical step towards building transceivers that are compatible with multiple standards. Highly integrated transceivers will also result in reduced package complexity, power consumption and cost. At present, most radio transceivers rely on a large number of discrete frequency-selection components, such as radio-frequency (RF) and intermediate-frequency (IF) band-pass filters, RF voltage-controlled oscillators (VCOs), quartz crystal oscillators and solid-state switches to perform the necessary analog signal processing. Figure 13.33 shows a schematic of a superheterodyne radio architecture, in which discrete components are shaded in dark color. Theses off-chip devices occupy the majority of the system area, thus severely hindering transceiver miniaturization. MEMS technology, however, offers a potential solution to integrate these discrete components onto silicon substrates with microelectronics, achieving a size reduction of a few orders of magnitude. It is therefore expected to become an enabling technology to ultimately miniaturize radio transceivers for future wireless communications.

MEMS Variable Capacitors

Integrated high-performance variable capacitors are critical for low-noise VCOs, antenna tuning, tunable matching networks, etc. Capacitors with high quality factors (Q), a large tuning range and linear characteristics are crucial for achieving system performance requirements. On-chip Si PN junction and MOS-based variable capacitors suffer from low quality factors (below 10 at 1 GHz), limited tuning range and poor linearity, thus are inadequate for building high-performance transceivers. MEMS technology has demonstrated monolithic variable capacitors achieving stringent performance requirements. These devices typically reply on an electrostatic actuation method to vary the air gap between a set of parallel plates [13.31–34], vary the capacitance area between a set of conductors [13.35] or mechanically displace a dielectric layer in an air-gap capacitor [13.36]. Improved tuning ranges have been achieved with various device configurations. Figure 13.34 shows SEM micrographs of an aluminum micromachined variable capacitor fabricated on a silicon substrate [13.31]. The device consists of a $200\,\mu m \times 200\,\mu m$ aluminum plate with a thickness of $1\,\mu m$ suspended above the bottom electrode by an air gap of $1.5\,\mu m$. Aluminum is selected as the structural material due to its low resistivity, which is critical for achieving a high quality factor at high frequencies. A DC voltage applied across the top and bottom electrodes introduces an electrostatic pull-down force, which pulls the top plate towards the bottom electrode, thus changing the device capacitance value. The capacitors are fabricated using aluminum-based surface-micromachining technology. Sputtered aluminum is used for building the capacitor top and bottom electrodes. Photoresist is served as the sacrificial layer, which is then removed through an oxygen-based plasma dry etch to release the microstructure. The processing technology requires a low thermal budget, thus allow-

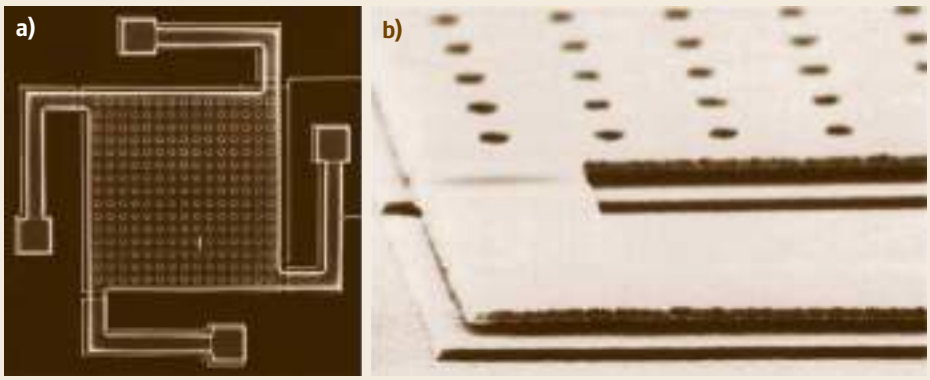


Fig. 13.34 **(a)** SEM micrograph of a top-view of an aluminum surface-micromachined variable capacitor; **(b)** SEM micrograph of a cross-sectional view of the variable capacitor (after [13.31])

ing the variable capacitors to be fabricated on top of wafers with completed electronic circuits without degrading the performance of active devices. Figure 13.35 presents an SEM micrograph of four MEMS tunable capacitors connected in parallel. This device achieves a nominal capacitance value of 2 pF and a tuning range of 15% with 3 V. A quality factor of 62 has been demonstrated at 1 GHz, which matches or exceeds that of discrete varactor diodes and is at least an order of magnitude larger than that of a typical junction capacitor implemented in a standard IC process.

MEMS tunable capacitors based upon varying the capacitance area between a set of conductors have been demonstrated. Figure 13.36 shows an SEM micrograph of such a device [13.35]. The capacitor comprises arrays of interdigitated electrodes, which can be electrostatically actuated to vary the electrode overlap area. A close-up view of the electrodes is shown in Fig. 13.37. The capacitor is fabricated using a silicon-on-insulator (SOI) substrate with a top silicon layer thickness around 20 μm to obtain a high aspect ratio for the electrodes, which is critical for achieving a large capacitance density and reduced tuning voltage. The silicon layer is etched to form the device structure followed by removing the underneath oxide to release the capacitor. A thin aluminum layer is then sputtered over the capacitor to reduce the series resistive loss. The device exhibits a quality factor of 34 at 500 MHz and can be tuned between 2.48 pF and 5.19 pF with an actuation voltage under 5 V, corresponding to a tuning range over 100%. Figure 13.38 shows the variation of electrode overlap area under different tuning voltages.

Tunable capacitors relying on a movable dielectric layer have been fabricated using MEMS technology. Figure 13.39 presents an SEM micrograph of a copper-based micromachined tunable capacitor [13.36]. The device consists of an array of copper top electrodes suspended above a bottom copper plate with an air gap of approximately 1 μm. A thin nitride layer is deposited, patterned, and suspended between the two copper layers



Fig. 13.35 SEM micrograph of four MEMS aluminum variable capacitors connected in parallel (after [13.31])

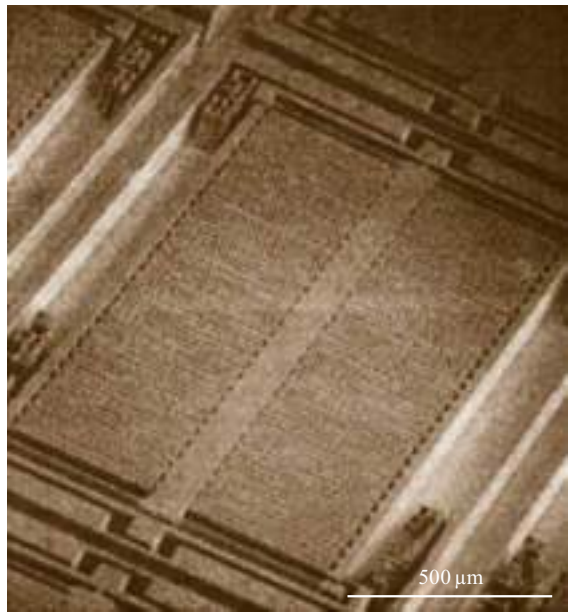


Fig. 13.36 SEM micrograph of a silicon tunable capacitor using comb drive actuator (after [13.35])

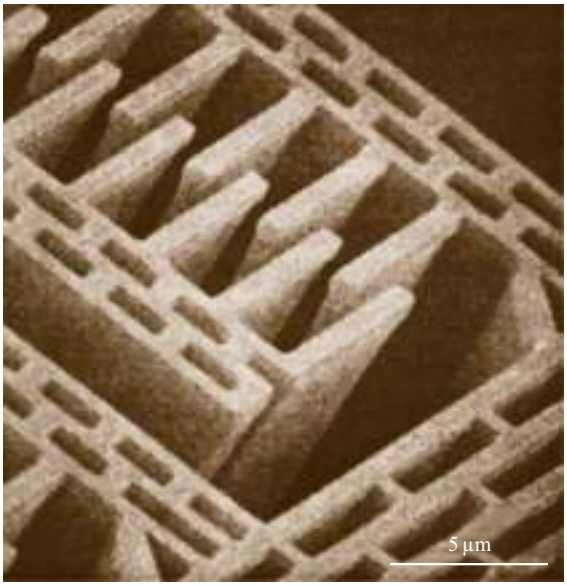


Fig. 13.37 SEM micrograph of a close-up view of a tunable capacitor comb fingers (after [13.35])

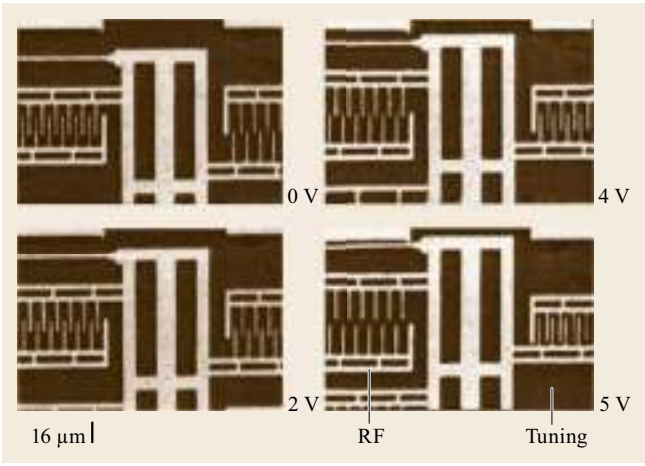


Fig. 13.38 Photos of comb fingers at different actuation voltages (after [13.35])

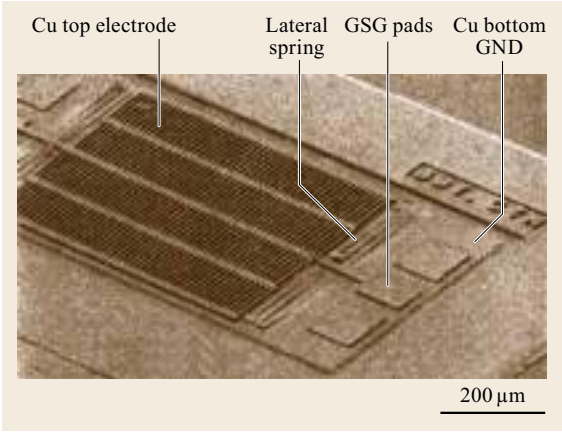


Fig. 13.39 SEM micrograph of a copper surface-micromachined tunable capacitor with a movable dielectric layer (after [13.36]); *GND*: ground, *GSG*: ground-signal-ground

by lateral mechanical spring suspensions after sacrificial release. A DC voltage applied across the copper layers introduces a lateral electrostatic pull-in force on the nitride, thus resulting in a movement that changes the overlapping area between each copper electrode and the bottom plate, and hence the device capacitance. The tunable capacitor achieves a quality factor over 200 at 1 GHz with 1 pF capacitance due to the highly conductive copper layers and a tuning range around 8% with 10 V.

Micromachined Inductors

Integrated inductors with high quality factors are as critical as the tunable capacitors for high-performance RF system implementation. They are the key components for building low-noise oscillators, low-loss matching networks, etc. Conventional on-chip spiral inductors suffer from limited quality factors of around 5 at 1 GHz, an order of magnitude lower than the required values from discrete counterparts. The poor performance is mainly caused by substrate loss and metal resistive loss at high frequencies. Micromachining technology provides an attractive solution to minimize these loss contributions, hence enhancing the device quality factors. Figure 13.40 shows an SEM micrograph of a 3-D coil inductor fabricated on a silicon substrate [13.37]. The device consists of 4-turn 5 µm-thick copper traces electroplated around an insulting core with a 650 µm × 500 µm cross-section. Compared to spiral inductors, this geometry minimizes the coil area, which is in close proximity to the substrate and hence the eddy-current loss, resulting in a maximized Q-factor and device self-resonant frequency. Copper is selected as the interconnect metal because of its low sheet resistance, which is critical for achieving a high Q-factor. The inductor achieves a 14 nH inductance value with a quality factor of 16 at 1 GHz. A single-turn 3-D device exhibits a Q-factor of 30 at 1 GHz, which matches the performance of discrete counterparts. The high-Q 3-D inductor and MEMS tunable capacitors, shown in Fig. 13.35, have been employed to implement a RF CMOS VCO achieving a low phase-noise performance suitable for typical wireless communication applications such as GMS cellular telephony [13.38].

Other 3-D inductor structures such as the levitated spiral inductors have been demonstrated using micromachining fabrication technology. Figure 13.41

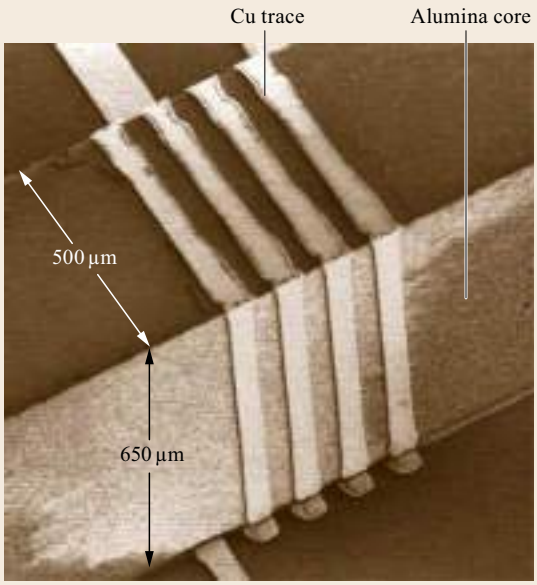


Fig. 13.40 SEM micrograph of a 3-D coil inductor fabricated on a silicon substrate (after [13.37])

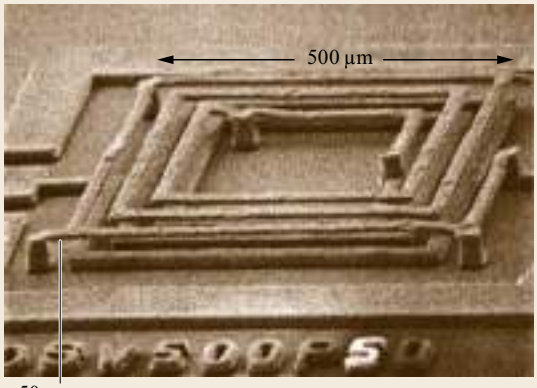


Fig. 13.41 SEM micrograph of a levitated spiral inductor fabricated on a glass substrate (after [13.39])

shows an SEM micrograph of a levitated copper inductor, which is suspended above the substrate through supporting posts [13.39]. The levitated geometry can minimize the substrate loss, thus achieving an improved quality factor. The inductor shown in the figure achieves a 14 nH inductance value with a Q-factor of 38 at 1.8 GHz using a glass substrate.

A self-assembled out-of-plane coil has been fabricated using micromachining technology. The inductor winding traces are made of refractory metals with controlled built-in stress such that the traces can curl out of the substrate surface upon release and interlock into

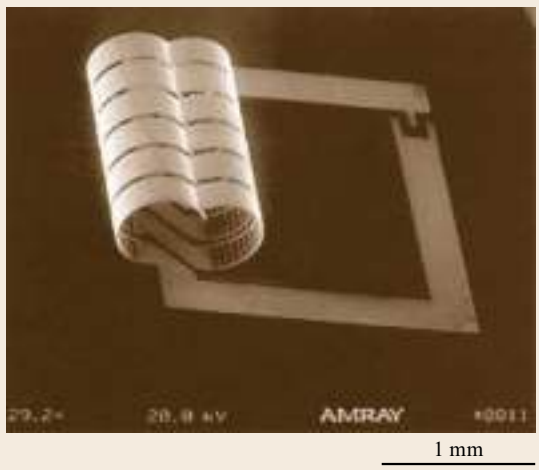


Fig. 13.42 SEM micrograph of a self-assembled out-of-plane coil inductor (after [13.40])

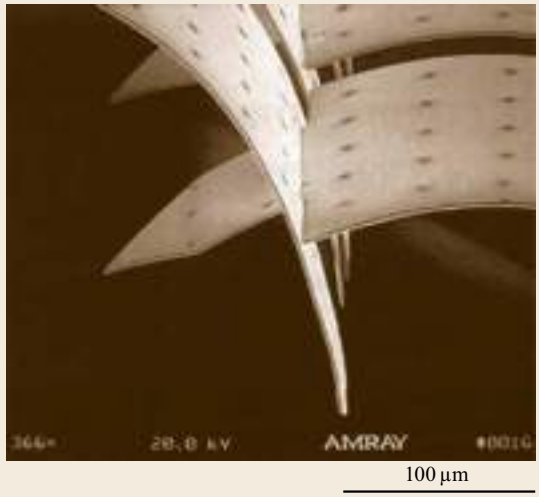


Fig. 13.43 SEM micrograph of an interlocking trace from a self-assembled out-of-plane oil inductor (after [13.40])

each other to form coil windings. Figure 13.42 shows an SEM micrograph of a self-assembled out-of-plane coil inductor [13.40]. A close-up view of an interlocking trace is shown in Fig. 13.43. Copper is plated on the interlocked traces to form highly conductive windings at the end of processing sequence. The inductor shown in Fig. 13.42 achieves a quality factor around 40 at 1 GHz.

MEMS Switches

The microelectromechanical switch is another potentially attractive miniaturized component enabled by micromachining technologies. These switches offer su-

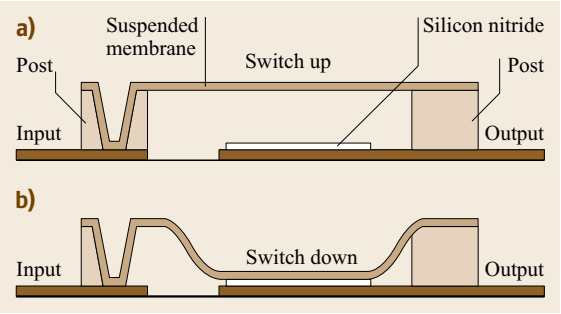


Fig. 13.44a,b Cross-sectional schematics of an RF MEMS capacitive switch; **(a)** Off state, **(b)** On state

perior electrical performance in terms of insertion loss, isolation, linearity, etc., and are intended to replace off-chip solid state counterparts, which provide switching between the receiver and transmitter signal paths. They are also critical for building phase shifters, tunable antennas, and filters. The MEMS switches can be characterized into two categories: capacitive and metal-to-metal contact types. Figure 13.44 presents a cross-sectional schematic of an RF MEMS capacitive switch. The device consists of a conductive membrane, typically made of aluminum or gold alloy suspended above a coplanar electrode by an air gap of a few micrometers. For RF or microwave applications, actual metal-to-metal contact is not necessary; rather, a step change in the plate-to-plate capacitance realizes the switching function. A thin silicon nitride layer with a thickness on the order of 1000 Å is typically deposited above the bottom electrode. When the switch is in the on-state, the membrane is high, resulting in a small plate-to-plate capacitance, and hence a minimum high-frequency signal coupling (high isolation) between the two electrodes. In the off-state (with a large enough applied DC voltage), the switch provides a large capacitance due to the thin dielectric layer, thus causing a strong signal coupling (low insertion loss). The capacitive switch consumes near-zero power, which is attractive for low-power portable applications. Switching cycles over millions for this type of device have been demonstrated. Figure 13.45 shows a top-view photo of a fabricated MEMS capacitive switch [13.41]. Surface-micromachining technology, using metal for the electrodes and polymer as the sacrificial layer, is used to fabricate the device. The switch can be actuated with a DC voltage on the order of 50 V and exhibits a low insertion loss of approximately −0.28 dB at 35 GHz and a high isolation of −35 dB at the same frequency.

Metal-to-metal contact switches are important for interfacing large bandwidth signals including DC. This type of device typically consists of a cantilever beam

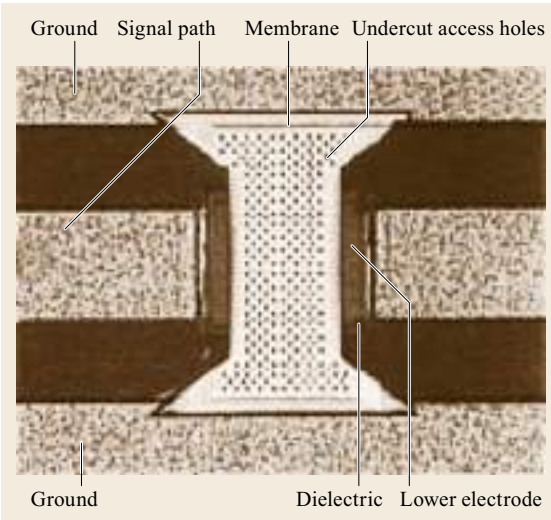


Fig. 13.45 Top-view photo of a fabricated RF MEMS capacitive switch (after [13.41])

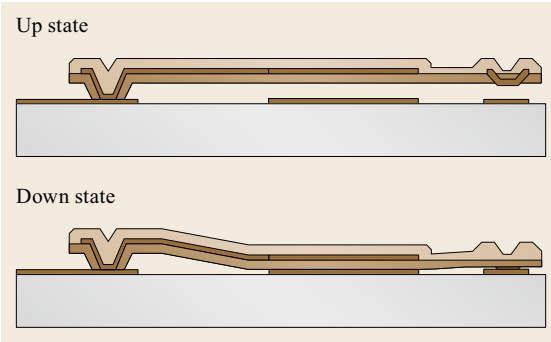


Fig. 13.46 Cross-sectional schematic of a metal-to-metal contact switch (after [13.44])

or clamped-clamped bridge with a metallic contact pad positioned at the beam tip or underneath the bridge center. Through an electrostatic actuation, a contact can be formed between the suspended contact pad and an underlying electrode on the substrate [13.42–44]. Figure 13.46 shows a cross-sectional schematic of a metal-to-metal contact switch [13.44]. The top-view of the fabricated device is presented in Fig. 13.47. The switch exhibits an actuation voltage of 30 V, a response time of 20 μs and mechanical strength to withstand 10^9 actuations. An isolation greater than 50 dB below 2 GHz and insertion loss less than 0.2 dB from DC through 40 GHz has been demonstrated.

MEMS Resonators

Microelectromechanical resonators based upon poly-Si comb-drive structures, suspended beams and center-pivoted disk configurations have been demonstrated



Fig. 13.47 Top-view photo of a fabricated metal-to-metal contact switch (after [13.44])

for performing analog signal processing [13.45–49]. These microresonators can be excited into mechanical resonance through an electrostatic drive. The mechanical motion causes a change of device capacitance resulting in an output electrical current when a proper DC bias voltage is used. The output current exhibits the same frequency as the mechanical resonance, thus achieving an electrical filtering function through the electromechanical coupling. Micromachined polysilicon flexural-mode mechanical resonators have demonstrated a quality factor greater than 80 000 in a 50 μTorr vacuum [13.50]. This level of performance is comparable to a typical quartz crystal and is thus attractive for implementing monolithic low-noise and low-drift reference signal sources. Figure 13.48 shows an SEM micrograph of a surface-micromachined comb-drive resonator integrated with CMOS sustaining electronics on a same substrate to form a monolithic high-Q MEMS resonator-based oscillator [13.45]. The oscillator achieves an operating frequency of 16.5 KHz with a clean spectral purity. A chip area of approximately 420 μm × 230 μm is consumed for fabricating the overall system, representing a size reduction by orders of magnitude compared to conventional quartz crystal oscillators.

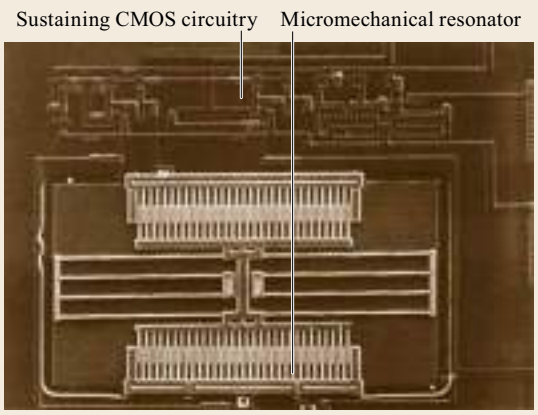


Fig. 13.48 SEM micrograph of a surface-micromachined comb-drive resonator integrated with CMOS sustaining electronics (after [13.45])

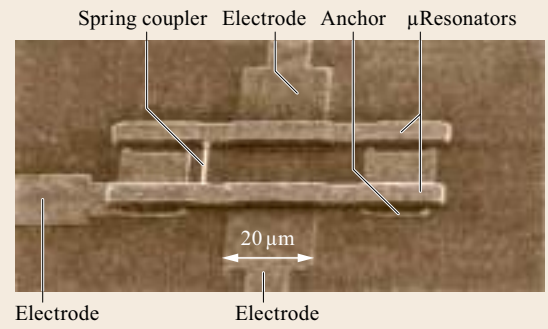


Fig. 13.49 SEM micrograph of a polysilicon surface-micromachined two-resonator spring-coupled bandpass micromechanical filter (after [13.47])

Micromachined high-Q resonators can be coupled to implement low-loss frequency selection filters. Figure 13.49 shows an SEM micrograph of a surface-micromachined polysilicon two-resonator, spring-coupled bandpass micromechanical filter [13.47]. The filter consists of two silicon micromechanical clamped-clamped beam resonators, coupled mechanically by a soft spring, all suspended 0.1 μm above the substrate. Poly-Si strip lines underlie the central regions of each resonator and serve as capacitive transducer electrodes positioned to induce resonator vibration in a direction perpendicular

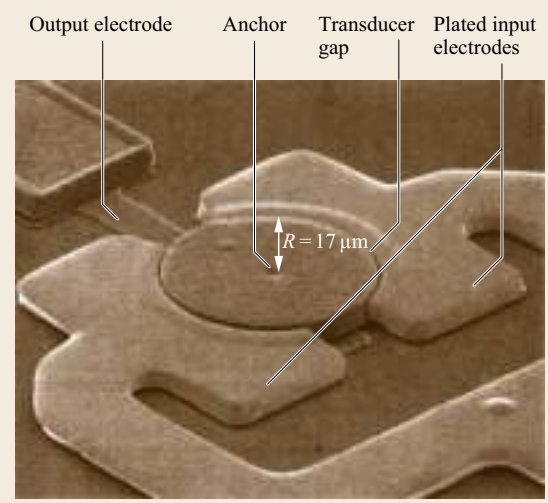


Fig. 13.50 SEM micrograph of a polysilicon surface-micromachined contour-mode disk resonator (after [13.49])

to the substrate. Under normal operation, the device is excited capacitively by a signal voltage applied to the input electrode. The output is taken at the other end of the structure, also by capacitive transduction. The filter achieves a center frequency of 7.81 MHz, a bandwidth of 0.23% and an insertion loss less than 2 dB. The achieved performance is attractive for implementing filters in the low MHz range.

To obtain a higher mechanical resonant frequency with low losses, a surface-micromachined contour-mode disk resonator has been proposed, as shown in Fig. 13.50 [13.49]. The resonator consists of a polysilicon disk suspended 5000 Å above the substrate with a single anchor at its center. Plated metal input electrodes surround the perimeter of the disk with a narrow separation of around 1000 Å, which defines the capacitive, electromechanical transducer of the device. To operate the device, a DC bias voltage is applied to the structure with an AC input signal applied to the electrodes, resulting in a time- varying electrostatic force acting radially on the disk. When the input signal matches the device resonant frequency, the resulting electrostatic force is amplified by the Q-factor of the resonator, producing expansion and contraction of the disk along its radius. This motion in turn produces a time-varying output current at the same frequency, thus achieving the desirable filtering. The prototype resonator demonstrates an operating frequency of 156 MHz with a Q-factor of 9400 in vacuum. The increased resonant frequency is comparable to the first intermediate frequency used in a typical wireless transceiver design and is thus suitable for implementing IF bandpass filters. A higher operating frequency is needed in order to potentially replace the RF frequency selection filters with microelectromechanical resonators.

13.2 NEMS Devices and Applications

Nanoelectromechanical systems (NEMS), if broadly defined, should include all structures and devices that satisfy (i) having both *electro* (conducting and/or semiconducting) and *mechanical* (movable) parts *and* having an *electromechanical transduction* function, and (ii) having at least one characteristic dimension in the $\approx 1-100$ nm size scale, according to the official National Nanotechnology Initiative (NNI) convention. For example, a simple cantilever- or beam-structured nanoelectromechanical resonator possesses both *electro* (resistance) and *mechanical* (suspended, movable beam) parts, and the beam's mechanical motion can be converted into an electrical signal and be read out with certain transduction schemes. Thus it is viewed as a NEMS device.

NEMS as an emerging and attractive field has very interesting historical perspectives, which could have important implications in transforming NEMS into enabling technologies. (i) From a physics viewpoint, the emergence of NEMS originated from some mesoscopic physics studies back in the late 1980s [13.51–53]. Even at that time, sub-100 nm and sub-10 nm structures (mesoscopic *systems*) were made by advanced nanofabrication for electronic transport measurements. In hindsight, it almost seemed natural to take further steps to suspend these mesoscopic systems and to excite them into mechanical motions for exploration of new physics and engineering. This has proven to be a significant and successful evolution. (ii) From the engineering point of view, NEMS could be viewed as MEMS further shrunk down to the nanoscale. This view is helpful in NEMS engineering as one can always first learn from the many techniques developed for MEMS (especially in fabrication processes and signal transduction), with the hope of benefiting the explorations with NEMS. A few classical and helpful introductory reviews of NEMS with intermingled physics and engineering viewpoints and focuses can be found in the literature [13.51, 54–58].

13.2.1 Fabrication Techniques for NEMS

Today, most genuine NEMS are based on nanostructures made either by *top-down* lithographical nanopatterning and nanomachining of thin films of crystalline materials that are also popular in MEMS (e.g., Si, SiC, AlN), or by *bottom-up* chemical synthesis and assembly, such as nanotubes (NTs), nanowires (NWs), nanoflakes, and even nanoparticles. In general, *top-down* approaches utilize nanometer-scale lithographic techniques to define and *carve* mechanically movable nanostructures from crystalline thin films (grown/deposited on substrates) or thick (*bulk*) substrates. *Bottom-up* approaches, meanwhile, build up nanoscale structures and devices in much the same way that mother nature constructs objects, by sequentially assembling atomic and molecular building blocks. During the last two decades, while many have optimistically envisioned or argued that bottom-up approaches will eventually dominate the fabrication of nanoelectronics and NEMS, most of the functional NEMS to date have been created by utilizing top-down techniques

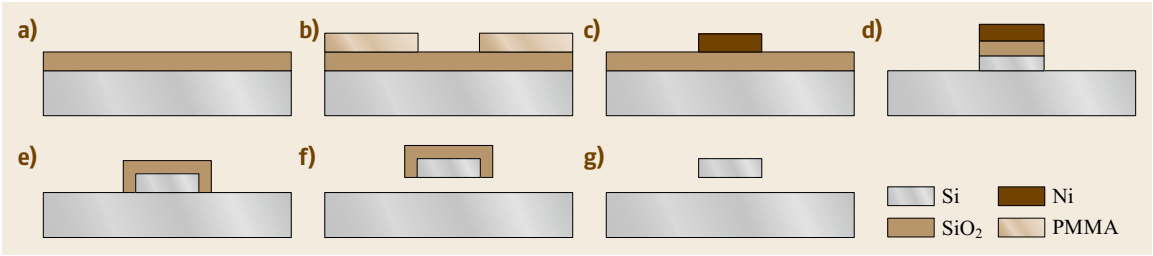


Fig. 13.51a–g Cross-sectional schematics of a process to fabricate nanomechanical beams directly from a silicon substrate: **(a)** after thermal oxidation; **(b)** after PMMA coating and electron-beam lithography; **(c)** after Ni deposition and lift-off; **(d)** after anisotropic SiO_2 and Si etching and Al removal; **(e)** after Ni etching, thermal oxidation and anisotropic SiO_2 etching; **(f)** after release by isotropic Si etching; **(g)** after oxide removal

that combine existing process technologies, such as submicron electron-beam lithography, mainstream film growth/deposition and chemical etching. Top-down approaches make integration with microscale packaging relatively straightforward since the only significant difference between the nanoscale and microscale processing steps is the method used to pattern the various features. We see exciting opportunities keep emerging in the regime where top-down meets bottom-up: many novel nanostructures and NEMS devices can be enabled by devising schemes to combine the bottom-up and top-down techniques, and leveraging their features.

Focusing on the majority and state of the art, NEMS has followed a developmental path similar to the route taken in the development of MEMS in that both have leveraged existing processing techniques. For instance, the electron-beam lithographic techniques used in top-down NEMS fabrication are converging with the techniques that have become standard in the fabrication of submicron ICs. Furthermore, the materials used in many of the first generation top-down NEMS devices (Si, GaAs, Si_3N_4, SiC) were first used in ICs and then in MEMS. Like the first MEMS devices, the first generation NEMS structures consist of free-standing nanomechanical beams, paddle oscillators, and tethered plates made using simple bulk and single-layer surface *nano*machining processes. As an introduction to this new and exciting class of devices, the following section reviews the development of some of these first-generation NEMS devices.

Like Si MEMS, Si NEMS capitalize on well-developed processing techniques for Si and the availability of high-quality substrates. *Cleland* et al. [13.59] reported a relatively simple process to fabricate nanomechanical clamped-clamped beams directly from single crystal (100) Si substrates. As illustrated in Fig. 13.51, the process begins with the thermal oxidation of a Si substrate (Fig. 13.51a). Large Ni contact pads were then fabricated using optical lithography and lift-off. A polymethyl methacrylate (PMMA) lift-off mold was then deposited and patterned using electron-beam lithography into the shape of nanomechanical beams (Fig. 13.51b). Ni was then deposited and patterned by lifting off the PMMA (Fig. 13.51c). Next, the underlying oxide film was patterned by RIE using the Ni film as an etch mask. After oxide etching, nanomechanical beams were patterned by etching the Si substrate using RIE, as shown in Fig. 13.51d. Following Si RIE, the Ni etch mask was removed and the sidewalls of the Si nanomechanical beams were lightly oxidized in order to protect them during the release step (Fig. 13.51e). After performing an anisotropic SiO_2 etch to clear any oxide from the field areas, the Si beams were released using an isotropic Si RIE step, as shown in Fig. 13.51f. After release, the protective SiO_2 film was removed by wet etching in HF (Fig. 13.51g). Using this process, the authors reported the successful fabrication of nanomechanical Si beams with micron-scale lengths ($\approx 8\,\mu m$) and submicron widths (330 nm) and heights (800 nm).

The advent of SOI substrates with high-quality, submicron-thick Si top layers enables the fabrication of nanomechanical Si beams with fewer processing steps than the aforementioned technique since the buried oxide layer makes these device structures relatively easy to pattern and release. Additionally, the buried SiO_2 layer electrically isolates the beams from the substrate. *Carr* et al. [13.60] detail a process that uses SOI substrates to fabricate submicron clamped-clamped mechanical beams and suspended plates with submicron tethers. The process, presented in Fig. 13.52, begins with the deposition of PMMA on an SOI substrate. The SOI substrate has a top Si layer that was either 50 nm or 200 nm in thickness. The PMMA is patterned into a metal lift-off mask by electron beam lithography (Fig. 13.52b). An aluminum film is then deposited and patterned by lift-off into a Si etch mask, as shown in Fig. 13.52c. The nanomechanical beams are then patterned by Si RIE and released by etching the underlying SiO_2 in a buffered hydrofluoric acid solution as shown

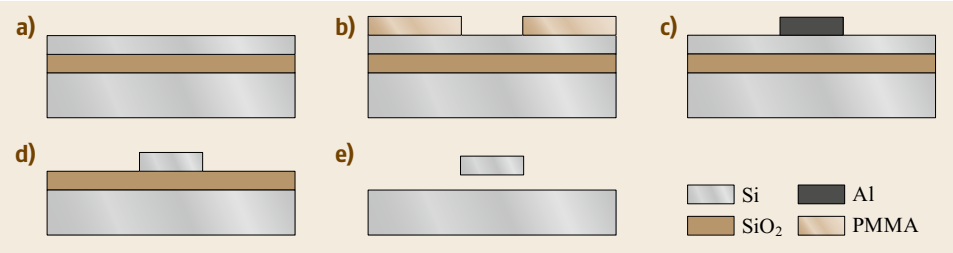


Fig. 13.52a–e Cross-sectional schematics of a process to fabricate nanomechanical structures using silicon-on-insulator (SOI) substrates: **(a)** starting SOI substrate; **(b)** after coating with PMMA and electron-beam lithography; **(c)** after Al deposition and lift-off; **(d)** after anisotropic Si etching and Al removal; **(e)** after release by SiO_2 etching

in Fig. 13.52d and Fig. 13.52e, respectively. Using this process, nanomechanical beams that were 7–16 μm in length, 120 to 200 nm in width and 50 or 200 nm in thickness were successfully fabricated.

Fabrication of NEMS structures is not limited to Si. In fact, III-V compounds, such as gallium arsenide (GaAs), make particularly good NEMS materials from a fabrication perspective because thin epitaxial GaAs films can be grown on lattice-matched materials that can be used as sacrificial release layers. A collection of clamped-clamped nanomechanical GaAs beams fabricated on lattice-matched sacrificial layers having micron-scale lengths and submicron widths and thicknesses is shown in Fig. 13.53. *Tighe* et al. [13.61] reported on the fabrication of GaAs plates suspended with nanomechanical tethers. The structures were made from single crystal GaAs films that were epitaxially grown on aluminum arsenide (AlAs) sacrificial layers. Ni etch masks were fabricated using electron beam lithography and lift-off as described previously. The GaAs films were patterned into beams using a chemically assisted ion beam etching (CAIBE) process and released using a highly selective AlAs etchant. Later, considerable efforts were made in engineering epitaxial GaAs heterostructures for making piezoelectric NEMS [13.62, 63], while GaAs has limited piezoelectric coupling coefficient and compromised elastic and acoustic properties (e.g., relatively low elastic modulus and speed of sound). It was then desirable to explore crystals with higher piezoelectric coupling efficiency, such as AlN, which has been more established in the MEMS areas [13.64, 65]. Among the principal challenges toward exploiting AlN for NEMS are obtaining nanometer-thick AlN films that retain excellent piezoelectric properties and devising efficient actuation and readout schemes. As very thin AlN piezolayers (such as down to 100 nm or sub-100 nm) are usually made to be sandwiched between two similarly thin electrode layers (e.g., Mo or Pt), considerable efforts have been made in designing and implementing multimask surface-nanomachining processes with nanoscale patterning techniques (e.g., electron-beam lithography) to enable AlN NEMS with reliable electrical access to both top and bottom electrode layers. For AlN layers down to 100 nm and thinner, very high-frequency (VHF) NEMS resonators have been demonstrated and it has been calibrated that the piezoelectric coupling coefficients are fairly close to that of bulk AlN crystals [13.66, 67].



Fig. 13.53 SEM micrograph of a set of nanomechanical GaAs clamped-clamped beam resonators (courtesy of Prof. Michael Roukes – Caltech)

As discussed above, it has been the combination and cooperation between both the bottom-up and top-down techniques that has enabled many milestone and representative devices to advance the field of NEMS. Nanowire NEMS is a family of important emerging NEMS devices for many emerging applications, promised by a number of attractive optical, electronic, mechanical and thermal properties of such wires. Regarding these devices, bottom-up grown nanowires can have some unique attributes such as being single-crystalline, atomically smooth and having naturally terminated surfaces [13.68]. In assembling high-quality crystals and as-grown nanowires into functional NEMS, an essential means is exploitation of both bottom-up and top-down nanofabrication and careful engineering in the regime

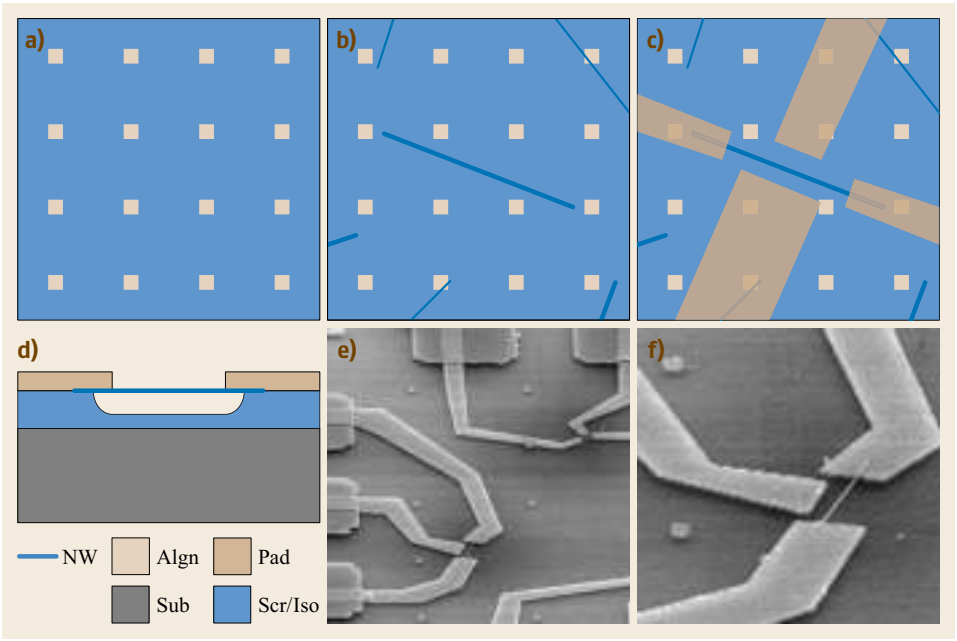


Fig. 13.54a–f Illustration of the fabrication process of conventional bottom-up chemically synthesized NWs and NW-based 1D NEMS devices. **(a)** Electron-beam lithography (EBL) patterned alignment (Algn) markers. **(b)** NWs are distributed on the substrate and coordinates are identified. **(c)** EBL patterning of electrodes. **(d)** Device suspension. **(e,f)** SEM images of representative devices. Reprinted with permission of CRC Press from [13.69] permission conveyed through Copyright Clearance Center, Inc.

where they meet [13.68]. Figure 13.54 briefly illustrates the process of making functional and high-quality nanowire NEMS using bottom-up chemically synthesized nanowires.

As examples of top-down fabricated NEMS and arrays of NEMS, Fig. 13.55 shows a small array of SiC NWs with in-plane coupled gate electrodes. These 50 nm-thin SiC NWs make robust resonators and demonstrate interesting frequency tunability and nonlinear behavior. These NWs and arrays can be interrogated by either optical or all-electronic signal transduction techniques. Engineering such NW devices may help the demonstrations of novel coupled and tunable NW resonator arrays, and arrayed NW NEMS switches.

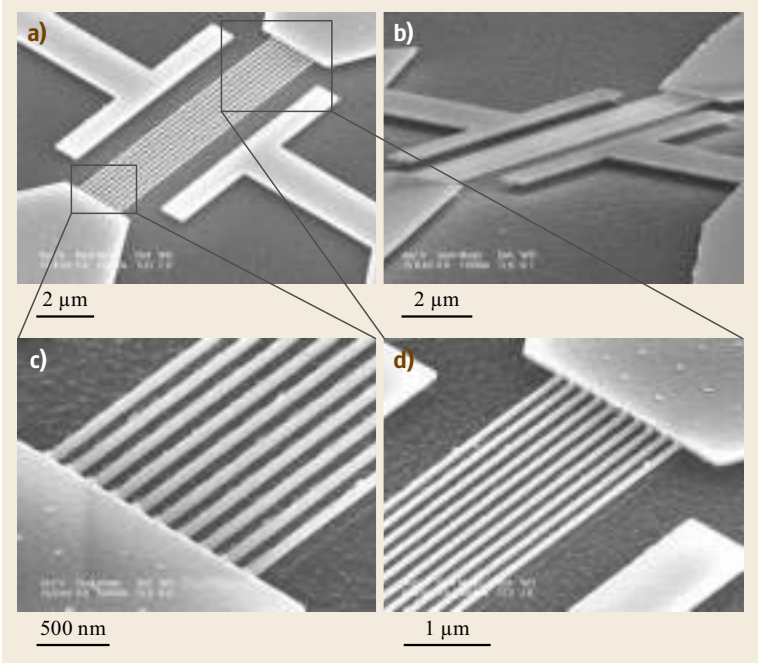


Fig. 13.55a–d Top-down lithographically patterned NW NEMS resonators and arrays. **(a,b)** Aerial views of device arrays. **(c,d)** Close-up views of an array of 10 NWs (50 nm-wide, 10 µm-long). The particular devices shown here have exhibited multiple resonances at ≈ 10 MHz with Qs of ≈ 4000 at room temperature. Reprinted with permission of CRC Press from [13.69] permission conveyed through Copyright Clearance Center, Inc.

13.2.2 NEMS Measurement and Signal Transduction

Several unique approaches have been developed to actuate and sense the motion of NEMS devices. Electrostatic actuation can be used to actuate beams [13.70], tethered meshes [13.71] and paddle oscillators [13.72]. *Sekaric* et al. [13.73] recently showed that low-power lasers can be used to drive paddle oscillators into self-oscillation by induced thermal effects on the structures. In these examples, an optical detection scheme based on the modulation of incident laser light by a vibrating beam is used to detect the motion of the beams. *Cleland* et al. [13.59] described a magnetomotive transduction technique that capitalizes on a time-varying Lorentz force created by an alternating current in the presence of a strong magnetic field. In this case, the nanomechanical beam is positioned in the magnetic field so that an AC current passing through the beam is transverse to the field lines. The resulting Lorentz force causes the beam to oscillate, which creates an electromotive force along the beam that can be detected as a voltage. Thus, in this method, the excitation and detection are performed electrically. In all of the abovementioned cases, the measurements were performed in vacuum, presumably to minimize the effects of squeeze-film damping as well as mass loading due to adsorbates from the environment.

13.2.3 NEMS Devices and Emerging Applications

For the most part, NEMS technology is still in the initial stage of development. Technological challenges related to fabrication and packaging will require innovative solutions before such devices make a significant commercial impact. Nevertheless, NEMS devices have already been used for precision measurements [13.74] enabling researchers to probe the properties of matter [13.75, 76] on a nanoscopic level. Whether or not NEMS makes an impact in the commercial sector, they will undoubtedly prove to be useful platforms for a host of experiments and scientific discoveries in fields ranging from physics to biology.

Radio Frequency (RF) NEMS Resonators and Oscillators

It has been technologically compelling and challenging to pursue RF NEMS resonators and self-sustained feedback oscillators. Intrinsically, aggressive scaling from MEMS to NEMS promises to further scale up the operating frequency and greatly reduce the required operating power levels for RF/microwave signal processing functions. Important initiative explorations have been made in developing RF/microwave NEMS resonators while maintaining high Qs. Given that crystalline materials often possess decent or high speeds of sound, simply scaling the device dimensions can very effectively scale up the device operating frequency up to the RF and microwave bands. Among many crystalline materials for making NEMS, SiC has clearly stood out for a unique combination of outstanding mechanical, electrical, optical, and thermal properties (e.g., superior over those of Si), with wafer-scale micro/nanomachining (e.g., well beyond the manufacturability of diamond). SiC NEMS were the first to break the GHz barrier, with mechanical resonators that operate at > 1 GHz in their fundamental flexural modes. Figure 13.56 highlights the experimental results from the first GHz SiC NEMS resonators. Given a specific mode shape, such as the fundamental flexural mode of a doubly-clamped beam, simply scaling the device length can efficiently scale up the operating frequency. Nonetheless, the Q factor of this given mode is often observed to decrease with the increasing frequency. This is particularly true in the VHF/ultrahigh-frequency (UHF)/microwave regime. It has been revealed that clamping losses in such flexural resonators can be important and even strongly dominating the energy dissipation. To mitigate the clamping losses, an effective approach is to emulate the tuning forks, and design mechanical structures and their modes, so that they are supported or clamped at the nodal points. Figure 13.57 demonstrates one of the first free-free beam VHF NEMS resonators and their in-plane mode shape. It has been shown that given the same dimensions and same fabrication process, free-free beam NEMS resonators can obtain Q factors enhanced by a factor of two or even higher, as compared to their doubly-clamped counterparts [13.77].

Active oscillators spontaneously generate self-sustaining periodic signals by extracting power from steady (DC) sources. This distinguishes them from passive resonators that are characterized by a damped response to impulsive stimuli, which makes them invaluable for applications in precision timekeeping, communications and sensing, which require continuous AC signals. Simply and more often from an electrical engineers' viewpoint, an RF self-sustaining, active oscillator converts DC power into RF power, and it functions as an RF signal generator. Over the last two decades, there has been considerable incentive to miniaturize mechanical resonators to replace conventional crystal resonators in order to integrate frequency selection and tuning functions on-chip with ICs.

SiC NEMS resonators have been exploited to make the first NEMS-referenced, self-sustained feedback oscillator at UHF. Figure 13.58 shows an autonomous

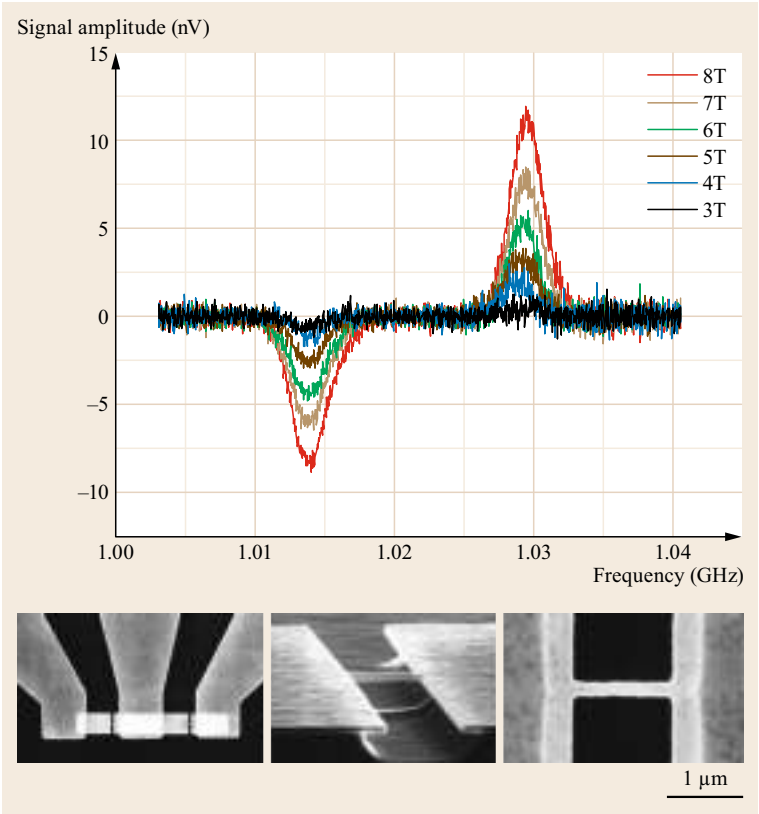


Fig. 13.56 The first resonant NEMS that beat the 1 GHz limit. The data plot shows the electrically measured ≈ 1 GHz NEMS resonances with amplitude dependence on sweeping excitation strength. The SEM images show the doubly-clamped beam SiC NEMS GHz resonators (after [13.78])

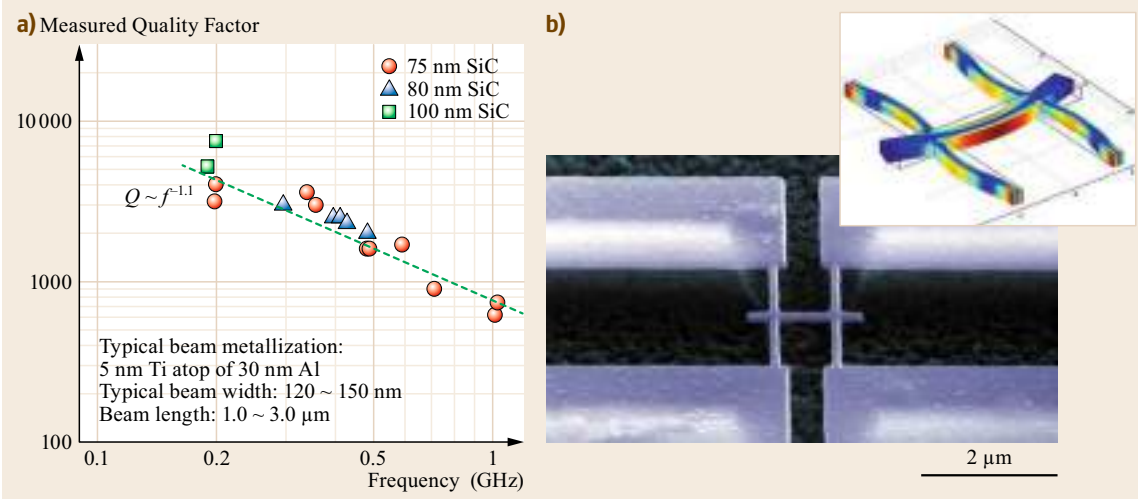


Fig. 13.57a,b Engineering with the frequency versus Q tradeoff. **(a)** The tradeoff between the resonance frequency and the Q factor of the fundamental mode resonance of doubly-clamped beam NEMS resonators, as the devices are scaled down in dimensions to scale up in frequency. **(b)** SEM image of a free-free beam VHF NEMS resonator (at ≈ 175 MHz) for reducing the clamping losses to boost Q [13.77]. The *inset* shows the in-plane free-free beam mode shape

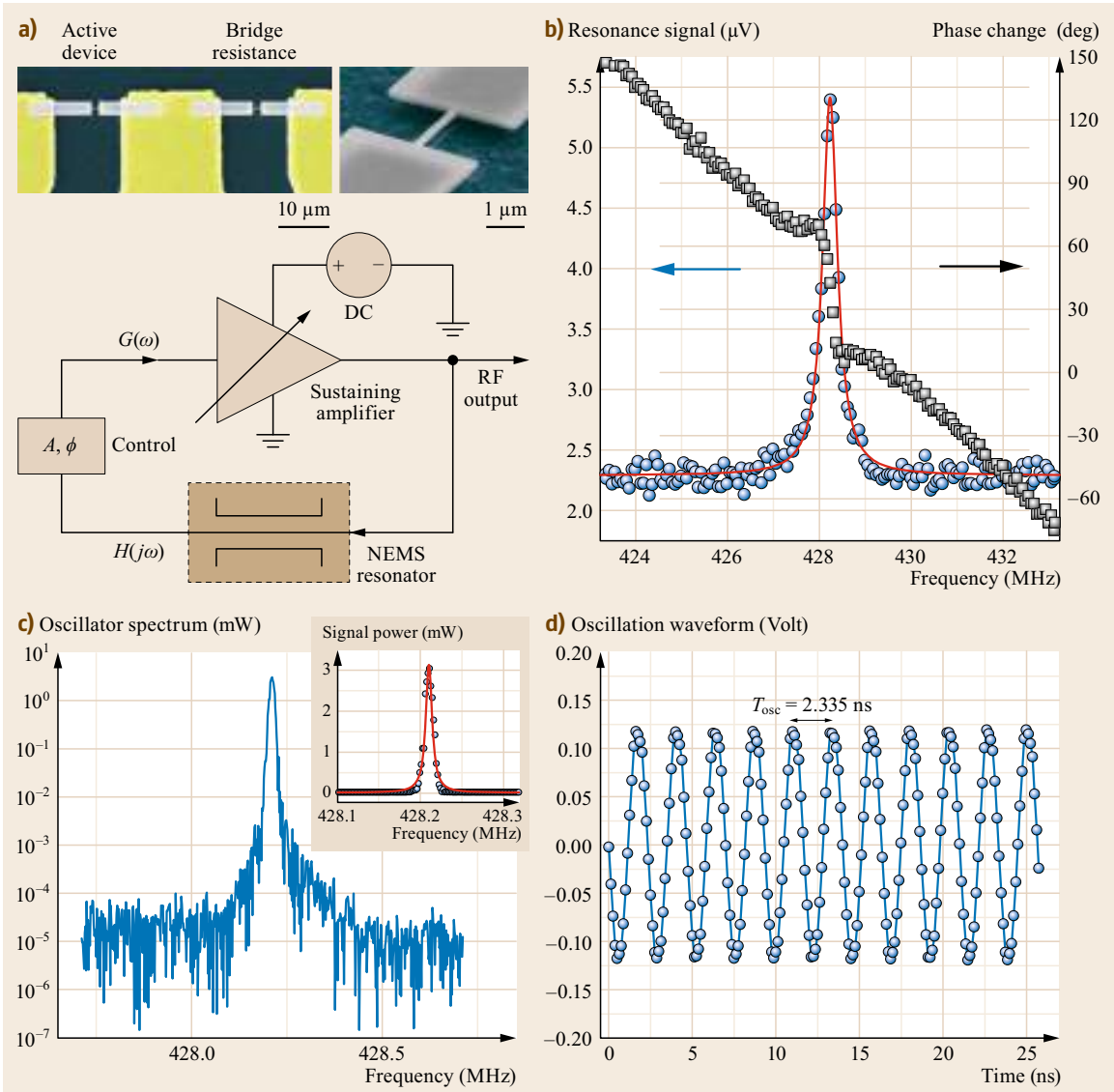


Fig. 13.58a–d Demonstration of an ultra-high frequency (UHF) self-sustaining electrical oscillator referred to a NEMS resonator. **(a)** Simplified feedback oscillator circuitry. The *inset* shows SEM images of UHF NEMS. **(b)** Open-loop resonance response of a 428 MHz NEMS resonator. **(c)** Output power spectrum of the NEMS oscillator (log scale). The *inset* shows the power spectrum on a linear scale and theoretical fit. **(d)** Stable time-domain oscillation waveforms of the closed-loop NEMS oscillator (after [13.79])

and self-sustaining NEMS oscillator that generates continuous UHF signals when powered by a steady DC source. The frequency-determining element in the oscillator is a 428 MHz NEMS resonator that is embedded within a tunable electrical feedback network to generate active and stable self-oscillation [13.79]. Beyond this demonstration, careful measurements and analyses have also been done on the frequency stability and phase noise performance of this UHF NEMS oscillator. An interesting linewidth-narrowing effect has also been observed in the self-sustaining UHF NEMS oscillator [13.79].

NEMS Resonators: From Quantum Limits to Resonant Sensors

Since the early days of explorations from the mesoscopic physics studies, NEMS have already been playing increasingly important roles as enabling tools for

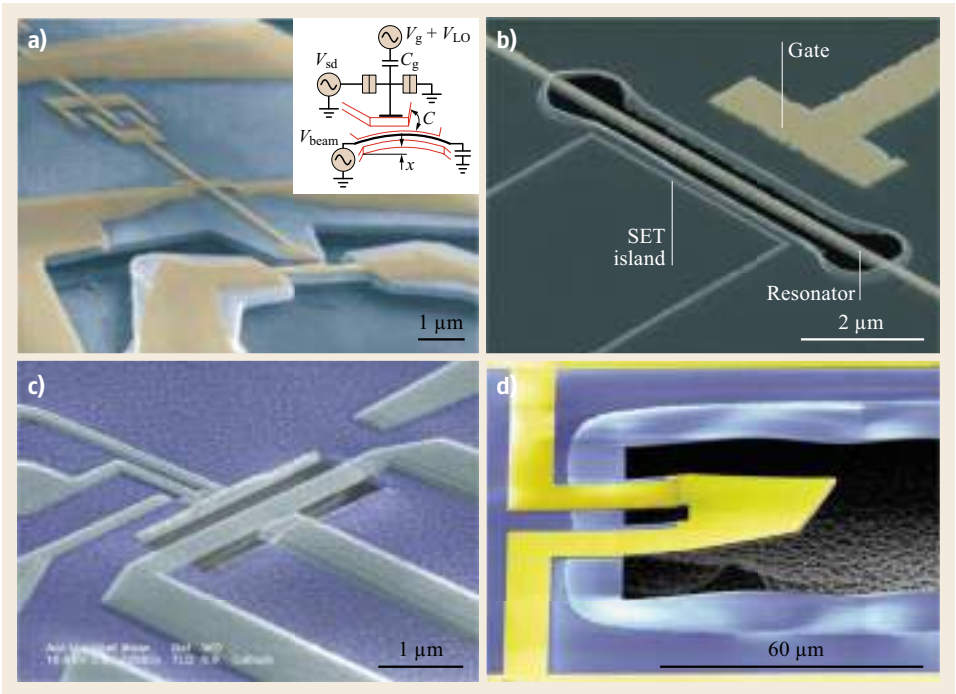


Fig. 13.59a–d Representative quantum NEMS devices that have enabled achieving key milestones in approaching the standard quantum mechanical limit (SQL) in human-made structures and devices. **(a)** SEM image of a doubly clamped GaAs beam with Al electrodes forming a single-electron transistor (SET) and beam electrode. *Inset*: a schematic of the mechanical and electrical operation of the device (after [13.80]). **(b)** A ≈ 20 MHz SiN nanomechanical resonator (200 nm wide, 8 μm long, coated with 20 nm Au atop 100 nm SiN), defined by etching through a suspended SiN membrane (after [13.81, 82]). **(c)** A SiN NEMS resonator with Al metallization, and coupled to the Al Cooper pair box (CPB) ≈ 300 nm away from the beam resonator (after [13.83]). **(d)** Colored SEM image of an AlN suspended film bulk acoustic resonator (130 nm Al on 330 nm AlN on 130 nm Al on 150 nm SiO_2) operating at the dilatational mode at ≈ 6.1 GHz (after [13.84])

scientific metrology, and for extending researchers' sensing and detection capabilities with unprecedented responsivities and sensitivities. In fundamental physics, NEMS make unique mesoscopic systems for probing quantum-limited phenomena [13.75, 85, 86] and for approaching and realizing bizarre quantum mechanical effects in manmade nano/microscale or even larger structures (contrary to just in fundamental particles) [13.86]. Toward such goals and to reach the standard quantum limit (SQL) in mesoscopic devices, significant progress has been made in the rapidly emerging fields of cavity quantum electromechanics (QEM) and cavity optomechanics. Remarkably, a few groups have recently reported cooling of NEMS/MEMS devices in their ground states, either by coupling superconducting quantum-bit (*qubit*) to UHF/microwave electromechanical (e.g., piezoelectric) MEMS resonators [13.84], by coupling mechanical resonators to microwave microstripline resonators [13.87, 88], or via optical cooling of strongly coupled optomechanical resonators [13.89, 90]. Figure 13.59 displays a few selected quantum NEMS devices that have enabled important milestones in this pursuit. These fundamental advances start to make it possible and promising for NEMS-enabled quantum information processing and technology.

In engineering and from a technological viewpoint, NEMS are highly attractive for resonant sensing and ultralow-power information processing in the classical domain. In particular, given their ultrasmall sizes, NEMS make exquisite sensors and detectors for various physical quantities. Numerous NEMS resonant sensors have been reported and many important and inspiring achievements have been made in the last decade. These include, but are not limited to, the following notable milestones:

1. Charge detection: ultrasensitive electronic charge detection with 0.1 electron/rtHz or better sensitivity [13.74].

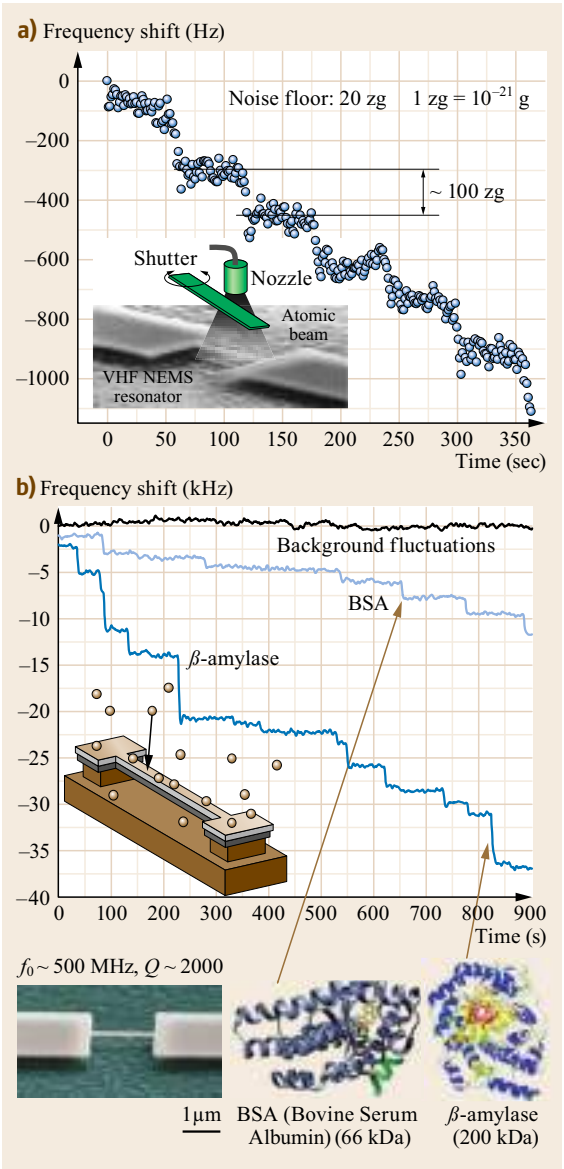


Fig. 13.60a,b Recent important milestones toward single-molecule and single-atom mass sensing and NEMS mass spectrometry. **(a)** Representative data demonstrating real-time sensing of Xe atoms landing on the surface of a very high-frequency (VHF) NEMS resonator, with a mass sensitivity of $\approx 7\,\mathrm{zg}$ ($1\,\mathrm{zg} = 10^{-21}\,\mathrm{g}$). Reprinted (adapted) with permission from [13.93]. Copyright 2006 American Chemical Society. **(b)** Representative data showing real-time detection of single protein molecules landing onto a UHF NEMS resonator, and the prototype of a NEMS mass spectrometry (after [13.94]) ►

2. Force sensing: attonewton-level ($1\,\mathrm{aN} = 10^{-18}\,\mathrm{N}$) force sensitivities that have led to magnetic resonance force detection of a single nuclear spin [13.91, 92].
3. Mass sensing: resonant mass sensing with imprecision at the level of a few zeptograms ($1\,\mathrm{zg} = 10^{-21}\,\mathrm{g}$, which is smaller than the mass of most proteins and biomolecules) and then quickly down to the levels of clusters of atoms [13.93–97] (all in vacuum), and attogram-level ($10^{-18}\,\mathrm{g}$) detection in fluids [13.98]. Figure 13.60 demonstrates the hallmark representative data of the zeptogram-scale resonant NEMS mass sensing, and the very first real-time detection of individual protein molecules landing on vibrating NEMS at UHF, as well as the first demonstration of the proof-of-concept NEMS mass spectrometry. Furthermore, these advances in sensitivities and devices technologies enabled new types of mass spectrometry with single-biomolecule resolution [13.94], a wide spectrum of NEMS-based biochemical sensing in fluids [13.99] and also a new avenue for probing the dynamics and fluctuations of physisorbed atoms/molecules on solid surfaces, with unprecedented sensitivities (enough to detect the effects of adsorption–desorption and diffusion at sub-monolayer coverage) [13.100].
4. Displacement and strain sensing: displacement detection in the quantum regime (as discussed above [13.80, 81]) has been rigorously pursued and it is now possible to achieve imprecision below the SQL. In the classical regime, femtometer ($1\,\mathrm{fm} = 10^{-15}\,\mathrm{m}$) displacement sensitivities have been achieved, as well as sensing of very low-level, fast-varying strains [13.101, 102].

NEMS Switches

Since early 2000s the idea of mechanical computing has been revisited with considerable efforts, enabled by scaling from MEMS to NEMS. In this context, contact-mode MEMS/NEMS logic switches (relays) have been rapidly emerging, and have been actively explored as a promising candidate for ultralow-power applications. Truly miniaturized and scaled, NEMS switches can offer compelling attributes and comparative advantages including ideally abrupt switching with minimal off-state leakage (zero-leakage when switched off), suitable for high temperature, extreme environments and small footprints. In this pursuit, however, major challenges remain: (i) all recently reported high-performance mechanical switches are still very much in the MEMS domain and are orders of magnitude larger in size/volume than today's mainstream transistors. In fact, recent MEMS switches with excellent cycle numbers have been all $\geq 10^3$ to $\geq 10^4$ times larger in *motional volume* than

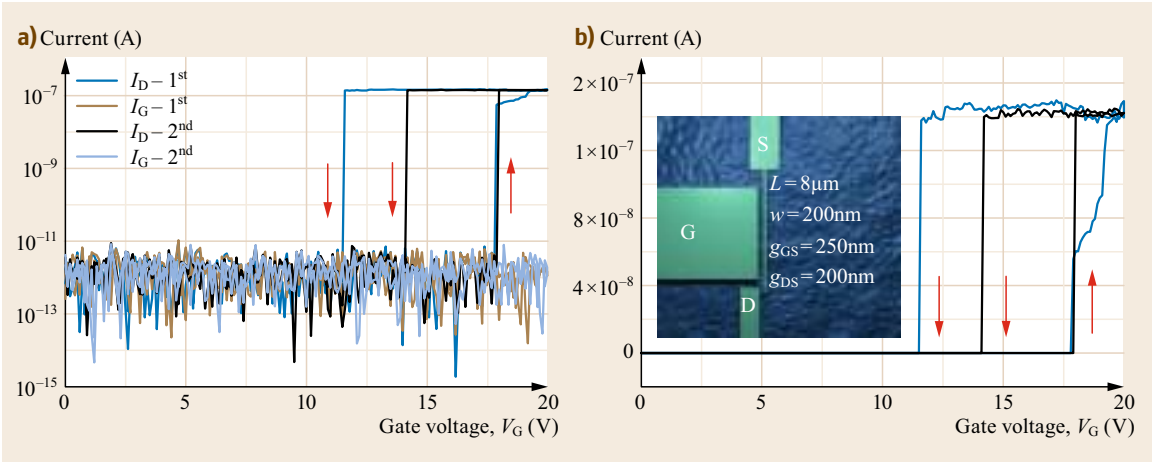


Fig. 13.61a,b First two switching cycles with hysteresis, measured in a SiC NEMS switch, with currents in logarithmic scale **(a)** and linear scale **(b)**, respectively (after [13.103])

genuine NEMS devices with nanometer-scale features. (ii) Most of today's truly nanoscale mechanical switches, however, suffer from a very short lifetime, and some can merely switch one or a few cycles before failure.

Among a number of cutting-edge efforts, the team at Case Western Reserve University [13.103–105] recently made efforts towards addressing one fundamental challenge in today's high-temperature, harsh-environment electronics – power consumption and off-state leakage in switching devices. The team has made SiC NEMS and exploit the unique properties of these SiC NEMS to develop logic building blocks from switches to inverters and subcircuits, which could offer new functionalities and advantages over traditional logic devices used in electronics at high temperatures and other extreme conditions. Here we briefly highlight the recently achieved milestones in the field:

1. SiC NEMS inverters that can operate at both room temperature and high temperature ($T \approx 500\,^{\circ}C$) [13.105]
2. NEMS switches based on ≈ 50 nm-thin nanowires with very low switch-on voltages ($V_{on} \approx 1$ V) [13.106]
3. Lateral three-terminal NEMS switches with robust performance and $> 10^6$ cycles of time evolution of the switching recorded in real time in ambient air [13.103]
4. Demonstration of dual-gate SiC NEMS switches and exploration of multiple-gate controls [13.107]
5. Design and implementation of various functional and reliable NEMS switches and logic gates [13.103, 105–108]
6. Nanoscale mechanical switches 10^3–10^4 times smaller in volume than state-of-the-art MEMS switches [13.103–108].

Figure 13.61 shows typical switching I–V curves measured from a generic three-terminal SiC NEMS switch, with $V_{on} = 17.9$ V and the on/off state current ratio $I_{on}/I_{off} > 10^5$. Beyond such generic lateral switch structures, we have also demonstrated four-terminal devices with a novel dual-gate configuration for enhanced gate control over electrostatic actuation. SiC NEMS switches and inverters have demonstrated very high cycle numbers in the millions and billions ($\approx 10^6$ to $\approx 10^9$) [13.105]. Given the ultrasmall sizes and volumes of these devices, such high performance is highly impressive, attractive and unique amongst NEMS switches [13.103]. Today, almost all contact-mode NEMS switches made of other important materials (Si, SiN, carbon nanotubes, etc.) still suffer from very low operating-cycle numbers. We have made considerable efforts on investigating nanoscale contacts in SiC NEMS, which are key to the devices' reliability and lifetime. We have recorded the complete time evolution of the switching events of robust SiC NEMS switches. Even in ambient air, for many devices, we have recorded $> 10^5$–10^6 switching cycles, without failure, over days of operations [13.103, 108].

13.3 Challenges and Perspectives

Although the field of MEMS has experienced significant growth over the past decade, many challenges still remain. These challenges can be grouped into three general categories:

1. Fabrication challenges
2. Packaging challenges
3. Application challenges.

Challenges in these areas will, in large measure, determine the commercial success of a particular MEMS device both in technical and economic terms. The following presents a brief discussion of some of these challenges as well as possible approaches to address them.

In terms of fabrication, MEMS is currently dominated by planar processing techniques, which find their roots in silicon IC fabrication. The planar approach and the strong dependence on silicon worked well in the early years, since many of the processing tools and methodologies commonplace in IC fabrication could be directly utilized in the fabrication of MEMS devices. This approach lends itself to the integration of MEMS with silicon ICs. Therefore, it is still popular for various applications. However, modular process integration of micromachining with standard IC fabrication is not straightforward and represents a great challenge in terms of processing material compatibility, thermal budget requirements, etc. Furthermore, planar processing places significant geometric restrictions on device designs, especially for complex mechanical components requiring high aspect ratio three-dimensional geometries, which are certain to increase as the application areas for MEMS continue to grow. Along the same lines, new applications will likely demand materials other than silicon, which may not be compatible with the conventional microfabrication approach, posing a significant challenge if integration with silicon microelectronics is required. Microassembly technique can become an attractive solution to alleviate these issues. Multifunctional microsystems can be implemented by assembling various MEMS devices and electronic building blocks fabricated through disparate processing technologies. Microsystems on a common substrate will likely become the ultimate solution. Development of sophisticated modeling programs for device design and performance will become increasingly important as fabrication processes and device designs become more complex. In terms of NEMS, the most significant challenge is likely the integration of nano- and microfabrication techniques into a unified process, since NEMS devices are likely to consist of both nanoscale and microscale structures. Integration will be particularly challenging for nanoscale devices fabricated using a *bottom-up* approach, since there is no analog in microfabrication. Nevertheless, hybrid systems consisting of nanoscale and microscale components will become increasingly common as the field continues to expand.

Fabrication issues notwithstanding, packaging is and will continue to be a significant challenge to the implementation of MEMS. MEMS is unlike IC packaging, which benefits from a high degree of standardization. MEMS devices inherently require interaction with the environment and since each application has in some way a unique environment, standardization of packaging becomes extremely difficult. This lack of standardization tends to drive up the costs associated with packaging, making MEMS less competitive with alternative approaches. In addition, packaging tends to negate the effects of miniaturization based upon microfabrication, especially for MEMS devices, requiring protection from certain environmental conditions. Moreover, packaging can cause performance degradation of MEMS devices, especially in situations where the environment exerts mechanical stresses on the package, which in turn results in a long-term device performance drift. To address many of these issues, wafer-level packaging schemes that are customized to the device of interest will likely become more common. In essence, the packaging of MEMS will move away from the conventional IC methods that utilize independently manufactured packages toward custom packages, which are created specifically for the device as a part of the batch fabrication process.

Without question, the increasing advancement of MEMS will open many new potential application areas to the technology. In most cases, MEMS will be one of several alternatives available for implementations. For cost-sensitive applications, the tradeoff between technical capabilities and cost will challenge those who desire to commercialize the technology. One of the biggest challenges to the field will be to identify application areas that are well suited for MEMS/NEMS technology and have no serious competitions. As MEMS technology moves away from component-level and more towards microsystems solutions, it is likely that such application areas will come to the fore.

References

13.1 M. Mehregany, S.F. Bart, L.S. Tavrow, J.H. Lang, S.D. Senturia: Principles in design and microfabrication of variable-capacitance side-drive motors, J. Vac. Sci. Tech. A **8**, 3614–3624 (1990)

13.2 Y.-C. Tai, R.S. Muller: IC-processed electrostatic synchronous micromotors, Sens. Actuators **20**, 49–55 (1989)

13.3 J.J. Sniegowski, S.L. Miller, G.F. LaVigne, M.S. Roders, P.J. McWhorter: Monolithic geared-mechanisms driven by a polysilicon surface-micromachined on-chip electrostatic microengine. In: *IEEE Solid-State Sens. Actuators Workshop* (1996) pp. 178–182

13.4 G.T.A. Kovacs: *Micromachined Transducer Sourcebook* (McGraw-Hill, Boston 1998)

13.5 S.D. Senturia: *Microsystem Design* (Kluwer Academic Publishers, New York 2001)

13.6 J.E. Gragg, W.E. McCulley, W.B. Newton, C.E. Derrington: Compensation and calibration of a monolithic four terminal silicon pressure transducer. In: *IEEE Solid-State Sens. Actuators Workshop* (1984) pp. 21–27

13.7 Y. Wang, M. Esashi: A novel electrostatic servo capacitive vacuum sensor. In: *IEEE Int. Conf. Solid-State Sens. Actuators* (1997) pp. 1457–1460

13.8 W.H. Ko, Q. Wang: Touch mode capacitive pressure sensors, Sens. Actuators **75**, 242–251 (1999)

13.9 H. Kapels, T. Scheiter, C. Hierold, R. Aigner, J. Binder: Cavity pressure determination and leakage testing for sealed surface micromachined membranes: a novel on-wafer test method. In: *Proc. 11th Annu. Int. Workshop Micro Electro Mech. Syst. (MEMS98), January 25–29, Heidelberg* (1998) pp. 550–555 doi:10.1109/MEMSYS.1998.659817

13.10 J.M. Bustillo, R.T. Howe, R.S. Muller: Surface micromachining for microelectromechanical systems, Proc. IEEE **86**(8), 1552–1574 (1998)

13.11 J.H. Smith, S. Montague, J.J. Sniegowski, P.J. McWhorter: Embedded micromechanical devices for the monolithic integration of MEMS with CMOS, IEEE Int. Electron Dev. Meet. (1993) doi:10.2172/114489

13.12 T.A. Core, W.K. Tsang, S.J. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. **36**(10), 39–47 (1993)

13.13 H. Kapels, R. Aigner, C. Kolle: Monolithic surface-micromachined sensor system for high pressure applications. In: *Transducers '01 Eurosensors XV*, ed. by E. Obermeier (Springer, Berlin, Heidelberg 2001) pp. 56–59

13.14 C. Lu, M. Lemkin, B.E. Boser: A monolithic surface micromachined accelerometer with digital output, IEEE J. Solid-State Circuits **30**(12), 160–161 (1995)

13.15 N. Yazdi, K. Najafi: An all-silicon single-wafer fabrication technology for precision microaccelerometers. In: *IEEE Int. Conf. Solid-State Sens. Actuators* (1997) pp. 1181–1184

13.16 M. Lemkin, M.A. Ortiz, N. Wongkomet, B.E. Boser, J.H. Smith: A 3-axis surface micromachined $\sigma\delta$ accelerometer. In: *IEEE Int. Solid-State Circuits Conf., 1997. Dig. Tech. Pap. 43rd ISSCC* (1997) pp. 202–203

13.17 T.B. Gabrielson: Mechanical-thermal noise in micromachined acoustic and vibration sensors, IEEE Trans. Electron Devices **40**(5), 903–909 (1993)

13.18 W.A. Clark, R.T. Howe: Surface micromachined z-axis vibratory rate gyroscope. In: *IEEE Solid-State Sens. Actuators Workshop* (1996) pp. 283–287

13.19 T. Juneau, A.P. Pisano: Micromachined dual input axis angular rate sensor. In: *Solid-State Sens. Actuators Workshop* (IEEE, Cleveland Heights 1996) pp. 299–302

13.20 R.S. Muller, K.Y. Lau: Surface-micromachined microoptical elements and systems, Proc. IEEE **86**(8), 1705–1720 (1998)

13.21 L.J. Hornbeck: Current status of the digital micromirror device (DMD) for projection television applications. In: *IEEE Int. Electron Devices Meet* (1993) pp. 381–384

13.22 P.F. Van Kessel, L.J. Hornbeck, R.E. Meier, M.R. Douglass: A MEMS-based projection display, Proc. IEEE **86**(8), 1687–1704 (1998)

13.23 M.J. Daneman, N.C. Tien, O. Solgaard, K.Y. Lau, R.S. Muller: Linear vibromotor-actuated micromachined microreflector for integrated optical systems. In: *IEEE Solid-State Sens. Actuators Workshop* (1996) pp. 109–112

13.24 M.S. Cohen, M.F. Cina, E. Bassous, M.M. Opyrsko, J.L. Speidell, F.J. Canora, M.J. DeFranza: Packaging of high density fiber/laser modules using passive alignment techniques, IEEE Trans. Comp. Hybrids Manuf. Technol. **15**, 944–954 (1992)

13.25 M.J. Wale, C. Edge: Self-aligned flip-chip assembly of photonic devices with electrical and optical connections, IEEE Trans. Comp. Hybrids Manuf. Technol **13**, 780–786 (1990)

13.26 K.S.J. Pister, M.W. Judy, S.R. Burgett, R.S. Fearing: Microfabricated hinges, Sens. Actuators (A) **33**(3), 249–256 (1992)

13.27 O. Solgaard, M. Daneman, N.C. Tien, A. Friedberger, R.S. Muller, K.Y. Lau: Optoelectronic packaging using silicon surface-micromachined alignment mirrors, IEEE Photon. Technol. Lett. **7**(1), 41–43 (1995)

13.28 S.S. Lee, L.S. Huang, C.J. Kim, M.C. Wu: 2x2 MEMS fiber optic switches with silicon sub-mount for low-cost packaging. In: *IEEE Solid-State Sens. Actuators Workshop* (1998) pp. 281–284

13.29 T. Akiyama, H. Fujita: A quantitative analysis of scratch drive actuator using buckling motion. In: *Proc. Micro Electro Mech. Syst., MEMS* (1995) pp. 310–315

13.30 V.A. Aksyuk, F. Pardo, D.J. Bishop: Stress-induced curvature engineering in surface-micromachined devices, Proc. SPIE **3680**, 984 (1999)

13.31 D.J. Young, B.E. Boser: A micromachined variable capacitor for monolithic low-noise VCOs. In:

IEEE Solid-State Sens. Actuator Workshop (1996) pp. 86–89

13.32 A. Dec, K. Suyama: Micromachined electro-mechanically tunable capacitors and their applications to RF IC's, IEEE Trans. Microw. Theory Tech. **46**, 2587–2596 (1998)

13.33 Z. Li, N.C. Tien: A high tuning-ratio silicon-micromachined variable capacitor with low driving voltage. In: *IEEE Solid-State Sens. Actuator Microsyst. Workshop* (2002) pp. 239–242

13.34 Z. Xiao, W. Peng, R.F. Wolffenbuttel, K.R. Farmer: Micromachined variable capacitor with wide tuning range. In: *IEEE Solid-State Sens. Actuator Workshop* (2002) pp. 346–349

13.35 J.J. Yao, S.T. Park, J. DeNatale: High tuning-ratio MEMS-based tunable capacitors for RF communications applications. In: *IEEE Solid-State Sens. Actuator Workshop* (1998) pp. 124–127

13.36 J.B. Yoon, C.T.-C. Nguyen: A high-Q tunable micromechanical capacitor with movable dielectric for RF applications. In: *IEEE Int. Electron Devices Meet* (2000) pp. 489–492

13.37 D.J. Young, V. Malba, J.J. Ou, A.F. Bernhardt, B.E. Boser: Monolithic high-performance three-dimensional coil inductors for wireless communication applications. In: *IEEE Int. Electron Devices Meet* (1997) pp. 67–70

13.38 D.J. Young, B.E. Boser, V. Malba, A.F. Bernhardt: A micromachined RF low phase noise voltage-controlled oscillator for wireless communication, Int. J. RF Microw. Comput.-Aided Eng. **11**(5), 285–300 (2001)

13.39 J.B. Yoon, C.H. Han, E. Yoon, K. Lee, C.K. Kim: Monolithic high-Q overhang inductors fabricated on silicon and glass substrates. In: *IEEE Int. Electron Devices Meet* (1999) pp. 753–756

13.40 C.L. Chua, D.K. Fork, K.V. Schuylenbergh, J.P. Lu: Self-assembled out-of-plane high Q inductors. In: *IEEE Solid-State Sens. Actuator Microsyst. Workshop* (2002) pp. 372–373

13.41 C.L. Goldsmith, Z. Yao, S. Eshelman, D. Denniston: Performance of low-loss RF MEMS capacitive switches, IEEE Microw. Guided Wave Lett. **8**(8), 269–271 (1998)

13.42 J.J. Yao, M.F. Chang: A surface micromachined miniature switch for telecommunication applications with signal frequencies from DC up to 40 GHz. In: *8th Int. Conf. Solid-State Sens. Actuators* (1995) pp. 384–387

13.43 P.M. Zavracky, N.E. McGruer, R.H. Morriosn, D. Potter: Microswitches and microrelays with a view toward microwave applications, Int. J. RF Microw. Comput.-Aided Eng. **9**(4), 338–347 (1999)

13.44 D. Hyman, J. Lam, B. Warneke, A. Schmitz, T.Y. Hsu, J. Brown, J. Schaffner, A. Walston, R.Y. Loo, M. Mehregany, J. Lee: Surface-micromachined RF MEMs switches on GaAs substrates, Int. J. RF Microw. Comput.-Aided Eng. **9**(4), 348–361 (1999)

13.45 C.T.C. Nguyen, R.T. Howe: CMOS microelectromechanical resonator oscillator. In: *IEEE Int. Electron Devices Meet* (1993) pp. 199–202

13.46 L. Lin, R.T. Howe, A.P. Pisano: Microelectromechanical filters for signal processing, IEEE J. Microelectromech. Syst. **7**(3), 286–294 (1998)

13.47 F.D. Bannon III, J.R. Clark, C.T.C. Nguyen: High frequency micromechanical filter, IEEE J. Solid-State Circuits **35**(4), 512–526 (2000)

13.48 K. Wang, Y. Yu, A.C. Wong, C.T.C. Nguyen: VHF free-free beam high-Q micromechanical resonators. In: *12th IEEE Int. Conf. Micro Electro Mech. Syst.* (1999) pp. 453–458

13.49 J.R. Clark, W.T. Hsu, C.T.C. Nguyen: High-Q VHF micromechanical contour-mode disk resonators. In: *IEEE Int. Electron Devices Meet* (2000) pp. 493–496

13.50 C.T.C. Nguyen, R.T. Howe: Quality factor control for micromechanical resonator. In: *IEEE Int. Electron Devices Meet* (1992) pp. 505–508

13.51 M.L. Roukes: Nanoelectromechanical systems face the future, Phys. World **14**, 25–31 (2001)

13.52 M.L. Roukes, A. Scherer, S.J. Allen Jr., H.G. Craighead, R.M. Ruthen, E.D. Beebe, J.P. Harbison: Quenching of the Hall effect in a onedimensional wire, Phys. Rev. Lett. (1987) doi:10.1103/PhysRevLett.59.3011

13.53 B.P. Van der Gaag, A. Scherer: Microfabrication below 10nm, Appl. Phys. Lett. (1990) doi:10.1063/1.102772

13.54 M.L. Roukes: Nanoelectromechanical systems. In: *Tech. Dig. 2000 Solid-State Sens. Actuator Workshop* (Transducers Research Foundation, Cleveland 2000) pp. 367–376

13.55 M.L. Roukes: *Plenty of room, indeed*, *Sci*, Am, Vol. 285, 2001) pp. 48–57

13.56 K.L. Ekinci, Y.T. Yang, M.L. Roukes: Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems, J. of Appl. Phys. **95**(5), 2682–2689 (2004)

13.57 K.L. Ekinci, M.L. Roukes: Nanoelectromechanical systems, Rev. of Sci. Instrum. (2005) doi:10.1063/1.1927327

13.58 H.G. Craighead: Nanoelectromechanical systems, Science **290**, 1532–1535 (2000)

13.59 A.N. Cleland, M.L. Roukes: Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals, Appl. Phys. Lett. **69**, 2653–2655 (1996)

13.60 D.W. Carr, H.G. Craighead: Fabrication of nanoelectromechanical systems in single crystal silicon using silicon on insulator substrates and electron beam lithography, J. Vac. Sci. Technol. B **15**, 2760–2763 (1997)

13.61 T.S. Tighe, J.M. Worlock, M.L. Roukes: Direct thermal conductance measurements on suspended monocrystalline nanostructures, Appl. Phys. Lett. **70**, 2687–2689 (1997)

13.62 S.C. Masmanidis, R.B. Karabalin, I.D. Vlaminck, G. Borghs, M.R. Freeman, M.L. Roukes: Multifunctional nanomechanical systems via tunably coupled piezoelectric actuation, Science (2007) doi:10.1126/science.1144793

13.63 H.X. Tang, X.M.H. Huang, M.L. Roukes, M. Bichler, W. Wegsheider: Two-dimensional electron-gas actuation and transduction for GaAs nano-

electromechanical systems, Appl. Phys. Lett. **81**, 3879–3881 (2002)
13.64 R. Ruby, P. Bradley, J.D. Larson, Y. Oshmyansky: PCS 1900 MHz duplexer using thin film bulk acoustic resonators (FBARs), Electron. Lett. (1999) doi:10.1049/el:19990559
13.65 G. Piazza, P.J. Stephanou, A.P. Pisano: Piezoelectric aluminum nitride vibrating contour-mode MEMS resonators, J. of Microelectromechanical Syst. (2006) doi:10.1109/JMEMS.2006.886012
13.66 R.B. Karabalin, M.H. Matheny, X.L. Feng, E. Defaÿ, G. Le Rhun, C. Marcoux, S. Hentz, P. Andreucci, M.L. Roukes: Piezoelectric nanoelectromechanical resonators based on aluminum nitride thin films, Appl. Phys. Lett. (2009) doi:10.1063/1.3216586
13.67 N. Sinha, G.E. Wabiszewski, R. Mahameed, V.V. Felmetsger, S.M. Tanner, R.W. Carpick, G. Piazza: Piezoelectric aluminum nitride nanoelectromechanical actuators, Appl. Phys. Lett. (2009) doi:10.1063/1.3194148
13.68 X.L. Feng, R.R. He, P.D. Yang, M.L. Roukes: Very high frequency silicon nanowire electromechanical resonators, Nano Lett. (2007) doi:10.1021/nl0706695
13.69 P.X.-L. Feng: Nanoscale electromechanical devices enabled by nanowire structures. In: *Microelectronics to Nanoelectronics: Materials, Devices & Manufacturability*, ed. by A.B. Kaul (CRC, Boca Raton 2012) pp. 109–128
13.70 D.W. Carr, S. Evoy, L. Sekaric, H.G. Craighead, J.M. Parpia: Measurement of mechanical resonance and losses in nanometer scale silicon wires, Appl. Phys. Lett. **75**, 920–922 (1999)
13.71 D.W. Carr, L. Sekaric, H.G. Craighead: Measurement of nanomechanical resonant structures in single-crystal silicon, J. Vac. Sci. Technol. B **16**, 3821–3824 (1998)
13.72 S. Evoy, D.W. Carr, L. Sekaric, A. Olkhovets, J.M. Parpia, H.G. Craighead: Nanofabrication and electrostatic operation of single-crystal silicon paddle oscillators, J. Appl. Phys. **86**, 6072–6077 (1999)
13.73 L. Sekaric, M. Zalalutdinov, S.W. Turner, A.T. Zehnder, J.M. Parpia, H.G. Craighead: Nanomechancial resonant structures as tunable passive modulators, Appl. Phys. Lett. **80**, 3617–3619 (2002)
13.74 A.N. Cleland, M.L. Roukes: A nanometre-scale mechanical electrometer, Nature **392**, 160–162 (1998)
13.75 K.C. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes: Measurement of the quantum of thermal conductance, Nature **404**, 974–977 (2000)
13.76 S. Evoy, A. Olkhovets, L. Sekaric, J.M. Parpia, H.G. Craighead, D.W. Carr: Temperature-dependent internal friction in silicon nanoelectromechanical systems, Appl. Phys. Lett. **77**, 2397–2399 (2000)
13.77 X.M.H. Huang, X.L. Feng, C.A. Zorman, M. Mehregany, M.L. Roukes: VHF, UHF and microwave frequency nanomechanical resonators, New J. of Phys. (2005) doi:10.1088/1367-2630/7/1/247
13.78 X.M.H. Huang, C.A. Zorman, M. Mehregany, M.L. Roukes: Nanoelectromechanical systems: Nanodevice motion at microwave frequencies, Nature **421**, 496 (2003)
13.79 X.L. Feng, C.J. White, A. Hajimiri, M.L. Roukes: A self-sustaining ultrahigh-frequency nanoelectromechanical oscillator, Nat. Nanotechnol. **3**, 342–346 (2008)
13.80 R.G. Knobel, A.N. Cleland: Nanometre-scale displacement sensing using a single electron transistor, Nature **424**, 291–293 (2003)
13.81 M.D. LaHaye: *The Radio-Frequency Single-Electron Transistor Displacement Detector*, Ph.D. Thesis (Department of Physics, University of Maryland, College Park 2005)
13.82 M.D. LaHaye, O. Buu, B. Camarota, K.C. Schwab: Approaching the quantum limit of a nanomechanical resonator, Science **304**, 74–77 (2004)
13.83 M.D. LaHaye, J. Suh, P.M. Echternach, K.C. Schwab, M.L. Roukes: Nanomechanical measurements of a superconducting qubit, Nature **459**, 960–964 (2009)
13.84 A.D. O'Connell, M. Hofheinz, M. Ansmann, R.C. Bialczak, M. Lenander, E. Lucero, M. Neeley, D. Sank, H. Wang, M. Weides, J. Wenner, J.M. Martinis, A.N. Cleland: Quantum ground state and single-phonon control of a mechanical resonator, Nature **464**, 697–703 (2010)
13.85 A. Cho: Researchers race to put the quantum into mechanics, Science (2003) doi:10.1126/science.299.5603.36
13.86 K.C. Schwab, M.L. Roukes: Putting mechanics into quantum mechanics, Phys. Today **58**, 36–42 (2005)
13.87 C.A. Regal, J.D. Teufel, K.W. Lehnert: Measuring nanomechanical motion with a microwave cavity interferometer, Nat. Phys. **4**, 555–560 (2008)
13.88 J.D. Teufel, T. Donner, D. Li, J.W. Harlow, M.S. Allman, K. Cicak, A.J. Sirois, J.D. Whittaker, K.W. Lehnert, R.W. Simmonds: Sideband cooling of micromechanical motion to the quantum ground state, Nature **475**, 359–363 (2011)
13.89 T.J. Kippenberg, K.J. Vahala: Cavity optomechanics, Opt. Express **15**, 17172–17205 (2007)
13.90 T.J. Kippenberg, K.J. Vahala: Cavity optomechanics: Back-action at the mesoscale, Science **321**, 1172–1176 (2008)
13.91 T.D. Stowe, K. Yasumura, T.W. Kenny, D. Botkin, K. Wago, D. Rugar: Attonewton force detection using ultrathin silicon cantilevers, Appl. Phys. Lett. **71**, 288–290 (1997)
13.92 D. Rugar, R. Budakian, H.J. Mamin, B.W. Chui: Single spin detection by magnetic resonance force microscopy, Nature **430**, 329–332 (2004)
13.93 Y.T. Yang, C. Callegari, X.L. Feng, K.L. Ekinci, M.L. Roukes: Zeptogram-scale nanomechanical mass sensing, Nano Lett. **6**, 583–586 (2006)
13.94 A.K. Naik, M.S. Hanay, W.K. Hiebert, X.L. Feng, M.L. Roukes: Toward single-molecule nanomechanical mass spectrometry, Nat. Nanotechnol. **4**, 445–450 (2009)

13.95 B. Lassagne, D. Garcia-Sanchez, A. Aguasca, A. Bachtold: Ultrasensitive mass sensing with a nanotube electromechanical resonator, Nano Lett. **8**, 3735–3738 (2008)

13.96 H.Y. Chiu, P. Hung, H.W. Ch Postma, M. Bockrath: Atomic-scale mass sensing using carbon nanotube resonators, Nano Lett. **8**, 4342–4346 (2008)

13.97 K. Jensen, K. Kim, A. Zettl: An atomic-resolution nanomechanical mass sensor, Nat. Nanotechnol. **3**, 533–537 (2008)

13.98 T.P. Burg, M. Godin, S.M. Knudsen, W.J. Shen, G. Carlson, J.S. Foster, K. Babcock, S.R. Manalis: Weighing of biomolecules, single cells and single nanoparticles in fluid, Nature **446**, 1066–1069 (2007)

13.99 J.L. Arlett, E.B. Myers, M.L. Roukes: Comparative advantages of mechanical biosensors, Nat. Nanotechnol. **6**, 203–215 (2011)

13.100 Y.T. Yang, C. Callegari, X.L. Feng, M.L. Roukes: Surface adsorbate fluctuations and noise in nanoelectromechanical systems, Nano Lett **11**, 1753–1759 (2011)

13.101 R.R. He, X.L. Feng, M.L. Roukes, P.D. Yang: Self-transducing silicon nanowire electromechanical systems at room temperature, Nano Lett. **8**, 1756–1761 (2008)

13.102 Z. Wang, J. Lee, P.X.-L. Feng: Spatial mapping of multimode Brownian motions in high frequency silicon carbide microdisk resonators, Nat. Commun. (2014) doi:10.1038/ncomms6158

13.103 T. He, R. Yang, S. Rajgopa, M. Tupta, S. Bhunia, M. Mehregany, P.X.-L. Feng: Robust silicon carbide (SiC) nanoelectromechanical switches with long cycles in ambient and high temperature conditions. In: *Proc. 26th IEEE Int. Conf. Micro Electro Mech. Syst., Taipei* (2013) pp. 516–519

13.104 T. He, R. Yang, V. Ranganathan, S. Rajgopal, M.A. Tupta, S. Bhunia, M. Mehregany, P.X.-L. Feng: Silicon carbide (SiC) nanoelectromechanical switches and logic gates with long cycles and robust performance in ambient air and at high temperature, Electron Devices Meet. (2013) doi:10.1109/IEDM.2013.6724562

13.105 T. Lee, S. Bhunia, M. Mehregany: Electromechanical computing at 500°C with SiC, Science **329**, 1316–1318 (2010)

13.106 X.L. Feng, M.H. Matheny, C.A. Zorman, M. Mehregany, M.L. Roukes: Low voltage nanoelectromechanical switches based on silicon carbide nanowires, Nano Lett **10**, 2891–2896 (2010)

13.107 T. He, R. Yang, S. Rajgopal, S. Bhunia, M. Mehregany, P.X.-L. Feng: Dual-gate silicon carbide (SiC) lateral nanoelectromechanical switches. In: *Nano/Micro Engineered and Molecular Systems (NEMS), 2013 8th IEEE International Conference* (IEEE, Suzhou 2013) doi:10.1109/NEMS.2013.6559791

13.108 T. He, V. Ranganathan, R. Yang, S. Rajgopal, S. Bhunia, M. Mehregany, P.X.-L. Feng: Time-domain AC characterization of silicon carbide (SiC) nanoelectromechanical switches toward high-speed operations. In: *Tech. Digest, 17th Int. Conf. on Solid-State Sensors, Actuators & Microsystems (Transducers 13)* (IEEE, Barcelona 2013) doi:10.1109/Transducers.2013.6626855

14. Single-Walled Carbon Nanotube Sensor Concepts

Cosmin I. Roman, Thomas Helbling, Miroslav Haluška, Christofer Hierold

Carbon nanotubes are excellent nanocomponents that offer unique properties that can be exploited in next-generation devices. Sensing applications is perhaps the class that has most to gain from single-walled carbon nanotubes (SWNTs); virtually any property of SWNTs, such as electronic, electrical, mechanical and optical, can result or has already resulted in sensor concept demonstrators. The basic questions this chapter will attempt to address are, *why* use SWNTs? and *how* can SWNTs be used? in sensing applications. A tour through the gallery of basic nanotube properties is given in order to reveal the richness and uniqueness of this material's intrinsic properties. Together with examples from the literature, showing performance of SWNT-based sensors at least comparable to – but oftentimes surpassing – state-of-the-art micro- or macrodevices, nanotube properties should explain *why* there is so much effort currently being invested in this field. As nanotubes, like any other nano-object, are not easy to probe, a versatile strategy for accessing their properties, via the carbon nanotube field-effect transistor (CNFET) concept, will be described in this chapter. Furthermore, descriptions of CNFET device fabrication together with examples of SWNT sensor demonstrators utilizing the CNFET principle will be followed by proposals of *how* nanotubes can be utilized in sensors.

14.1 **Sensor Design Considerations** 432
14.1.1 CNT Properties for Sensing 432
14.1.2 Carbon Nanotube FET Structures 436
14.1.3 Sensor Characterization 438

14.2 **Sensor Fabrication: SWNT Synthesis and Integration** 439
14.2.1 Methods for SWNT Production 440
14.2.2 Strategies for SWNT Assembly into Devices 441
14.2.3 Process Contaminations and Sensor Performance 444

14.3 **Summary of State-of-the-Art, Applications Examples** 445
14.3.1 Chemical and Biochemical Sensors 445
14.3.2 Piezoresistive Sensors 447
14.3.3 Resonant Sensors 449

14.4 **Concluding Remarks** 451

References 451

Sensors are only one possible application of SWNTs. Other notable applications include field emission devices, energy storage, composites, and nanoelectronics [14.1]. For example, *carbon-based nanoelectronics* has been identified by the International Technology Roadmap 2013 (ITRS) – edited by a consortium of scientists from all major semiconductor manufacturers and academic institutions – as the top contender to Silicon CMOS (complementary metal–oxide–semiconductor) technology. The focus of this chapter is however on sensing devices, for which carbon nanotubes (CNTs) present several key advantages including:

1. *Nanometer feature size*: Useful for building highly localized sensing units (active spots) and for large-scale integration (sensor arrays)
2. *Ultrahigh sensitivity and resolution*: Stemming from their unique structural and electronic properties (see Sect. 14.1.1) and high surface-to-volume ratio
3. *Ultralow power consumption*: Whether operating as transistors or mechanical resonators, the power needed to excite or probe an SWNT is in the order of 10 nW.

Because of these and other advantages, SWNTs have, in the relatively short time since their discovery, resulted in sensor device demonstrators that have been greeted with optimism worldwide. Many of these investigations have been published in prestigious research journals such as Science and Nature. There are, however, still some challenges to overcome before broader

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_14

acceptance in industrial product development activities will be observed. In fact, carbon nanotubes are today at a crossroad between basic science and engineering. Carbon nanotube device demonstrators and theoretical extrapolations often surpass in performance state-of-the-art devices, more than motivating any future attempts to solve the remaining issues. On the basis of the steep evolutionary slope experienced so far, it may not take long until carbon nanotube-based sensors appear on device and product roadmaps.

The remainder of this chapter is structured as follows. In Sect. 14.1 design considerations for SWNT sensors are described, starting with a brief survey of useful SWNT properties for sensing. The CNFET is introduced in Sect. 14.1.2 as a platform enabling access to individual SWNT properties during the sensing process. The current status of CNFET-based sensor characterization is captured in Sect. 14.1.3. Methods for fabricating, or supporting the fabrication of SWNT field-effect transistors (FETs) are reviewed in Sect. 14.2. Finally, Sect. 14.3 will be devoted to examples of CNT-based sensors, encompassing three main case studies, namely (bio)chemical, piezoresistive and resonator sensors, and recent advancements in those fields.

14.1 Sensor Design Considerations

The range of sensing schemes involving carbon nanotubes is considerable. A rough classification of CNT-based sensors can be done according to: *material*, *input*, and *output*. The carbon nanotube *material* utilized in sensors can vary from individual SWNTs, MWNTs and bundles, to CNT networks and composites and even to bulk (forests) CNT material (Table 14.1). Sensors based on nanotubes can respond to a wide range of *inputs*, including (bio)chemical (molecules), mechanical (deformation), optical (radiation), and electrical (fields by charges). Also different transduction mechanisms have been employed in CNT sensors to generate *outputs* such as electrical (conductive, capacitive), mechanical (resonance frequency), and optical (luminescence). A nonexhaustive CNT sensor catalog with references is given in Table 14.1.

For reasons of clarity and concreteness, in this chapter the discussion will be restricted to a particular class of CNT sensors, namely carbon nanotube field-effect transistor (CNFET) sensors. This class refers to FET configurations with *an individual single-walled carbon nanotube* (SWNT) channel that transforms *input stimuli of different origin* into *electrical signals at the output*. The CNFET is one of the simplest means to probe properties of an individual SWNT, and, at the same time, perhaps the most versatile building block available for engineering sensing devices. Furthermore, most of the knowledge gained in studying the rich variety of SWNT FET sensors is transferable to more complex devices involving MWNTs, bundles, or networks.

In the next subsection, some of the most important SWNT properties and property modulation mechanisms useful for sensing are listed and exemplified. The operation and particularities of CNFETs are reviewed in Sect. 14.1.2, whereas in Sect. 14.1.3, the current status of sensor characterization is briefly discussed.

14.1.1 CNT Properties for Sensing

SWNTs have many interesting properties for nanodevices in general, and nanosensors in particular. The properties of SWNTs are discussed in many textbooks [14.21–23]. In the following, we condense the most relevant SWNT properties, and support with examples on how these have been employed in sensor devices.

SWNTs are Nanometer-Thin Semiconducting or Metallic Wires

Structurally, SWNTs are molecular cylinders with monoatomic-thick walls, resembling a honeycomb lattice of carbon atoms (graphene) rolled into a tube (Fig. 14.1). The structure of an SWNT is uniquely identified by chiral indices (n, m) that give the so-called chiral vector $\boldsymbol{C}_\mathrm{h} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$ determining the circumference of the nanotube upon rolling ($\boldsymbol{a}_{1,2}$ are the 2-D graphene lattice vectors, Fig. 14.1a). The (n, m) indices can be interchanged with (d_t, θ), where d_t is the nanotube diameter given by $d_\mathrm{t} \approx (a/\pi)\sqrt{n^2 + m^2 + nm}$ (nm) and θ is the chiral angle defined via $\tan\theta = \sqrt{3}m/(2n + m)$ ($a = 0.249$ nm is the graphene lattice constant). SWNTs can be classified with the help of θ, into zigzag ($\theta = 0$; $m = 0$), armchair ($\theta = \pi/6$; $n = m$), or chiral ($0 < \theta < \pi/6$; $n \neq m \neq 0$). In practice, SWNTs have diameters ranging from 0.4–3 nm, and lengths around a few μm, although tubes of almost cm in length have been produced [14.24].

From the *electronic* point of view, depending on chiral indices (n, m), SWNTs can be either semiconducting or metallic (hereafter labeled s-SWNT (semiconducting single-walled carbon nanotube) or m-SWNT (metallic single-walled carbon nanotube)). A simple model for the electronic structure of SWNTs (zone-folding

Table 14.1 Nonexhaustive catalog of CNT-based sensors with references

Sensor type	Sensor input	Sensor output	CNT material
Chemical sensors [14.2–4]	Gas molecules (i. e., NO_2, NH_3, O_2); organic vapor	Conductance change; CNFETs: Shift of gate threshold	Individual SWNT
Chemical sensors [14.5, 6]	Gas molecules, organic vapor	Conductance change; CNFETs: Shift of gate threshold	CNT networks; functionalization
Biochemical sensors [14.7]	Biomolecules (liquid phase)	Conductance change; CNFETs: Shift of gate threshold	Individual SWNT, CNT networks
Electromechanical sensors [14.8–12]	Pressure, displacement, strain	Conductance change	Individual SWNT
Pressure sensors [14.13]	Pressure	Conductance change	Macroscopic CNT block CNT composite
Resonant cantilevers [14.14, 15]	Molecules (mass loading)	Resonance frequency change, readout via field emission	MWNT, DWNT
Doubly clamped resonators [14.16–18]	Molecules (mass loading), strain	Resonance frequency change; readout via conductance change	Individual SWNT
Optical sensors [14.19, 20]	Photons, light	Photocurrent	Individual SWNT

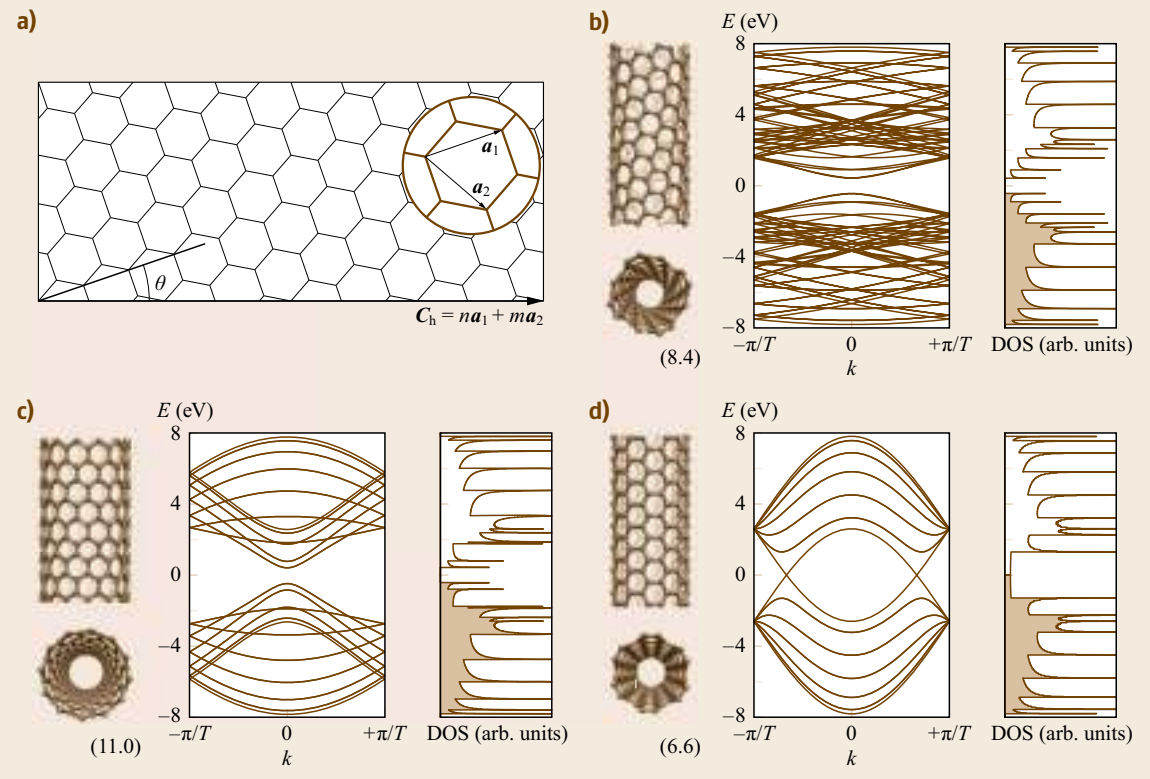


Fig. 14.1a–d Structural and electronic properties of SWNTs. **(a)** The *unfolded* unit cell of the chiral (8,4) CNT showing chiral vector $\boldsymbol{C}_h$, chiral angle θ, and 2-D graphene lattice vectors $\boldsymbol{a}_{1,2}$ (in *inset*). **(b)** The *folded* (8,4) structure, electronic band structure, and density of states (DOS). Bands are obtained via the zone-folding procedure applied to tight-binding dispersion relations (after [14.25]). **(c,d)** Band structure and DOS for zigzag (11,0) and armchair (6,6) nanotubes, respectively

of graphene tight-binding π bands) [14.25] predicts that those tubes for which $p \equiv (n-m) \bmod 3 = 0$ are metallic (1/3 of all SWNTs), the rest ($p = \pm 1$) being semiconducting (2/3 of all SWNTs). Structure, bands, and densities of states for three selected SWNTs (a chiral, a zigzag, and an armchair tube) are displayed in Fig. 14.1b–d. As for any 1-D structure, the density of states of SWNTs is singular at energies corresponding to sub-band extrema (van Hove singularities). A rough estimation for the electronic band gap of s-SWNTs is $E_g \approx (2at_0)/(\sqrt{3}d_t)$ (eV) [14.25] ($t_0 = 2.6$ eV is the so-called hopping tight-binding parameter for π orbitals). More accurate electronic structure calculations, taking into account the surface curvature of nanotubes, revealed that in fact only armchair (n,m) SWNTs are truly metallic; while other tubes with $p = 0$ actually do have a small band gap $E_g \approx 40/d_t^2$ [meV] [14.26] [these tubes are labeled SGS-SWNT (small-gap semiconducting single-walled carbon nanotube)].

The mentioned structural and electronic properties have resulted in a few SWNT sensor concepts. For example, FETs based on s-SWNTs have been utilized as charge detectors in flowmeters [14.27] or (bio)chemical sensors [14.2, 7] (more on this in Sect. 14.3.1). The high curvature ($d_t \approx 1$ nm) of carbon nanotubes (both tips and bodies) has been exploited for example in gas ionization sensors [14.28] and capacitance gas sensors [14.29].

SWNTs are Diamond-Stiff, Ultralight Strings

The basic *inertial* and *mechanical* properties of SWNTs are: linear mass density $\rho_L = 2.33 d_t$ (zg/nm) (1 zg = 10^{-21} g, and d_t is the tube diameter in nm as given above), Young's modulus E in the range of $\approx$ 1.25 TPa [14.30], max tensile strain 6%, and strength $\approx$ 45 GPa [14.31, 32]. These properties promote carbon nanotubes as ideal nanosized beams for mechanical sensors. For example, consider a straight, doubly clamped SWNT. Assuming that the SWNT can be treated as a continuum elastic beam, the resonance frequency of the fundamental flexural mode is given by [14.33]

$$\omega = \frac{4\pi^2}{L^2}\sqrt{\frac{EI}{3\rho A}\left(1 + \frac{L^2\sigma}{4\pi^2 EI}\right)},$$

where L is the tube length, I the moment of inertia, A the cross section area, ρ the mass density ($\rho_L = \rho A$), and σ the initial tension in the beam.

The dependence of ω on tension σ, has been utilized by *Sazonova* et al. [14.16] and others [14.17, 18] to demonstrate *tunable* CNT resonators. These devices can thus be employed as sensitive strain/stress sensors. Conversely, the attachment of a small mass on the tube increases the effective mass of the beam and as a result downshifts ω (via the ρ term in the above formula). This principle, known as resonant inertial balance, has been utilized in [14.14, 15, 17] to detect minute mass loading on a CNT. Sensing mechanisms for strain sensors and inertial balances are sketched in Fig. 14.2, and are discussed in more detail in Sect. 14.3.3.

SWNTs' Electronic Properties Can Be Tuned Mechanically

In analogy to crystalline semiconductors such as silicon, straining carbon nanotubes modifies their electronic band gap. With an extended version of the model that was used to obtain the bands in Fig. 14.1, *Yang* et al. [14.34] calculated the modulation of band gaps E_g of SWNTs of different chiralities under tensile (axial) and shear (torsional) strain (Fig. 14.3c,d). A linear approximation of the band gap change ΔE_g (in eV) with axial ε and torsional γ strain, was derived by the same group in [14.35]

$$\Delta E_g(\varepsilon,\gamma) \approx 3t_0 \mathrm{sgn}(2p+1) \times [(1+v)\varepsilon\cos 3\theta + \gamma \sin 3\theta]$$

with hopping parameter t_0, family p and chirality θ introduced earlier, and $v = 0.2$ the Poisson's ratio. The sign of ΔE_g depends on p, being positive for $p \in \{0, 1\}$ (band gap increases with strain), and negative for $p = -1$ (band gap decreases with strain). Also, the sensitivity with respect to strain, depends on the chiral angle; as such, armchair tubes ($\theta = \pi/6$) are insensitive to axial strain but maximally sensitive to torsional strain. Zigzag

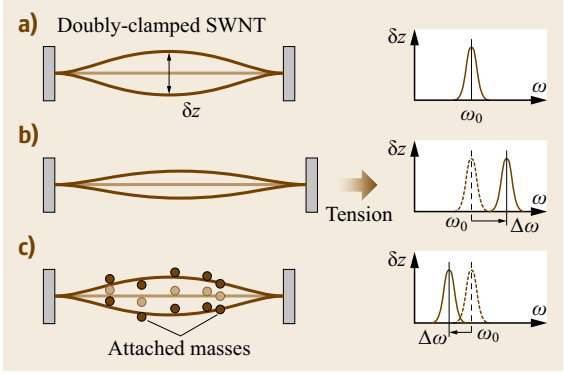


Fig. 14.2 **(a)** Sketch illustrating a doubly clamped SWNT beam and a resonance in the frequency response spectrum corresponding to the fundamental bending mode ω_0. **(b)** With applied tension the resonance frequency ω shifts upwards, whereas **(c)** with attached particles (mass loading) ω shifts downwards

tubes ($\theta = 0$) are just the opposite, whereas chiral tubes are sensitive to both types of strain.

For ordinary semiconductors such as silicon the conductance varies exponentially with band gap, i. e., $G \approx \exp(-E_g/(k_B T))$. This hindsight, has led to many sensor concepts (strain, torsion, force, pressure, etc.) that involve SWNTs as strain gauges [14.8–12]. More details about piezoresistive properties of SWNT in CNFET configurations will follow in Sect. 14.3.2.

SWNTs are Optically Active, Direct Gap Materials

Inspecting Fig. 14.1 reveals that sub-band extrema come in pairs at the same lattice wave number k (electron-hole symmetry). SWNTs are therefore optically active materials. The absorption of photons generates electron-hole (e^-–h^+) pairs, and conversely, e^-–h^+ recombination occurs over the nanotube band gap via photon emission. Resonant absorption of photons happens at energies corresponding to transitions between symmetric (with respect to E_F) van Hove singularities. These transition energies are labeled $E_{nn}^{s/m}$, where s/m denotes semiconducting or metallic, and $n = 1, 2, \ldots$ is the index of the van Hove singularity in increasing energy order ($E_{n+1n+1}^{s/m} > E_{nn}^{s/m}$, with $E_{11}^{s} \equiv E_g$).

Excited e^-–h^+ pairs with energies $E_{nn}^{s/m} > E_{11}^{s/m}$ (i. e., $n > 1$) de-excite nonradiatively into lower energy pairs whenever there is a continuum of states below $E_{nn}^{s/m}$, until an energy gap is reached. Therefore, no photoluminescent emission is possible for m-SWNTs, whereas s-SWNTs emit mainly over their band gap. Transition energies $E_{nn}^{s/m}$ cover a wide spectrum ($\subset (-9, 9)$ eV), thus SWNTs absorb light in UV (> 3.1 eV), visible and IR (< 1.8 eV) domains, but they emit mostly in IR ($d_t > 0.5$ nm corresponds to $E_g < 1.5$ eV). The basic optical properties of SWNTs are summarized in Fig. 14.4. In addition, because of the 1-D character, SWNTs only absorb and emit light linearly polarized along their axes. Optical properties of SWNTs have been exploited in devices such as polarized photodetectors [14.19, 20].

SWNTs are Sensitive to Molecule Adsorption

SWNTs are hollow structures, made of surface only. Molecular adsorption is therefore expected to have a huge impact on CNT electronic properties. Theoretically, the interaction of molecules with carbon nanotubes is a complex topic that can only be treated ab initio from quantum mechanics [14.36]. Generally speaking, molecules either physisorb or chemisorb at the CNT surface. Bond breaking is rare, as the sp^2 hybridized carbon network is very stable. Local sp^3 hybridization is however possible, and this adsorption mechanism is enhanced by curvature in small-diameter carbon nanotubes [14.37]. Defect sites in the nanotube wall greatly enhance adsorption as well [14.38].

At low molecular coverage, doping is the main effect, which is equivalent to a simple charge neutrality level shift E_F with respect to mid-gap (Fig. 14.5a). However, at large coverage densities, the bands of the tube itself can be distorted (by mixing with molecular orbitals). In addition to the charge neutrality level shift, in this regime, new bands may appear inside the nano-

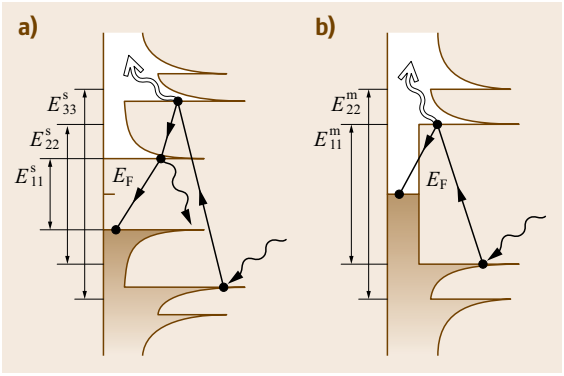


Fig. 14.4a,b Optical-related processes taking place in an SWNT after photon absorption. *Wiggly single lines* represent photons, solid lines electron transitions (excitation/de-excitation) and *wiggly double lines* stand for nonradiative processes (e.g., phonons). **(a)** A semiconducting tube can re-emit photons over the band gap, whereas in **(b)** metallic tubes, de-excitation is mostly nonradiative

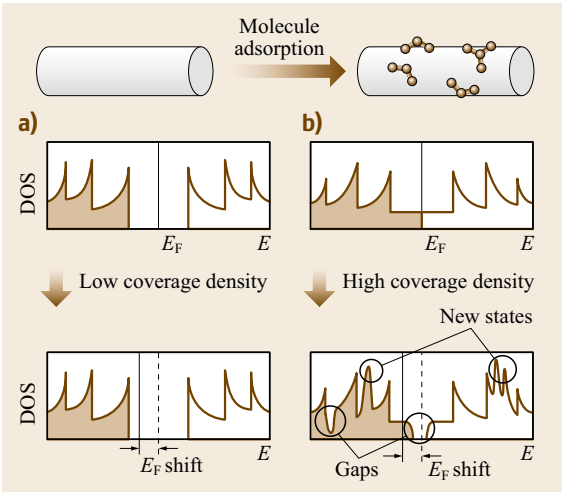


Fig. 14.5a,b The effect of molecular adsorbates on SWNT electronic properties. **(a)** An s-SWNT in the low coverage limit, suffers only a shift in the Fermi level E_F. **(b)** In the high coverage limit, new gaps and states, E_F shifts, and even work function changes are possible

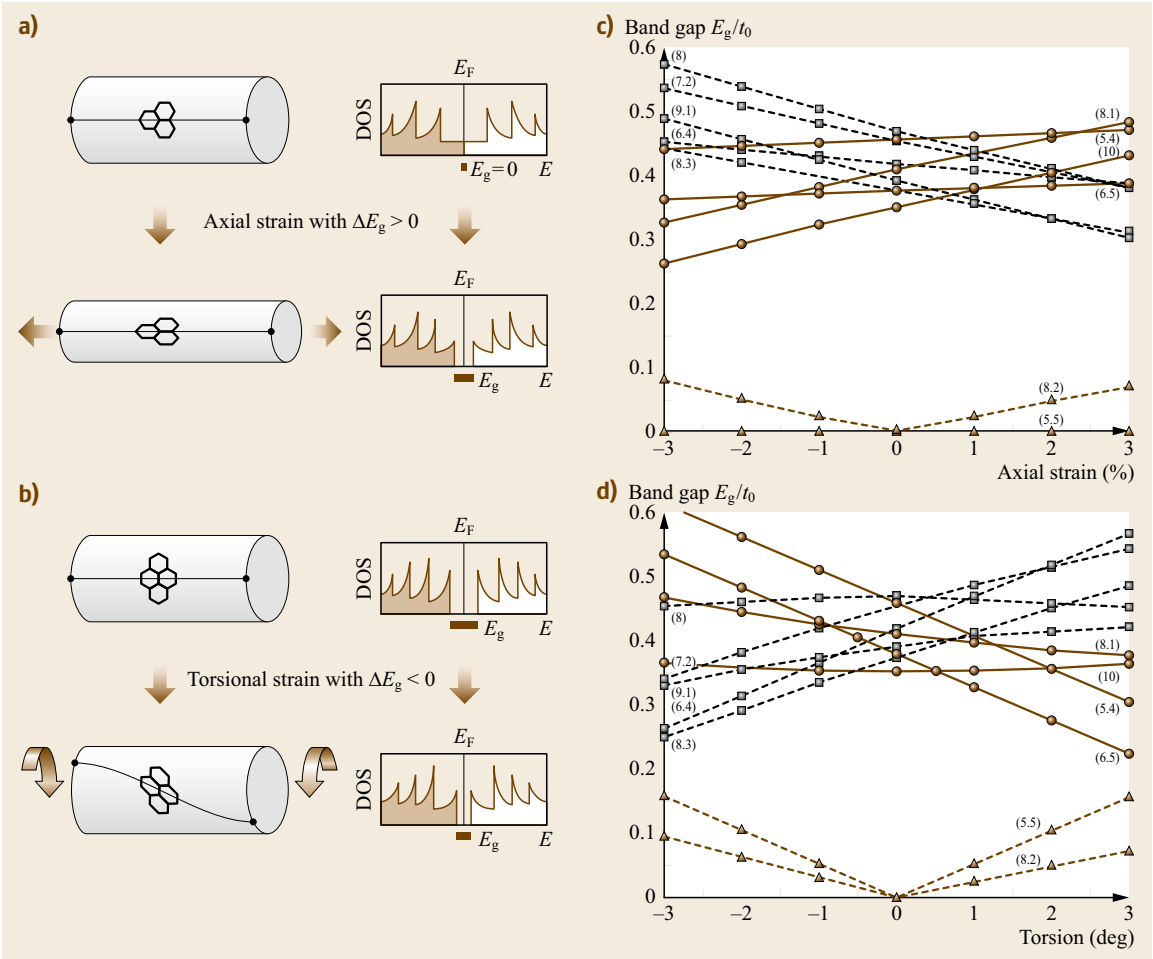


Fig. 14.3a–d Influence of strain on electronic properties of SWNTs. **(a,b)** Density of states modulation by strain. The example in **(a)** corresponds to tensile strain applied to a metallic zigzag CNT $p = 0$ (band gap opening), and the one in **(b)** to torsional strain applied to a chiral CNT with $p = 1$ (band gap closing). **(c,d)** Calculated band gaps as a function of tensile **(c)** and torsional **(d)** strain, for several chiralities belonging to each one of the three families ($p = 0$ *dotted*, $p = 1$ *dashed*, and $p = -1$ *solid lines*) (**(c,d)** after [14.34])

tube gap, or new gaps may open within the conduction or valence bands [14.39] (Fig. 14.5b). Even the electron affinity/work function of the nanotube can be modified in the process [14.37]. In m-SWNTs (as well as in the valence or conduction band of s-SWNTs), adsorbing molecules may create a disorder potential which results in band gap opening at the Fermi level.

Chemical and biochemical sensors based on the modulation of SWNTs' electronic properties upon exposure to target molecules, have been demonstrated in [14.2, 7]. The most widely used method for reading out these modulations is via transport measurements (CNFETs). However, capacitive [14.29] or optical [14.40] detection have also been employed. Chemical sensors will be treated in Sect. 14.3.1.

14.1.2 Carbon Nanotube FET Structures

The preceding section has focused on the properties of SWNTs and property modulation subject to mechanical, chemical, or optical stimuli. However, these properties are intrinsic to isolated SWNTs, and in order to become useful they need to be probed from the macroscopic world. A simple way to probe CNT properties is to electrically contact the nanotube with metallic leads. In combination with a nearby gate electrode it is a configuration known as the carbon nanotube field-effect transistor (CNFET). A basic CNFET is sketched in Fig. 14.6, and consists of an individual SWNT (the channel) placed on an insulator (typically SiO_2), contacted by source and drain electrodes, and gated by the

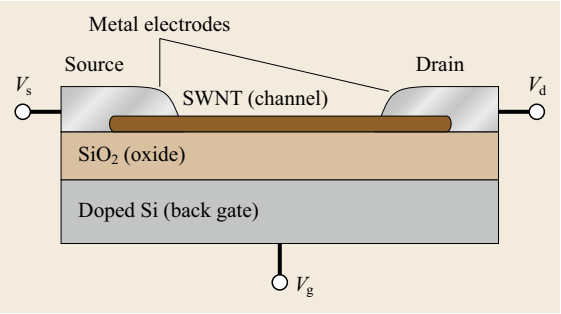


Fig. 14.6 Schematic representation of a simple SWNT field-effect transistor in cross-sectional view

conducting substrate (typically highly doped Si). The fabrication of such a structure (and of more advanced CNFETs) is covered in Sect. 14.2.2. Before giving examples of sensor demonstrators in Sect. 14.3 we review the basic operation and particularities of CNFETs. Understanding CNFET operation is important in identifying to what extent contacting SWNTs by metal leads affects their properties.

The right frameset to study transport in carbon nanotubes is mesoscopic physics, since nanotubes are crystalline, low defect density conductors, for which quantum mechanical effects are expected to play a role. Now, in a general conductor with length below the elastic mean free path (ℓ_e) and phase relaxation length (L_φ), at low temperature and low bias (quasi-equilibrium), each mode (Bloch wave) carries a current of $2e/h$. Therefore, the total source and drain current I_{sd}, can be obtained by summing the average number of modes M, going from source to drain, with energies in the range (eV_s, eV_d) centered at the Fermi level E_F. This yields $I_{sd} = (2e^2/h)MV_{sd}$ [14.41], where e, h are the usual electron charge and Planck's constant respectively, and $V_{sd} = V_s - V_d$ is the applied source-drain (low) bias. In other words, the sample conductance $G = I_{sd}/V_{sd}$, is simply MG_0, where the quantum conductance $G_0 = (2e^2/h) = (1/12.9)\,\mathrm{k\Omega}^{-1}$ was introduced.

For an m-SWNT, at any energy close to E_F (to within ± 1 eV), there are precisely two subbands (Fig. 14.1d), meaning that $M = 2$ and the predicted conductance of a metallic tube is $G = 2G_0\,(= 1/6.5\,\mathrm{k\Omega}^{-1})$. Conductance values approaching $2G_0$, have indeed been measured experimentally [14.42, 43] for m-SWNTs. However, in most situations, G is well below $2G_0$. One reason is that the metallic leads used to contact the tube differ in geometry, lattice and electronic properties (most notably work function and effective mass) from SWNTs. The heterojunction between a metal and the SWNT incorporates interface barriers that introduce scattering and reduce the electron transfer rates. Another reason for reduced conductance $G < 2G_0$ is disorder in the nanotube, caused by structural defects, adsorbed molecules, or simply the random potential profile of the underlying oxide (trapped charges, dangling bonds, etc.). Accordingly, in order to include interface and disorder, the conductance is reduced by a multiplicative factor $\mathcal{T}\,(G \to \mathcal{T}\,2G_0)$, describing the average transmission probability for an electron to propagate from source to drain. At room temperature, for long enough nanotubes, phonon scattering leads to incoherent, classically diffusive transport, respecting Ohm's law ($\mathcal{T} \approx 1/L$).

As mentioned previously, for m-SWNTs the transmission $\mathcal{T}$ can approach 1, which corresponds to thin contact barriers, easy to tunnel by electrons, and clean SWNT surfaces. For s-SWNTs, the situation is different. The physics of heterostructures predicts that at the interface between a metal and a semiconductor, Schottky barriers (SB) decaying slowly within the semiconductor may arise. Figure 14.7a depicts schematically the band diagram at the interface between a metal, of work function ϕ_m, and an intrinsic s-SWNT, of electron affinity χ_s ($\chi < \phi_m < \chi_s + E_g/2$). There are two such barriers, one at the source and the other at the drain. For simplicity, assume low V_{sd} bias (quasi thermal equilibrium), that allows defining a flat Fermi level throughout the structure. Assume as well that the conductance of the device can be written as $1/G = 1/G_s + 1/G_t + 1/G_d$, with $G_{s,t,d}$ the conductance of the source barrier, nanotube channel and drain barrier, respectively. The last assumption ignores interference effects and holds only if the length L of the nanotube segment exceeds L_φ [14.41]. Although still debated, this seems to be the case for SWNTs with $L \geq 1\,\mu\mathrm{m}$ at room temperature [14.44]. Within this picture, $G_d = G_s$ and we can focus on just one interface.

At finite temperature ($T \neq 0$), the charge flow across any metal-semiconductor junction, characterized by some interface barrier ϕ_B, has two components: tunneling through ϕ_B and thermionic emission over ϕ_B. At $T = 0$ only direct tunneling at E_F is possible. There are two such barriers, one for electrons ϕ_B^n and one for holes ϕ_B^p. In Fig. 14.7a, corresponding to $V_{gs} = 0$ and an intrinsic s-SWNT, $\phi_B^n = E_g/2$, $\phi_B^p = E_g + \chi_s - \phi_m$ and only thermionic emission is possible. In this situation I_{sd} is practically 0. A positive voltage $V_{gs} > 0$ applied on the gate (Fig. 14.7b) shifts downwards the bands in the bulk of the nanotube, but not the position of the band edges at the interface with respect to E_F. Thermionic emission of electrons increases while it decreases for holes. At some point, when the nanotube conduction band E_C becomes aligned to E_F, $\phi_B^n \equiv \phi_{SB}^n = \phi_m - \chi_s$ and $\phi_B^p = E_g$. The barrier ϕ_B^n is also a lot thinner than at $V_{gs} = 0$ because the electron density inside the nanotube is higher, screening any interface

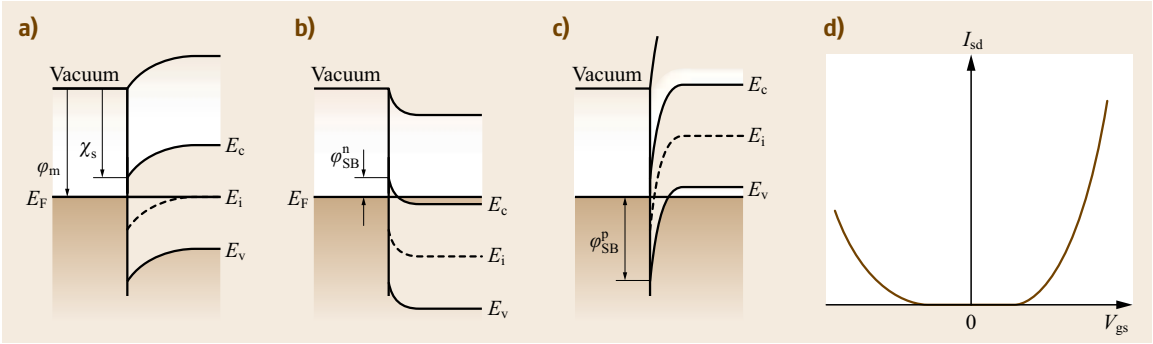


Fig. 14.7a–d Band structure diagram at the interface between a metal and an s-SWNT. **(a)** No gate bias applied $V_{gs} = 0$, **(b)** $V_{gs} > 0$ right above the threshold for electron conduction, and **(c)** $V_{gs} < 0$ right below the threshold for hole conduction. **(d)** The $I_{sd}(V_{gs})$ characteristic of the transistor, emphasizing asymmetric electron versus hole conductivity, due to $\phi^p_{SB} > \phi^n_{SB}$

potential more effectively. Tunneling of electrons increases and will eventually exceed thermionic emission (see below). Above this threshold, increasing V_{gs} further thins down the barrier, increasing I_{sd}. The barrier width depends, besides the position of E_F, on the geometry of the leads and oxide thickness [14.45]. Similarly, for negative, large enough gate voltage, the bulk valence band E_V becomes aligned to E_F (Fig. 14.7c). In this case $\phi^p_B \equiv \phi_{SB} = E_g + \chi_s - \phi_m$, $\phi^n_B = E_g$, and hole tunneling dominates. The inequality between the electron and hole barriers (in this example $\phi^p_{SB} > \phi^n_{SB}$) translates into asymmetric electron versus hole conductivity (the two branches of the $I_{sd}(V_{gs})$ characteristic), sketched qualitatively in Fig. 14.7d.

As opposed to normal, planar junctions (e.g., metal-silicon), Fermi level pinning is proposed not to exist in metal-CNT junctions, because of the particular tubular geometry of the CNTs [14.46]. As a result, different metals lead to barriers of different heights, which should also depend on the CNT diameter d_t (d_t determines the band gap E_g as discussed in Sect. 14.1.1). This fact, evidenced experimentally [14.47], is a strong confirmation of the Schottky barrier model of CNFETs. Consequently, it seems possible to completely eliminate the Schottky barrier by either choosing large work function metals such as Au, Pt, Pd ($\phi_m \approx 5.1$ eV) to cancel ϕ^p_{SB}, or low work function metals such as Al ($\phi_m \approx 4.2$ eV) to cancel ϕ^n_{SB}, knowing that the mid gap work function of SWNTs is around 4.5 eV. On the basis of this principle *Javey* et al. [14.48] have demonstrated G within 50% of the ideal $2G_0$. More experimental evidence in support of the SB-CNFET model has been brought by *Appenzeller* et al. [14.49, 50]. By measuring the temperature dependence of the $I_{sd}(V_{gs})$ characteristic [14.50], they concluded that tunneling is the main injection mechanism in SB-CNFETs and not thermionic emission. Work function alone, however, is not enough, as for example Pt ($\phi_m \approx 5.6$ eV) yields poor contacts, likely due to poor wetting or native oxide tunnel barriers [14.51].

For certain CNFET sensor devices, the Schottky barrier FET model is essential in understanding the device operation and sensing mechanism. This is the case of (bio)chemical sensors to be discussed in Sect. 14.3.1. In other devices, Schottky barriers, if present at all, contribute only to the contact resistance that is in a first approximation constant during the sensing process, and can be thus factored-out from the sensing mechanism. This is the case with piezoresistive gauges and resonators as presented in Sects. 14.3.2 and 14.3.3, respectively. For these devices, the transport picture introduced in this section, in terms of thermionic emission and tunneling, is still useful in explaining the operation. The only difference is that potential barriers for electrons and holes will be found on the nanotube and not only at the interface (Fig. 14.16c,d). Further insight into the physics of CNFETs in different regimes (large bias, low-temperature phenomena such as Luttinger liquid, Coulomb blockade, or orbital Kondo effects, etc.) may be obtained from [14.52–54].

14.1.3 Sensor Characterization

Sensor characterization refers to assessing performance metrics for a certain sensing technology. Some of the most important performance metrics are: sensitivity, signal-to-noise ratio (SNR), limit-of-detection (LoD), cross-sensitivity/selectivity, signal rise/fall time (speed), repeatability, offset/sensitivity, drift, hysteresis, and lifetime/robustness. So far, mainly *sensitivity* and *dynamic properties e.g., signal rise/fall time* have been targeted by the CNT sensor community (see results in Sect. 14.3). This is a natural situation, considering that the technology for fabricating individual

SWNT FETs is not yet scalable (Sect. 14.2). The efforts have been then focused on demonstrating that CNT nanosensors are superior to existent devices, so as to motivate and justify future investment. Conversely, the lack of batch fabrication processes has hindered gathering statistical information about characteristics of CNT sensors in operation; most publications on individual SWNT sensors (in particular FET devices) refer to data obtained from just a few fabricated samples. Therefore, *reproducibility* of sensor characteristics has been ignored, or marginally addressed.

To better illustrate matters, consider the story of chemical CNFET sensors. In 2000 *Kong* et al. [14.2] first report detecting NO_2 and NH_3 with individual SWNT FETs. In 2008, *Cho* et al. (in collaboration with Kong) published first results for a hybrid CNT chemical sensor with a CMOS interface chip [14.55]. The fabrication of CNT sensor arrays is still far from scalable, yet by this approach the interface captures signals from up to 24 CNFETs in parallel, further multiplexed and sent to analysis. The authors manage to measure 414 devices which results in a histogram of the distribution of resistance, spread over 6 orders of magnitude. This large dispersion is the result of varying number and type of CNTs in the devices, which is not controlled during fabrication. The dispersion is also visible in the recorded sensor response on the measurand. This kind of platform is a first step towards capturing device performance statistics, a prerequisite for process control and optimization.

Acquiring sensor *signal-to-noise-ratio* is currently under-progress. Electronic noise in CNFETs has received a lot of attention lately. *Collins* et al. [14.56] were the first to measure $1/f$ noise in SWNTs, both individual tubes and mats. The noise spectrum was found to agree with the classical noise power formula $S_V(f) = (A/f^{\beta})V^2$, where A is the noise amplitude, f the frequency, $\beta \approx 1$, and V the biasing voltage. Furthermore, the noise amplitude seems to obey the empirical Hooge law $A = \alpha_H/N$, with α_H the Hooge constant and N the number of carriers in the channel, because on average A decreases with the number of tubes in the sample [14.56]. Interestingly, the $S_V(f)$ formula is expected to hold for classically diffusive transport which is questionable for CNFETs. However, *Appenzeller* et al. [14.57] have calculated the average number of carriers inside SWNT ballistic FETs, for different contact metals, and found the Hooge law to be valid even in this regime, with a fitted α_H value of $\approx 2 \times 10^{-3}$ (same as for bulk silicon). Upon closer examination, it has been observed by Tersoff [14.58] that the Hooge law is not accurate in the subthreshold region of a CNFET. He proposed adding a phenomenological term to the noise power proportional to $(dI_{sd}/dV_{gs})^2$, that accounts for gating of the CNFET by fluctuating charges in the vicinity of the tube. The model has been recently confirmed experimentally in [14.59]. Overall, regardless of the model utilized, the noise in CNFETs is found to be significant, which limits the SNR of CNT sensors. The fact that noise is proposed to be mainly extrinsic [14.57], i. e., caused by external fluctuators, is however encouraging since it sets clear technical objectives in achieving better control of the CNT environment.

14.2 Sensor Fabrication: SWNT Synthesis and Integration

Fabrication of CNT devices involves many different aspects. These can roughly be grouped in two main tasks, namely *synthesis* of SWNTs and *integration* (or assembly) of nanotubes into devices. *Synthesis* is concerned with production of SWNTs with controlled properties such as diameter, length, defect densities, electrical conductivity type or even chirality. Sometimes synthesis is followed by postsynthesis methods for CNT purification, sorting, and most importantly functionalization. Conversely, *integration* refers to techniques and methods for placing SWNTs at predefined locations on a substrate with controlled number of nanotubes, orientation and slack (straightness). Postassembly, other processing steps such as nanotube electrical contacting or device encapsulation may follow. Also note that the distinction is sometimes artificial, e.g., in situ nanotube synthesis can be simultaneously both a synthesis and integration method. Complete processes for fabrication of micro- and nanosystems that integrate nanotube devices/sensors with acceptable yield, remains one of the key topics in CNT research. This section attempts to survey some of the available methods and processes aimed at controlled *diameter*, *chirality*, *location*, and *orientation*. Moreover, some consideration will be given to process *cleanliness*. In the last few years it has become increasingly clearer that sensor device performance is heavily affected by contamination of the nanotubes during processing.

As discussed in Sect. 14.1, the intrinsic properties and property modulations of SWNTs depend directly on the nanotube chiral indices. Another, more easily controllable parameter is SWNT *diameter* d_t. For example, the band gap E_g of an s-SWNT is inversely proportional to d_t, and so are Schottky barriers $\phi_{SB}^{n/p}$

in CNFETs. Other properties, most notably piezoresistance, depend as well on *chiral angle* θ. Since d_t and θ of an SWNT are determined during synthesis (production), methods for nanotube production are briefly surveyed in Sect. 14.2.1. *Location* and *orientation* control during synthesis or/and assembly, are important for building complex structures, for developing batch fabrication processes, or for large-scale device integration (sensor arrays). Section 14.2.2 will thus review processes for the assembly of SWNTs into devices, together with some general strategies in achieving location and orientation control. Section 14.2.3 will briefly discuss the importance of maintaining nanotube cleanliness throughout the process.

14.2.1 Methods for SWNT Production

Carbon nanotube production methods can be classified in two categories, namely high-temperature (*arc discharge* and *laser ablation*, $T =$ above 3000 °C) and medium temperature (*catalytic chemical vapor deposition* (CCVD), $T = 400–1100$ °C). Details about each method can be obtained from textbooks [14.21, 23] and also Chap. 8 of this book. Here, only those features that are relevant for CNT devices and integration are discussed.

Because of higher production temperature potentially annealing defects, *arc discharge* and *laser ablation* are expected to produce SWNTs with better crystallinity. However, with these methods it is not possible to produce SWNTs directly on the target substrate for device fabrication; only postsynthesis assembly is possible. Typically, as-produced nanotubes are first dispersed in a liquid that can be subsequently utilized to deposit SWNTs on the target substrate (Sect. 14.2.2). Also, the distribution of diameters is difficult to control with these methods during SWNT production. Nevertheless, after dispersing nanotubes in a liquid, different techniques can be employed to separate semiconducting from metallic and sort by diameter or even by chirality CNTs, as explained below. Sorting and purification techniques are reviewed in [14.60] and separation techniques in [14.61].

In *catalytic chemical vapor deposition*, SWNTs are synthesized on catalyst nanoparticles, particles or seeds and from a carbon-containing gas feedstock [14.62, 63]. Since CCVD is a catalytic process, patterning of a catalyst area can be utilized to grow nanotubes at selected locations directly on preprocessed (silicon) chips (more on this in Sect. 14.2.2). The key observation for CCVD is that *the size of the catalyst particle correlates with SWNT diameter* [14.64], and the *catalyst particle density determines* the final *CNT density* [14.65]. In [14.64] *Cheung* et al. have prepared Fe nanoparticles, with narrow diameter distributions centered around 3, 9 and 13 nm, on SiO_2. The grown CNT diameter distribution mirrors the initial particle size distribution, with a standard deviation roughly 30% of the average. However, 3 nm particles produce SWNTs with double-walled CNTs (DWNT), 9 nm produce D/MWNT with just a few SWNTs, and 13 nm produce only MWNTs, showing that typically a particle size < 3 nm is required for CCVD synthesis of SWNTs. However, as shown in [14.66] even larger particles can produce relatively small-diameter SWNTs by the so-called *perpendicular mode* growth mechanism. For SWNT-based FETs, a good compromise between low contact resistance (low $\phi_{SB}^{n/p}$) and high gap E_g (large ON/OFF ratio) would correspond to diameters in the range 1.5–2 nm [14.47]. Obtaining such small catalyst nanoparticles, while keeping a narrow size distribution, is challenging. *Li* et al. [14.67] have demonstrated SWNT diameter distributions of either 1.5 ± 0.4 nm or 3.0 ± 0.9 nm (again standard deviation/average $\approx$ 30%). Their method involved loading the hollow cavity of the iron-storage protein Ferritin (internal cavity diameter 8 nm) with Fe, followed by deposition on a substrate and calcination of the protein shell. Particle size was controlled by the iron loading time.

Currently, much activity is going into trying to achieve SWNTs with acceptable diameter distributions or even with one or few chiral indices. For CCVD grown tubes, the most promising approach to control nanotube diameter seems to remain improving particle size distribution. Conversely, there are few approaches to produce SWNTs with specific chiral indices via *cloning* [14.68, 69], specific seeds [14.70, 71] or catalysts with defined morphology/structure [14.72, 73].

Postsynthesis, for SWNTs dispersed in solution, regardless of their origin (arc discharge, laser ablation, or CCVD), several methods have been proposed to narrow down the diameter distribution. For example, *Tromp* et al. [14.74] have recently demonstrated diameter separation of SWNTs by noncovalent functionalization with anchor molecules optimally attaching to specific tube diameters (1.2 nm). This functionalization improves solubility of the selected tubes, which are subsequently separated from the nonsoluble bulk via sonication and centrifugation. By using DNA-wrapping and density gradient ultracentrifugation (DGU), *Arnold* et al. [14.75], have shown a good separation by diameter, for SWNTs with diameters below 1 nm. Soon after, the same group [14.76] extended the technique to surfactants, and managed to isolate narrow distributions of SWNTs with $> 97\%$ within 0.02 nm diameter range, with e.g., 84% of the tubes being of (6,5) chiral indices. Furthermore, by using mixtures of surfactants, they were able to produce predominantly

semiconducting or metallic SWNTs. Utilizing different DNA sequence motifs *Tu* et al. [14.77] obtained chirality-selected SWNTs with purity up to 90% and with the separation yield up to 0.8%. Combined with recently proposed continued or cloning growth methods [14.78–80] the DGU method might one day achieve macroquantities of essentially single-chirality SWNTs.

However, to date, achieving SWNTs with a single chirality in reasonable amounts and with simultaneously fulfilled requests for length, structural quality, and cleanliness remains the key challenge in CNT production.

14.2.2 Strategies for SWNT Assembly into Devices

Perhaps the simplest way to survey SWNT assembly strategies is to split the discussion into two parts by considering the nanotube source: *liquid suspension* (laser, arc, or CCVD SWNTs dispersed in liquid) or in situ growth (CCVD SWNTs grown directly on the target substrate). In-depth reviews of the various assembly strategies can be found elsewhere [14.81, 82]. Here, we focus only on those techniques resulting directly in CNFET device structures, or have some relevance to this topic. Regarding liquid suspension or in situ growth, there is an ongoing debate as to which one produces the best results. Table 14.2 compares the two in several aspects, showing a good balance of advantages and disadvantages. In general, each process type could be preferred depending on specific application requests.

Liquid Suspension Assembly Methods

Historically, liquid suspension assembly methods were the first to appear. The very first CNFETs, demonstrated in 1998 [14.83, 84], were built via a similar, simple process. In [14.84] laser-ablation SWNTs were dispersed by sonication in dichloroethane, and then spread over Au electrodes predefined using electron beam lithography (EBL), on a doped Si wafer substrate covered by a thick gate oxide film (SiO_2). The Si substrate, served as a back-gate to all devices. A schematic cross section of a CNFET stack, together with an AFM image of a fabricated device, is shown in Fig. 14.8 [14.83]. The same process, extended with a patterned poly(methyl methacrylate) (PMMA) top layer has been used in demonstrating one of the first CNT logic-inverters [14.85]. The only controlled parameter in this process is tube density, which can be adjusted – taking into account the electrode design – to obtain devices with only one nanotube bridging the correct electrodes, albeit at a very low yield.

It was soon realized that surface functionalization can significantly help in guiding SWNT depo-

Table 14.2 A comparison between liquid suspension and in situ growth assembly methods

SWNT source	CNT quality[c]	Average diameter control[e]	Diameter standard deviation[h]	Chirality control[i]	FET performance[j]	Alignment precision	Ease in integration
Liquid suspension[a]	+/−	−[f]	++	++	−[k]	−/+[l]	+/−[m]
In situ growth[b]	+/−[d]	++[g]	−/+	−/+	+	+	+[n]

[a] CNTs as produced by arc discharge or laser ablation
[b] In situ growth mainly by CCVD (Sect. 14.2.1)
[c] SWNT quality is attributed to crystallinity and defect density (Sect. 14.2.1)
[d] Progress in CCVD leads to constant increase in CNT quality. Measures can include postannealing, H_2 treatment
[e] SWNT diameters influence Schottky barrier heights in CNFETs (Sect. 14.1.2)
[f] The average diameter of arc discharge and laser ablation SWNTs is between 1.2 and 1.5 nm and is difficult to control
[g] The average diameter of SWNTs correlates with catalyst particle size and growth conditions and can be tuned (Sect. 14.2.1)
[h] A small diameter standard deviation should reduce performance variations in CNFET sensor devices
[i] Functionalization of SWNTs, sonication and centrifugation postsynthesis narrows diameter distribution. Individual chiralities may be selected (Sect. 14.2.1). Only applicable to SWNTs assembled from liquid suspension
[j] FET performance for sensing applications refers to low contact resistance, low noise, and high ON/OFF ratio
[k] Liquid suspension involves CNT surface treatment by (strongly binding) surfactants, sonication, and centrifugation. These steps may degrade CNFET performance
[l] Localized surface functionalization and dielectrophoresis increases the alignment precision. However, bundling of individual CNTs is still an issue in these processes
[m] Wet processing is better compatible with CMOS IC substrates, but wet processing may incorporate impurities on SWCNT and influence the SWNTs' electronic properties
[n] In situ growth is compatible with MEMS (microelectromechanical system) substrates including suspended MEMS and NEMS (nanoelectromechanical system) structures

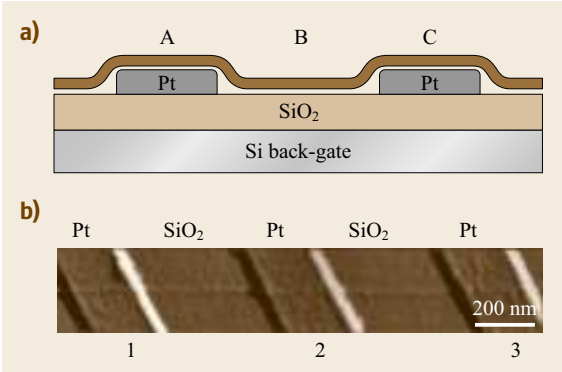


Fig. 14.8a,b First CNFET, **(a)** schematic view, and **(b)** AFM image (after [14.83])

sition. In [14.86] *Liu* et al. observed that SWNTs preferentially adsorb on amino-functionalized surfaces. Subsequently, they developed a process to define amino-functionalized regions where tubes are adsorbed, onto a tube-repelling surface (a trimethylsilyl self-assembled monolayer (SAM)). The process summarized in Fig. 14.9a, was utilized to place an SWNT between two contacts, as shown by the AFM image in Fig. 14.9b. Since then, other groups have utilized the same principle but with different nanotube-surface chemistries, for example the hydroxamic acid group-Al_2O_3 [14.87], patterned aminosilane monolayer [14.3], or polar-nonpolar interfaces [14.88]. Using an approach based on ion-exchange chemistry, *Park* et al. [14.89] recently demonstrated the first high-density integration of carbon nanotubes into transistors for logic applications.

Another important process family for SWNT assembly from solution is based on electric fields. *Krupke* et al. [14.90] used AC dielectrophoresis to place SWNT bundles from suspension, between predefined electrodes. Electrode arrays were defined using EBL, and connected to a frequency generator. On top of the electrodes a droplet of the CNT suspension was applied and evaporated. In general, $\approx 70\%$ of the electrodes were bridged by at least one bundle, and even individual SWNTs bridged occasionally. The same group [14.91], proposed that in D_2O (heavy water) with dielectric constant $\varepsilon_{D_2O} = 80$, s-SWNTs ($\varepsilon_s < 5$) will experience negative dielectrophoresis (repulsion from electrodes) and m-SWNTs ($\varepsilon_m \to \infty$) positive dielectrophoresis (attraction to electrodes). They showed that under these conditions m-SWNTs are indeed deposited on electrodes, while s-SWNTs remain in solution, thus providing a basic method to separate metallic from semiconducting tubes. Further progress in assembly processes based on dielectrophoresis are reviewed in [14.92].

In Situ Growth Assembly Methods

The second major class of assembly methods is based on *in situ growth* (via CCVD) of SWNTs directly on targeted (silicon) chips/wafers. CCVD has one major advantage over liquid suspension deposition regarding location control, namely the catalyst can be patterned [14.93]. The first process for fabricating CNFETs using CCVD was proposed by *Soh* et al. [14.94]. EBL was first used to pattern wells into which catalyst was deposited. A resist lift off left isolated islands with catalysts, from which SWNTs were grown by CCVD. Metal electrodes are defined on top of the catalyst is-

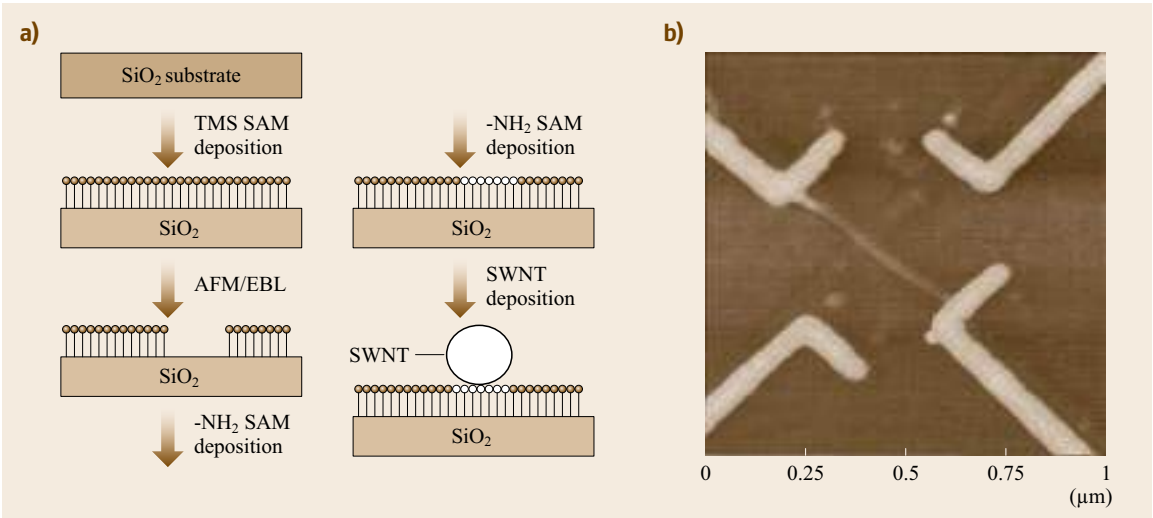


Fig. 14.9 **(a)** Schematic diagram of process for depositing SWNTs on functionalized surfaces. **(b)** AFM image of a device made on Au electrodes (after [14.86])

lands by lithography, contacting some of the grown tubes. The location of tubes is thus approximately controlled, as is the number of nanotubes between two electrodes by changing catalyst particle density and tuning growth parameters. Nevertheless, considerable progress has been made in catalyst deposition and patterning. *Javey* and *Dai* [14.95] developed a method that allows positioning of *individual* catalyst particles with EBL resolution. Furthermore, tube-to-particle number ratio approaching 1/1, i. e., close to 100% SWNT growth yield was achieved.

Since nanotube location control can be achieved by catalyst patterning, a lot of work subsequently focused on orienting the CCVD growth. Electric *field-directed growth*, has been pioneered by *Zhang* et al. [14.96], exploiting the large anisotropic polarizability of SWNTs. Polysilicon elevated structures were first defined onto a quartz substrate by optical lithography, with a catalyst layer on top, transferred by contact printing. Electric fields were applied during growth via outer poly-Si pads contacted by metal leads (electrodes). In Fig. 14.10 both the process flow and growth results are displayed, with/without applied fields. Field-directed growth was subsequently demonstrated on flat surfaces as well [14.97], and catalyst patterning and field-directed growth have been combined to build CNFETs by *Dittmer* et al. [14.98]. Another approach for orienting CNT growth was proposed in [14.99], where a strong correlation between the nanotube orientation and feedstock *gas flow direction* was observed. Finally, numerous recent studies show that *surface-directed growth* is also possible on substrates such as A- and R-planes of sapphire [14.100] or miss-cut C-plane sapphire [14.101]. The oriented SWNTs on sapphire have been contacted into FET configurations by *Liu* et al. [14.102]. These and other approaches are reviewed in [14.103].

As discussed in Sect. 14.1.1, SWNTs have interesting electromechanical properties, which often require free-standing (suspended) nanotube segments in order to manifest. Assembly of SWNTs into suspended microstructures, tends to be easier with *in situ growth* than from *liquid suspension*, because of complications arising from capillary forces. Directed growth of SWNTs from predefined silicon towers, was first shown by *Cassell* et al. [14.104] and [14.105], via a process that preceded the field-directed growth in Fig. 14.10. Basic electromechanical structures were presented later by the same group, with a process involving patterned CCVD growth directly from the surface of Mo, a refractory *metal* capable of withstanding CCVD temperatures ($\approx 900\,^{\circ}\mathrm{C}$) [14.106] and inhibiting catalyst particle diffusion. Integration of CCVD nanotube growth into a state-of-the-art fabrication process for microelectromechanical systems (MEMSs) has been demonstrated recently by *Jungen* et al. [14.107]. Both location and orientation control were achieved by confining the growth of the tubes geometrically between sharp poly-Si tips.

We end the survey of CNT assembly methods with one of the oldest processes for building SWNT devices, yet still the Swiss knife for rapid demonstration (proof of concepts) and basic investigations. The method proceeds from SWNTs already present in low density on the substrate, whether from liquid suspension or grown in situ. Location of nanotubes is recorded by AFM imaging, and electrodes are defined on nanotubes by e-beam lithography, metal evaporation and lift off. Since

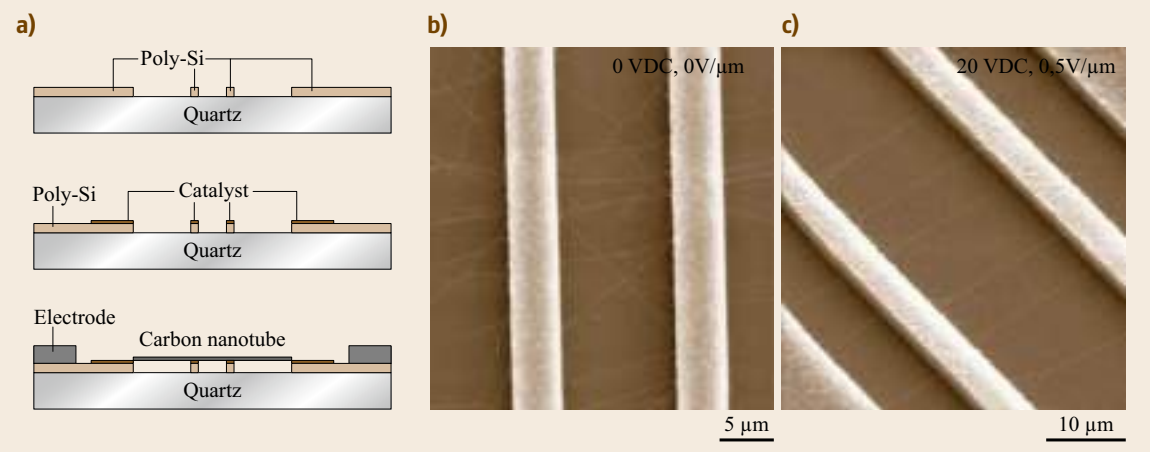


Fig. 14.10 **(a)** Schematic diagram process flow for electric-field-directed growth of SWNTs. **(b,c)** SEM images of SWNTs, grown in various fields. At zero field, tubes grow randomly **(b)**, whereas alignment is seen in 0.5 V/µm **(c)** (after [14.96])

the location and orientation of the tube is known, some basic yet precise micromachining is also possible in this approach. For example, *Witkamp* et al. [14.18] use a PMMA layer to define an etch mask over a top-contacted CNFET. The SiO_2 underneath the SWNT is etched away in HF, releasing the central part of the tube (Fig. 14.11). The resulting suspended nanotube device is an electromechanical resonator to be discussed in Sect. 14.3. The process is time-consuming (because of AFM imaging and EBL), area inefficient (because of nonaligned sparse tubes), and therefore nonscalable.

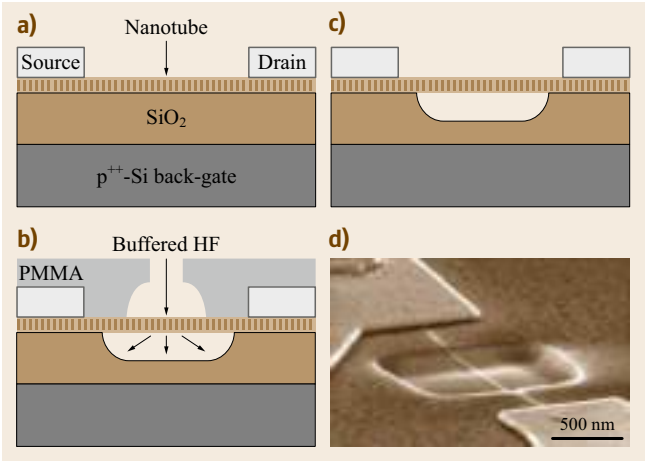


Fig. 14.11 (a–c) Schematic overview of an EBL fabrication process for SWNT resonators. (d) SEM image of a device (after [14.18])

However, currently it is the only process to guarantee individual SWNT devices with optimum alignment precision with respect to electrodes and other postdefined structures.

14.2.3 Process Contaminations and Sensor Performance

Though it was quickly established that structural defects strongly affect the properties of SWNT, the potential impact of process contaminations took more time before receiving proper attention. Indeed, the last years have produced a consistent body of evidence pointing in the same direction: integration processes are equally important to synthesis processes in regard to device quality.

A first striking example came with the work of *Muoth* et al. [14.108] who demonstrated *hysteresis-free* CNFETs in normal, humid air conditions (Fig. 14.12). This was made possible by growing individual, suspended ultraclean SWNT between two suspended tips which were subsequently metallized via shadow-masked metal deposition. This process avoids any wet chemistry or contact with lithography resists and, as such, it leaves the surface of the SWNTs, pristine, uncontaminated.

If a suspended SWNT architecture is incompatible with the application, there are some measures that can improve the quality of substrate-bound CNT devices as well. *Lee* et al. [14.109] and *Liu* et al. [14.110]

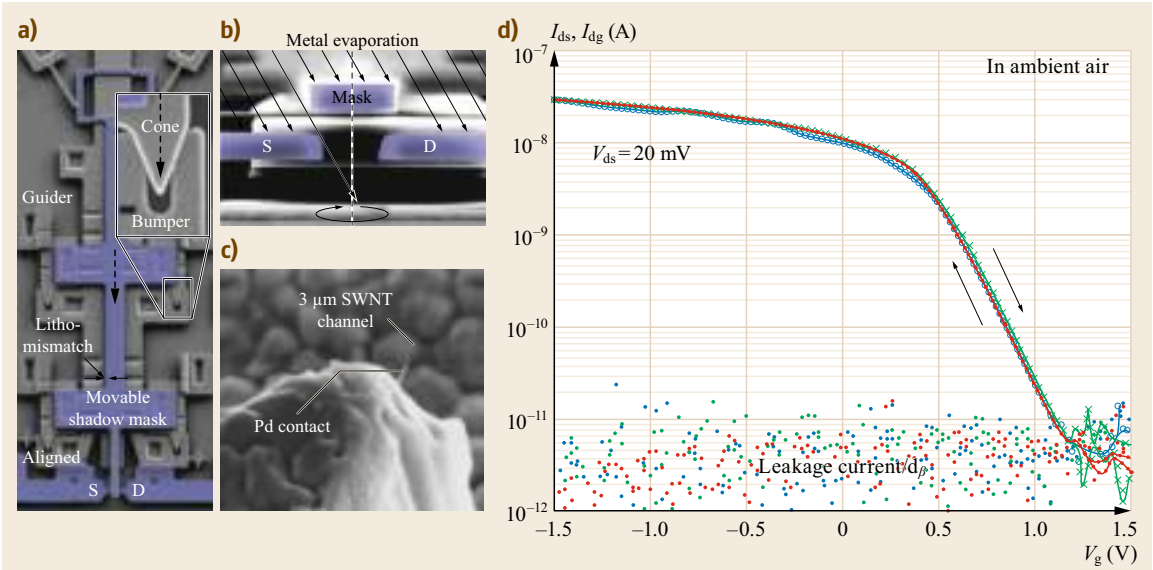


Fig. 14.12a–d Fabrication and demonstration of a hysteresis-free SWNT FET (a–c) SEM images of the movable shadow mask (a), tilted metal evaporation concept (b), and needle-like metallized nanotube contact (c). The transfer characteristic in (d), recorded under ambient conditions, shows no visible sign of hysteresis (after [14.108])

made use of a protective layer of Al_2O_3. This layer was deposited by atomic layer deposition right after the growth of SWNTs onto an SiO_2 substrate. All subsequent lithography steps – e.g., structuring of the top metal contacts – were performed on top of this protective layer. If needed the alumina layer could be stripped by H_3PO_4. This process results in reductions in hysteresis width down to 0.5 V in ambient air [14.109] and in improvements of device resistance [14.110], achieving 1.8 times reduction in the median value but also of standard deviation. The message is that if resists come into contact with CNTs, they are subsequently very hard to remove. The presence of resist residuals on the surface of SWNTs results in a significant degradation of device performance and should be avoided.

14.3 Summary of State-of-the-Art, Applications Examples

In the first section the intrinsic properties of SWNTs and the operation principle of CNFETs were introduced. Fabrication of CNFET devices, including a wide range of different SWNT production methods and various assembly strategies, was covered in the preceding Sect. 14.2. Here we finally show, by way of examples, how various properties of SWNTs assembled in CNFET configurations can be turned into sensor functions. This section will explain the concept and operation, and discuss sensing mechanisms for a few selected device demonstrators, rather than provide a thorough state-of-the-art review. References to review articles are given if available. An attempt will be made to discuss the operation of the selected SWNT sensor examples through the perspective of the CNFET model laid out in Sect. 14.1.2. In support, whenever possible, band diagrams (Schottky barriers, nanotube body bands) and their modulation by stimuli and/or bias conditions will be provided. The following subsections will cover *chemical and biochemical sensors* (Sect. 14.3.1), and *physical sensors* (strain-, pressure-, force- and mass) either piezoresistive (Sect. 14.3.2) or resonators (Sect. 14.3.3).

14.3.1 Chemical and Biochemical Sensors

Shortly after the first CNFETs were demonstrated, it was noticed that the electronic properties of these devices are very sensitive to environmental conditions. Specifically, *Collins* et al. [14.4] have shown that the resistance of SWNTs changed reversibly by up to 15% upon cycling the environment from air to vacuum, a phenomenon attributed to O_2 doping. Since then, both the number and diversity of proposed (bio)chemical SWNT sensors have grown tremendously; perhaps, this class of sensors is the largest in terms of number of publications. Excellent reviews on chemical and biochemical sensors based on carbon nanotubes are given in [14.111–113].

The first gas sensor based on an individual SWNT FET was demonstrated by *Kong* et al. [14.2]. The device is a back-gated FET, fabricated with a process as described in [14.94]. Such CNFET devices were exposed to NO_2 and NH_3 diluted in Ar or air, resulting in a significant change in their conductance. Figure 14.13a shows the response of a sensor to 200 ppm of NO_2. The conductance increased over three orders of magnitude in the time range of 2−10 s. The same device was exposed, after recovery, to 1% NH_3 (Fig. 14.13b), resulting in a 100-fold conductance decrease within 1−2 min. Upon measuring the $I_{sd}(V_{gs})$ characteristic (Fig. 14.13c) the authors conclude that NH_3 depletes the initially p-type SWNT of holes, shifting E_F away from the valence band, whereas NO_2 does the contrary, enhancing the hole concentration and pushing E_F closer to the valence band.

The exact mechanism responsible for the change in the $I_{sd}(V_{gs})$ characteristic – attributed to doping by *Kong* et al. in their original paper [14.2] – remains a controversial issue. Molecules can adsorb on the nanotube or onto the metal contacts or both, but transport in these different situations is expected to change in a qualitatively different fashion. If the molecules adsorb on the nanotube body, doping the nanotube and thus shifting its charge neutrality level E_F, the result is simply a *rigid shift* of the $I_{sd}(V_{gs})$ characteristic as sketched in Fig. 14.14b. Conversely, molecules attaching mainly to the contacts modify the metal work function ϕ_m. This modifies the relative height of the Schottky barriers $\phi_{SB}^{n/p}$, which translates into a changed polarity of the device, as illustrated by the $I_{sd}(V_{gs})$ characteristic in Fig. 14.14a. Molecules could also adsorb on the nanotube regions close to the contacts, changing the electron affinity of the SWNT χ_s or the depletion region, which also affects the electron and hole Schottky barrier heights $\phi_{SB}^{n/p}$ or widths. These two signatures were observed experimentally in [14.45] and assigned to one of the two mechanisms with the help of device simulations. Exposing a CNFET to potassium showed the *tube doping* signature, and oxygen showed the *contact adsorption* signature (Fig. 14.14c,d).

The sensing mechanism becomes further convoluted by the presence of the gate oxide. *Helbling*

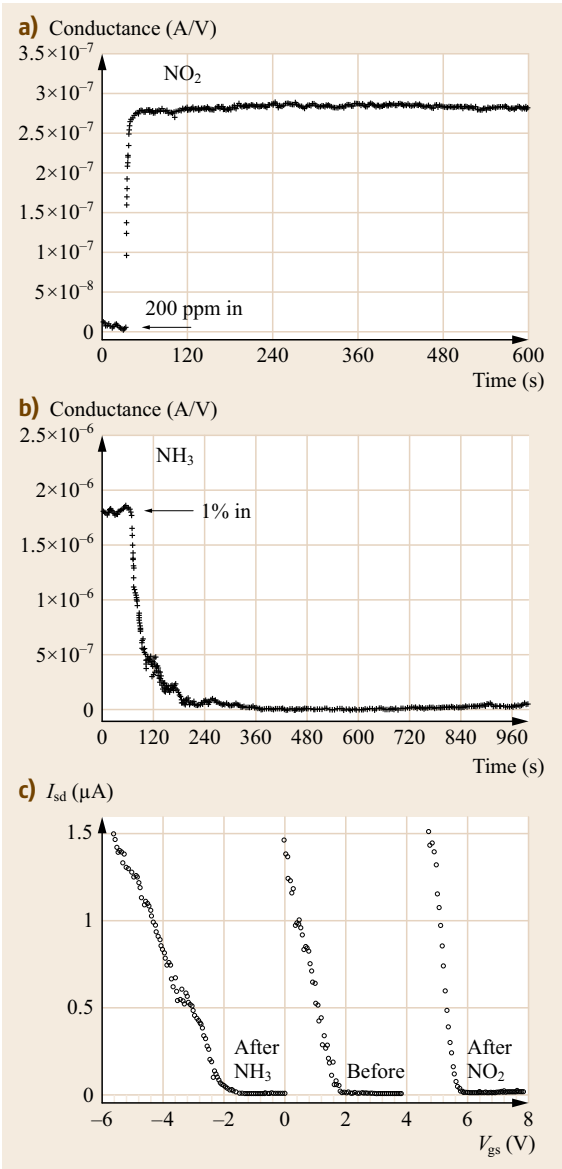


Fig. 14.13a–c SWNT FET chemical sensor measurements. **(a,b)** Time response of the same device upon exposure to 200 ppm NO_2 **(a)** and after recovery to 1% NH_3 **(b)**. **(c)** $I_{sd}(V_{gs})$ characteristics, before and after exposure to gas, showing a gate threshold shift (after [14.2])

et al. [14.114] showed that etching away the underlying SiO_2 from a CNFET results in chemically sensitive devices. Conversely, *Auvray* et al. [14.3] observed that SWNTs on aminosilane are 4–5 times more sensitive to triethylamine than SWNTs on SiO_2. These results suggest that, though the substrate may influence the sensitivity of the device, the substrate is not necessary for the sensor operation.

However, the debate concerning the role played by the metal contacts in gas sensing is still ongoing. In order to shed light on this issue, several groups have designed experiments in which only selected regions of the CNFET have been exposed to the molecules [14.115–121]. This was achieved via local passivation of the metal contacts, which leaves part of the nanotube surface exposed to the environment. There are mainly two opposing views on this issue. The first, most recently articulated in [14.121], holds that the dominant sensing mechanism is adsorption of the oxidizing molecules, NO_2 or NH_3, on the metal contacts leading to a change in the metal work function, whereas adsorption on the nanotube leads to smaller changes. Evidence is provided by gas measurements, Kelvin probe microscopy, and transistor modeling. The second view, summarized recently in [14.120], holds that in sensors with metal contacts passivated by gas-impermeable layers the sensor signal comes from adsorption of the molecules onto the nanotube itself. Evidence is provided by gas measurements, fully passivated control devices, and ab initio calculations [14.36]. Neither of the two views excludes the other. The debate concerns the dominant mechanism, but both admit the existence of the other mechanism. It is important to note that the interpretation of results may depend on factors such as device architecture, fabrication process, and even gas concentration range. Thus, one is advised to familiarize oneself with both views and assign the dominant mechanism based on the actual experimental conditions.

For CNFETs to work as biochemical sensors, some modifications are required. First, because of operation in aqueous solutions that contain ions, the back-gate is replaced by a reference electrode known as the liquid gate [14.122]. With this modification, CNFETs exhibit similar $I_{sd}(V_{lg})$ characteristics to devices in normal, gaseous environment ($V_{lg} \Leftrightarrow V_{gs}$). Second, to facilitate adsorption of biomolecules onto SWNTs, functionalization is often required [14.123]. The first biochemical CNFET sensor was demonstrated by *Besteman* et al. [14.7]. Glucose oxidase was attached to an SWNT, which subsequently showed response to both pH and concentration of glucose (0.1 M). In this study functionalization is an enabler; un-functionalized devices simply do not respond. More generally, functionalization is essential both in liquid and gas phase sensors for increasing sensitivity and selectivity [14.124]. For biosensors, the debate concerning the sensing mechanism follows the one for chemical sensors. Combining simulation and measurements, *Heller* et al. [14.125] found that both electrostatic gating and Schottky barriers are contributing, and that substrate also influences the response. The current understanding of the sens-

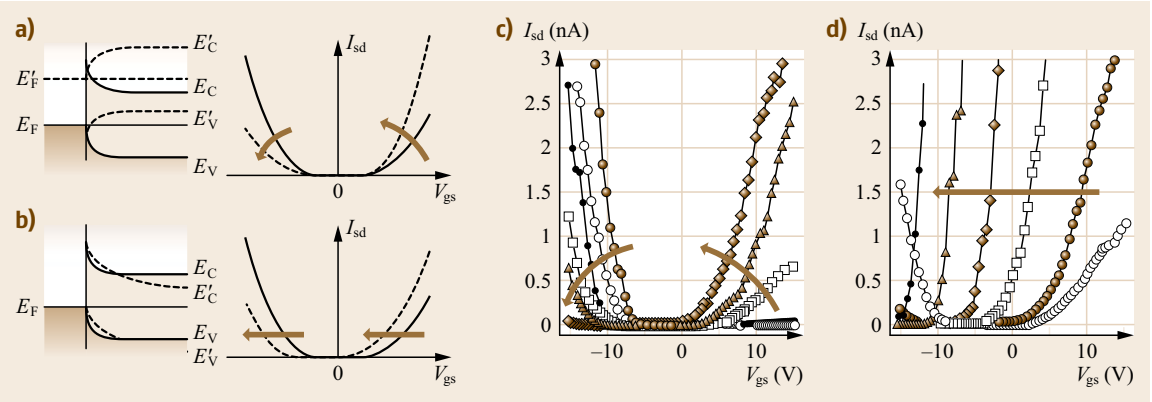


Fig. 14.14a–d Mechanisms encountered in CNT chemical sensors. **(a,b)** Band diagrams and corresponding $I_{sd}(V_{gs})$ characteristics, before (*solid line*) and after (*dashed line*) adsorption of molecules, for the case of adsorption at the contacts **(a)** and the case of nanotube doping **(b)**. In **(a)** the metal work function ϕ_m is modified, whereas in **(b)** the charge neutrality level changes in the CNT. **(c,d)** Measured $I_{sd}(V_{gs})$ characteristics subject to oxygen exposure **(c)**, assigned to mechanism **(a)**, and potassium doping **(d)**, assigned to mechanism **(b)**. *Arrows* in **(a–d)** indicate the change trend in $I_{sd}(V_{gs})$ characteristics (**(c,d)** after [14.45])

ing mechanisms as well as recent advancements in nanotube-based biosensors are reviewed in [14.126].

Concerning the sensing performance of chemical CNFET sensors, from the initial publication of *Kong* et al. [14.2], reporting concentration limits of detection of ≈ 2 ppm for NO_2 and $\approx 0.1\%$ for NH_3, there has been significant progress. Naturally, the limits of detection depend on the gas being sensed. *Chen* et al. [14.127] have measured concentrations as low as 40 ppt NO_2 and 200 ppt NH_3 with carbon nanotube mats under continuous UV illumination. With individual SWNT devices, the limit of detection was two orders of magnitude higher, but nonetheless, detecting 5 ppb of NO_2 is remarkable and well within the range that makes this technology interesting for environmental gas monitoring. Regarding CNFET biosensors, *Besteman* et al. [14.7] speculated that CNFET sensors have the potential to measure the enzymatic activity of even a single redox enzyme. Since then, single-molecule and single-charge resolution have been achieved experimentally by *Choi* et al. [14.128] and *Sharf* et al. [14.129], respectively.

Besides sensitivity, the other attractive performance metric of SWNT sensors is their energy efficiency. Biosensors are very often utilized for single measurements, diminishing the importance of recovery. But gas sensors are expected to work continuously and lots of energy is typically required to desorb attached molecules. *Chikkadi* et al. [14.119] managed to demonstrate that reproducible sensor recovery is possible with only 2.9 μW for 10 min (Fig. 14.15). Considering the power required to read out a nanotube gas sensor of around 32 μW [14.55], sensor systems operating with a few tens of μW are now possible based on this technolog1y, paving the way to tiny, ultralow power gas sensors for wireless sensor applications.

14.3.2 Piezoresistive Sensors

The electronic structure and thus the electrical conductance of all three classes of SWNTs (metallic, semiconducting, and small band gap semiconducting) depends greatly on the mechanical deformation of the SWNT structure as theoretically predicted in [14.26, 35] and briefly outlined in Sect. 14.1.1 for axial and torsional strain.

This theoretical expectation was first experimentally corroborated in pioneering work by *Tombler* et al. [14.8], for a suspended m-SWNT clamped to and contacted by two metal electrodes. Upon deforming and straining the nanotube by an AFM tip, the overall conductance of the device decreased reversibly by two orders of magnitude (strain $\varepsilon \approx 3\%$). The authors attributed this large conductance change to a mechanically induced modification of the atomic structure of the SWNT region right below the AFM tip, which could raise a tunnel barrier in the path of free carriers.

A suspended doubly clamped CNFET configuration to measure the piezoresistance (i. e., change of the band gap as a function of axial strain $dE_g/d\varepsilon$) was first proposed by *Minot* et al. [14.9]. The experimental setup was similar to that of Tombler. The CNFETs were strained by an AFM tip while the tip was simultaneously used as a local gate in the center region of the suspended SWNT (Fig. 14.16a). The overall CNFET conductance was tuned by the gate voltage at

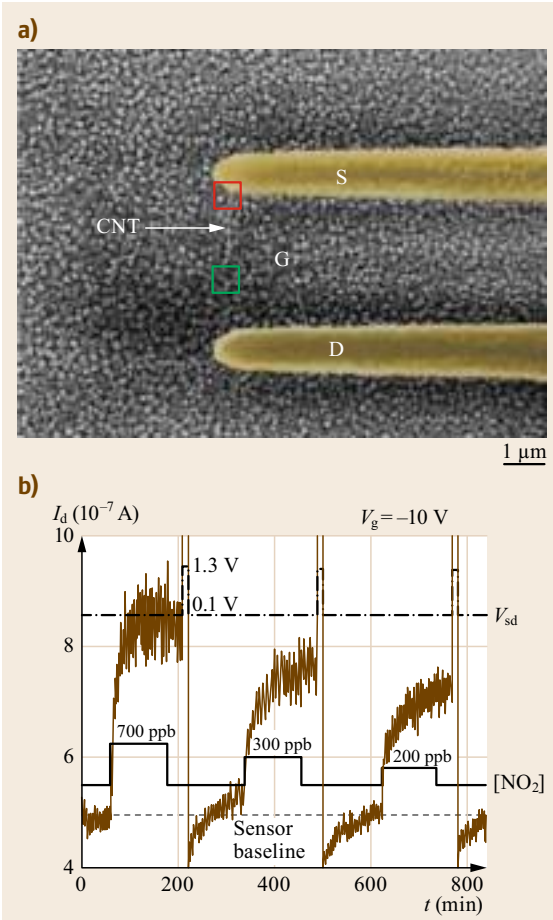


Fig. 14.15a,b Demonstration of an ultralow power NO_2 gas sensor based on SWNT FETs. **(a)** Top-view SEM image of a fabricated device consisting of suspended nanotube between source and drain contacts. **(b)** Sensor recovery by Joule self-heating. After each distinct, decreasing concentration a heating-pulse is applied and the sensor base-line is recovered. Reprinted from [14.119] with permission of AIP publishing

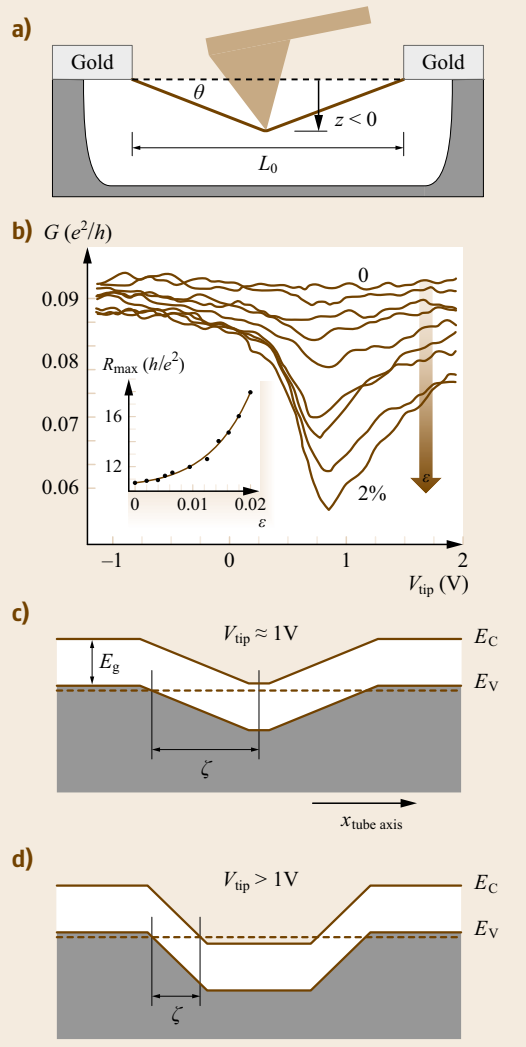


Fig. 14.16a–d Strain-dependent conductance measurement on a doubly clamped fully suspended CNFET. **(a)** Schematic of the experimental setup; an AFM tip strains the nanotube. **(b)** $G(V_{tip})$ measurements for an intrinsic metallic SWNT for 0 to 2% strain (*arrow* indicates increasing strain). *Inset* shows $R_{max}(\varepsilon)$ for $V_{tip} \approx 1$ V. **(c,d)** Band diagrams at the *center* of the SWNT at $V_{tip} \approx 1$ V **(c)** and $V_{tip} > 1$ V **(d)** displaying a barrier of width ζ below the AFM tip (after [14.9])

the tip (V_{tip}). Figure 14.16b shows a series of sweeps of V_{tip} on an intrinsic m-SWNT when no strain is applied. By gradually straining the suspended doubly clamped CNFET a small band gap opens ($dE_g/d\varepsilon > 0$), according to the theory of SGS-SWNTs ($p = 0$), and a conductance dip forms in the gate characteristic at $V_{tip} \approx 1$ V. In this configuration it is assumed that the tip gate is acting locally in the middle part of the SWNT, while the sections close to the contacts are p-type (this is because for $V_{tip} = 0$ V the CNFET is in the on-state hole conductance). The strained CNFET characteristic can be explained starting from this initial situation. For a small positive gate voltage ($0\,\text{V} \le V_{tip} < 1\,\text{V}$) a barrier for holes is created leading to a reduction of the total current (Fig. 14.16c). The hole current reaches its minimum (I_{min}) when the barrier formed by V_{tip} reaches E_g. Up to this point only thermionic emission is accounted for, since no tunneling is expected due to the width of the barrier formed by the tip gate (> 20 nm). For $V_{tip} > 1$ V the CNFET section of the SWNT reaches in-

version and transport increases due to tunneling through the thinned barriers ζ (Fig. 14.16d). Since the barrier height for holes at $I_{\min}$ is approximately equal to the band gap E_g of the CNFET, and since thermionic emission dominates, the overall change in the band gap of the CNFET due to strain can be extracted by employing the thermally activated conductance formula $G \approx \exp(-E_g(\varepsilon)/(K_B T))$. As a result an overall change of the band gap of $dE_g/d\varepsilon \approx 35\,\text{meV}/\%$ was extracted. This value is below the theoretical prediction of a maximum band gap change of $dE_{g\max}/d\varepsilon \approx 94\,\text{meV}/\%$ [14.35].

In contrast to doubly clamped suspended SWNTs the electromechanical properties of CNFETs adhering to thin film membranes were investigated in two studies [14.10, 11], confirming piezoresistivity also in this situation. *Grow* et al. [14.10] apply the gate voltage at the metallized backside of an SiN_x membrane. For SGS-SWNT FETs the $I_{sd}(V_{gs})$ characteristic shows that the unstrained device is p-type with no or only small Schottky barriers for the holes (small ON-state resistance). The applied strain increases the band gap ($dE_g/d\varepsilon > 0$). By increasing the back-gate voltage V_{gs} a small but wide hole barrier (zero tunneling) is formed along the length of the CNFET channel reducing the hole current, and reaching $I_{\min}$ when the barrier is close to E_g. The largest change in the channel conductance at various strain levels happens then at $I_{\min}$.

The sensitivity of the devices can be expressed as $|dE_g/d\varepsilon|$ in $\text{meV}/\%$ or alternatively by the gauge factor (GF) which is usually used for classical piezoresistors. The GF is defined by the relative change of resistance divided by the strain of the SWNT and it is only constant for small strains ($d\varepsilon$) around a working point ε_0. *Minot* et al. [14.9] reported a maximum sensitivity of $54\,\text{meV}/\%$ for the suspended CNFET while *Grow* et al. [14.10] extracted $180\,\text{meV}/\%$ for an SGS-SWNT adhering to a substrate. For a metallic SWNT a sensitivity of 340–430 $\text{meV}/\%$ was reported [14.11]. Such high sensitivities of CNFETs in contact with the substrate may be explained by additional mechanical distortions of the SWNTs due to interaction with the dielectric substrate. Some reported GF values include 856 for a membrane-based unstrained CNFET [14.10] and 2900 for a prestrained suspended metallic SWNT with $\varepsilon_0 \approx 4\%$ [14.12].

Research on the piezoresistivity of SWNTs has continued in the past few years in two directions. On the one hand, the integration of SWNTs into functional devices such as pressure sensors has been pursued [14.130, 131]. On the other hand, research into furthering understanding of the piezoresistive effect at the atomistic level has continued [14.132]. By exploiting the extraordinary gauge factors of SWNT, *Helbling* et al. [14.130] have demonstrated ultra-miniaturized pressure sensors with diaphragm diameters as small as $40\,\mu\text{m}$. *Burg* et al. [14.131] have extended this work, by using a process flow based on dielectrophoresis, to assemble the SWNT strain gauges at the high-strain locations on the membrane. At the fundamental level, *Muoth* et al. [14.132] have integrated individual, ultraclean SWNTs into MEMS tensile actuators. This approach significantly improved the control of the applied uniaxial strain in the nanotubes, while giving simultaneous access to the transport properties under strain and to the chiral indices by in situ TEM electron diffraction. Their investigations established that SWNTs withstand 4.5% strain repeatedly and confirmed the sign of the band gap modulation with strain according to the family index as predicted by the theory outlined in Sect. 14.1.1.

In general, the piezoresistive effect of SWNTs has been experimentally manifested in CNFETs by loading the tubes with axial strain. Though there are still quantitative discrepancies between theory and experiment regarding the band gap change rate with strain, this is one of the fortunate cases in which experiment exceeds theoretical expectations. From a practical point of view, piezoresistive CNFETs excel with regard to very high sensitivity already at small strain ($< 0.2\%$), nanoscale size, and ultralow power consumption (nW).

14.3.3 Resonant Sensors

Along with piezoresistive sensors, resonant sensors are useful in measuring mechanical stimuli such as strain and stress. In addition to that, resonators can be utilized to measure inertial properties (mainly mass), in which case they are known as inertial balances. The core principle, briefly explained in Sect. 14.1.1, exploits the dependence of the resonant frequency – typically the fundamental flexural mode – of an SWNT beam on quantities such as tension σ and mass density $\rho, \omega = \omega(\sigma, \rho)$.

For measuring mass, via $\omega(\rho)$, either a cantilever configuration or doubly clamped beam can be employed. Indeed, the cantilever configuration has been utilized right from the beginning to acquire CNT Young's moduli [14.133] or to demonstrate mass sensing. *Poncharal* et al. [14.14] were the first to show that with a MWNT cantilever, a mass as small as $22 \pm 6\,\text{fg}$ ($1\,\text{fg} = 10^{-15}\,\text{g}$) can be measured. *Jensen* et al. [14.15], claim to have detected only 51 Au atoms, corresponding to a mass of roughly 17 zg ($1\,\text{zg} = 10^{-21}\,\text{g}$), with a DWNT cantilever inside a field-emitting device.

Conversely, the doubly clamped beam configuration is in addition useful for measuring stress and strain, not possible with cantilevers. Moreover, for doubly clamped beams, biasing and measuring currents

through the nanotube is straightforward, which is a prerequisite for electronic readout. The first electromechanical resonator based on a doubly clamped carbon nanotube was proposed by *Sazonova* et al. [14.16]. Structurally, the device is a standard back-gate CNFET with an SWNT channel doubly clamped by top electrodes and partly or totally suspended over a trench (Fig. 14.17a and Fig. 14.11d [14.18]). The suspended region is the vibrating part of the nanotube.

The mechanical actuation of the nanotube is electrostatic. An applied gate voltage V_{gs} induces a capacitive Coulomb force F deflecting the suspended nanotube segment. V_{gs} contains a DC component to prestrain the nanotube ($\sigma \neq 0$), and a harmonic at frequency ω to excite the nanotube beam vibration. At frequencies close to the fundamental flexural mode ω_0, the nanotube behaves like a forced classical beam [14.134], oscillating with an amplitude δz that follows a classical Lorentzian centered at ω_0 (Fig. 14.2).

The readout of the nanotube displacement is done via the channel conductance. For a CNFET the conductance G depends both on the gate voltage V_{gs} and the distance of the nanotube to the gate z through the gate capacity C_{gs}. This is so because the conductance is proportional to the free-carrier concentration in the channel ($G \approx Q = C_{gs}V_{gs}$). At frequencies close to ω_0, the oscillation amplitude δz becomes important, resulting in a detectable conductance oscillation $\delta G(t)$. Theoretically $\delta G(t)$ is as easy to read as applying a source-drain voltage V_{sd} and measuring the current I_{sd}. However, because $f_0 = \omega_0/(2\pi) > 10\,\text{MHz}$, *Sazonova* et al. [14.16] have proposed using the CNFET as a mixer to downconvert $\delta G(t)$ from ω to $\Delta\omega$, conveniently chosen at around 10 kHz. In Fig. 14.17b, an intensity map of the current amplitude $\delta I_{sd}(V_{gs}, \omega)$ is shown [14.16]. As expected, ω increases with V_{gs}, since the static gate voltage changes the tension σ in the nanotube. Branches visible below the fundamental ω_0 in this map, are due to initial slack. Indeed, *Witkamp* et al. [14.18] have shown that for straight tubes (no slack) the modes below the fundamental are suppressed.

Regarding the sensing capabilities of CNT resonators, *Sazonova* et al. [14.16] have estimated the *force* resolution to be around $1\,\text{fN}/\sqrt{\text{Hz}}$ ($1\,\text{fN} = 10^{-15}\,\text{N}$), within a factor of ten of the best measured sensitivities at room temperature. At cryogenic temperatures of about 1.2 K, this resolution can be improved by 5 orders of magnitude, down to $12\,\text{zN}/\sqrt{\text{Hz}}$, as shown by *Moser* et al. in [14.135].

In [14.17], *Peng* et al. loaded a CNFET resonator with thermally evaporated Fe. Assuming a uniform coating of 2 nm of Fe ($\approx 3.5 \times 10^{-17}\,\text{g}$), they extrapolated a minimum detectable mass in the order of 1 ag ($10^{-18}\,\text{g}$). A series of improvements have followed ever since. *Lassagne* et al. [14.136] demonstrated a measurable mass limit of 25 zg at room temperature and of 1.4 zg at 5 K which corresponds roughly to the mass of only 15 Cr atoms. On the basis of measured noise power spectral densities at 20 mK, *Hüttel* et al. [14.137] estimated a mass resolution of $7\,\text{yg}/\sqrt{\text{Hz}}$ which would enable the detection of a single He atom. *Chaste* et al. [14.138] reported a mass resolution of only 1.7 yg, the mass of a single proton, at liquid helium temperatures.

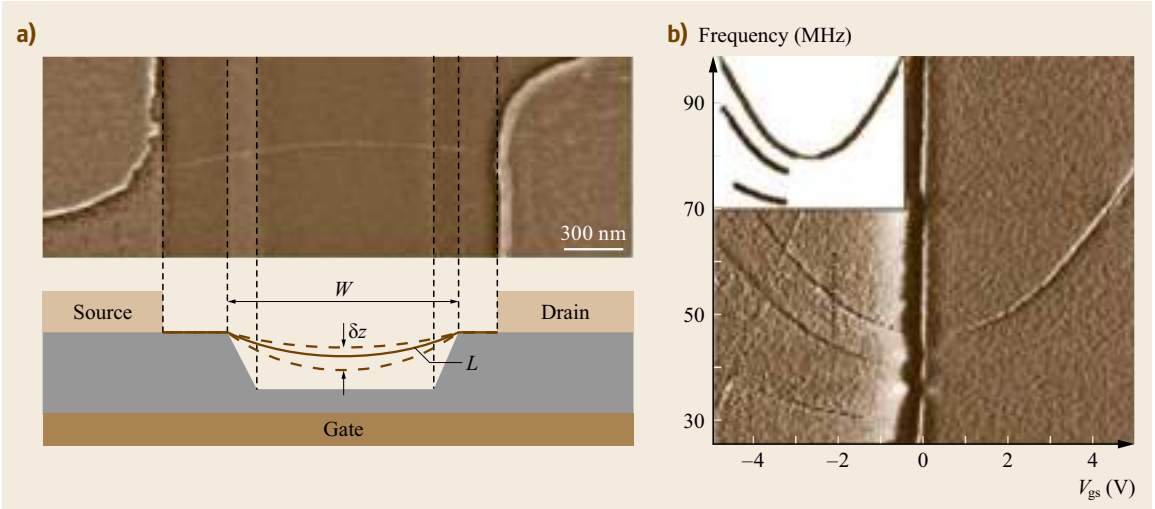


Fig. 14.17a,b Suspended SWNT resonator. **(a)** SEM image and cross-section schematic (*scale bar*, 300 nm). **(b)** Detected current as a function of gate voltage and frequency. *Inset* shows the extracted positions of the resonance frequency ω_0 (after [14.16])

Initially the limiting factor for both inertial balances and force sensors has been the quality factor Q, which in early reports was below 300 [14.16–18]. However, Q factors as high as 10^5 were demonstrated by *Hüttel* et al. [14.137], close to the intrinsic damping limit expected for SWNTs. Again, as emphasized by the authors, the key to obtaining such high quality factors, are suspended, defect-free SWNT that are not contaminated by resist residues. The measurements were performed in high-vacuum and at mK temperatures. The Q factor shows a $T^{-0.36}$ power law dependence on temperature, going from $\approx 10^5$ at 20 mK to less than 30 000 at 1 K. The authors did not continue the measurement up to room temperature, but *Lassagne* et al. [14.136] did find that the Q factor of one of their devices decreases from 865 at 5 K to 50 at 300 K. The adverse effect of pressure on Q factor has been known since the first paper on the subject, by *Sazonova* et al. [14.16], with resonance lost above pressures of 10 Torr. The scaling of the Q factor with temperature and pressure, especially the maximum achievable Q in ambient conditions is an open research question. On the practical side, devices based on carbon nanotube resonators with integrated axial MEMS actuators are well underway [14.139, 140]. Axial actuation – as opposed to electrostatic actuation – has already shown surprising capabilities, with 6 times [14.140] to 20 times [14.139] resonance frequency tuning achievable.

14.4 Concluding Remarks

Since CNTs are at the crossroad between fundamental science and engineering, transdisciplinary development in this field is highly demanded. Concerning device fabrication, synthetic chemists and process engineers still have challenges ahead of them, in trying to develop strategies for controlling nanotube electronic properties and local integration in functional systems. Even though we do not yet have chirality control, this doesn't mean that technology transfer on the midterm is not possible. In fact, the sensor concepts presented in Sect. 14.3 – with the exception of piezoresistive gauges – do not require chirality control, but band gap control (diameter and metallic versus semiconducting separation). With state-of-the-art methods for SWNT synthesis, separation and sorting, and CCVD, catalyst control and in situ growth, band gap control is almost here. Actually, one shouldn't rule out even piezoresistive gauges, since calibration, tuning and performance characterization, as usually done for sensors, may result in acceptable product yield. Still, as has become clearer in the past few years, besides controlled nanotube growth and integration, processes that do not contaminate the nanotubes after growth are of major importance and may see a surge of interest in the near future. On the application side, nanosystems technology may become the first user of SWNTs because in contrast to mainstream technologies (CMOS logic, memories) for which ultra large scale integration (ULSI) is the goal, sensors will require the integration of individual and *just a few* structures at the wafer level, only. An ideal platform for exploring new sensing devices is the CNFET sensor tool box as presented in this chapter.

References

14.1 R.H. Baughman, A.A. Zakhidov, W.A. de Heer: Carbon nanotubes – The route toward applications, Science **297**, 787–792 (2002)

14.2 J. Kong, N.R. Franklin, C. Zhou, M.G. Chapline, S. Peng, K. Cho, H. Dai: Nanotube molecular wires as chemical sensors, Science **287**, 622–625 (2000)

14.3 S. Auvray, V. Derycke, M. Goffman, A. Filoramo, O. Jost, J.-P. Bourgoin: Chemical optimization of self-assembled carbon nanotube transistors, Nano Lett. **5**, 451–455 (2005)

14.4 P.G. Collins, K. Bradley, M. Ishigami, A. Zettl: Extreme oxygen sensitivity of electronic properties of carbon nanotubes, Science **287**, 1801–1804 (2000)

14.5 J. Li, Y. Lu, Q. Ye, M. Cinke, J. Han, M. Meyyappan: Carbon nanotube sensors for gas and organic vapor detection, Nano Lett. **3**, 929–933 (2003)

14.6 Q.F. Pengfei, O. Vermesh, M. Grecu, A. Javey, Q. Wang, H. Dai, S. Peng, K.J. Cho: Toward large arrays of multiplex functionalized carbon nanotube sensors for highly sensitive and selective molecular detection, Nano Lett. **3**, 347–351 (2003)

14.7 K. Besteman, J.O. Lee, F.G.M. Wiertz, H.A. Heering, C. Dekker: Enzyme-coated carbon nanotubes as single-molecule biosensors, Nano Lett. **3**, 727–730 (2003)

14.8 T. Tombler, C. Zhou, L. Alexeyev, J. Kong, H. Dai, W. Liu, C.S. Jayanthi, M. Tang, S.Y. Wu: Reversible electromechanical characteristics of carbon nanotubes under local-probe manipulation, Nature **405**, 769–772 (2000)

14.9 E.D. Minot, Y. Yaish, V. Sazonova, J.Y. Park, M. Brink, P.L. McEuen: Tuning carbon nanotube band gaps with strain, Phys. Rev. Lett. **90**(156401), 1–4 (2003)

14.10 R.J. Grow, Q. Wang, J. Cao, D. Wang, H. Dai: Piezoresistance of carbon nanotubes on deformable thin-film membranes, Appl. Phys. Lett. **86**(093104), 1–3 (2005)

14.11 C. Stampfer, T. Helbling, D. Obergfell, B. Schoberle, M.K. Tripp, A. Jungen, S. Roth, V.M. Bright, C. Hierold: Fabrication of single-walled carbon-nanotube-based pressure sensors, Nano Lett. **6**, 233–237 (2006)

14.12 C. Stampfer, A. Jungen, R. Linderman, D. Obergfell, S. Roth, C. Hierold: Nano-electromechanical displacement sensing based on single-walled carbon nanotubes, Nano Lett. **6**, 1449–1453 (2006)

14.13 V.L. Pushparaj, L. Ci, S. Sreekala, A. Kumar, S. Kesapragada, D. Gall, O. Nalamasu, A.M. Pulickel, J. Suhr: Effects of compressive strains on electrical conductivities of a macroscale carbon nanotube block, Appl. Phys. Lett. **91**(153116), 3 (2007)

14.14 P. Poncharal, Z.L. Wang, D. Ugarte, W.A. de Heer: Electrostatic deflections and electromechanical resonances of carbon nanotubes, Science **283**, 1513–1516 (1999)

14.15 K. Jensen, K. Kim, A. Zettl: An atomic-resolution nanomechanical mass sensor, Nat. Nanotechnol. **3**, 533–537 (2008)

14.16 V. Sazonova, Y. Yaish, H. Üstünel, D. Roundy, T.A. Arias, P.L. McEuen: A tunable carbon nanotube electromechanical oscillator, Nature **431**, 284–287 (2004)

14.17 H.B. Peng, C.W. Chang, S. Aloni, T.D. Yuzvinsky, A. Zettl: Ultrahigh frequency nanotube resonators, Phys. Rev. Lett. **97**(087203), 1–4 (2006)

14.18 B. Witkamp, M. Poot, H.S.J. van der Zant: Bending-mode vibration of a suspended nanotube resonator, Nano Lett. **6**, 2904–2908 (2006)

14.19 M. Freitag, Y. Martin, J.A. Misewich, R. Martel, P. Avouris: Photoconductivity of single carbon nanotubes, Nano Lett. **3**, 1067–1071 (2003)

14.20 X. Qiu, M. Freitag, V. Perebeinos, P. Avouris: Photoconductivity spectra of single-carbon nanotubes: Implications on the nature of their excited states, Nano Lett. **5**, 749–752 (2005)

14.21 M.S. Dresselhaus, G. Dresselhaus, P. Avouris: *Carbon Nanotubes: Synthesis, Structure, Properties and Applications* (Springer, Berlin 2001)

14.22 R. Saito, G. Dresselhaus, M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College, London 2001)

14.23 A. Jorio, G. Dresselhaus, M.S. Dresselhaus: *Carbon Nanotubes: Advanced Topics in the Synthesis, Structure, Properties and Applications* (Springer, Berlin, Heidelberg 2008)

14.24 S. Li, Z. Yu, C. Rutherglen, P.J. Burke: Electrical properties of 0.4 cm long single-walled carbon nanotubes, Nano Lett. **4**, 2003–2007 (2004)

14.25 C.T. White, J.W. Mintmire: Density of states reflects diameter in nanotubes, Nature **394**, 29–30 (1998)

14.26 A. Kleiner, S. Eggert: Curvature, hybridization, and STM images of carbon nanotubes, Phys. Rev. B **64**(113402), 1–4 (2001)

14.27 B. Bourlon, J. Wong, C. Mikó, L. Forró, M. Bockrath: A nanoscale probe for fluidic and ionic transport, Nat. Nanotechnol. **2**, 104–107 (2006)

14.28 A. Modi, N. Koratkar, E. Lass, B. Wei, P.M. Ajayan: Miniaturized gas ionization sensors using carbon nanotubes, Nature **424**, 171–174 (2003)

14.29 E.S. Snow, F.K. Perkins, E.J. Houser, S.C. Badescu, T.L. Reinecke: Chemical detection with a single-walled carbon nanotube capacitor, Science **307**, 1942–1945 (2005)

14.30 A. Krishnan, E. Dujardin, T.W. Ebbesen, P.N. Yianilos, M.M.J. Treacy: Young's modulus of single-walled nanotubes, Phys. Rev. B **58**, 14013 (1998)

14.31 D.A. Walters, L.M. Ericson, M.J. Casavant, J. Liu, D.T. Colbert, K.A. Smith, R.E. Smalley: Elastic strain of freely suspended single-wall carbon nanotube ropes, Appl. Phys. Lett. **74**, 3803–3805 (1999)

14.32 M.-F. Yu, B.S. Files, S. Arepalli, R.S. Ruoff: Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties, Phys. Rev. Lett. **84**, 5552–5555 (2000)

14.33 H.W.C. Postma, I. Kozinsky, A. Husain, M.L. Roukes: Dynamic range of nanotube- and nanowire-based electromechanical systems, Appl. Phys. Lett. **86**(223105), 1–3 (2005)

14.34 L. Yang, M.P. Anantram, J. Han, J.P. Lu: Band-gap change of carbon nanotubes: Effect of small uniaxial and torsional strain, Phys. Rev. B **60**, 13874–13878 (1999)

14.35 L. Yang, J. Han: Electronic structure of deformed carbon nanotubes, Phys. Rev. Lett. **85**, 154–157 (2000)

14.36 J.M. Kroes, F. Pietrucci, K. Chikkadi, C. Roman, C. Hierold, W. Andreoni: The response of single-walled carbon nanotubes to NO_2 and the search for a long-living adsorbed species, Appl. Phys. Lett. **108**(033111), 1–3 (2016)

14.37 P. Ruffieux, O. Gröning, M. Bielmann, P. Mauron, L. Schlapbach, P. Gröning: Hydrogen adsorption on sp^2-bonded carbon: Influence of the local curvature, Phys. Rev. B **66**(245416), 1–8 (2002)

14.38 S. Picozzi, S. Santucci, L. Lozzi, L. Valentini, B. Delley: Ozone adsorption on carbon nanotubes: The role of Stone-Wales defects, J. Chem. Phys. **120**, 7147–7152 (2004)

14.39 D.L. Carroll, P. Redlich, X. Blase, J.-C. Charlier, S. Curran, P.M. Ajayan, S. Roth, M. Rühle: Effects of nanodomain formation on the electronic structure of doped carbon nanotubes, Phys. Rev. Lett. **81**, 2332–2335 (1998)

14.40 R.J. Chen, N.R. Franklin, J. Kong, J. Cao, T. Tombler, Y. Zhang, H. Dai: Molecular photodesorption from single-walled carbon nanotubes, Appl. Phys. Lett. **79**, 2258–2260 (2001)

14.41 S. Datta: *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge 1997)

14.42 J. Kong, E. Yenilmez, T.W. Tombler, W. Kim, H. Dai, R.B. Laughlin, L. Liu, C.S. Jayanthi, S.Y. Wu: Quantum interference and ballistic transmission in

nanotube electron waveguides, Phys. Rev. Lett. **87**(106801), 1–4 (2001)
14.43 W. Liang, M. Bockrath, D. Bozovic, J.H. Hafner, M. Tinkham, H. Park: Fabry-Perot interference in a nanotube electron waveguide, Nature **411**, 665–669 (2001)
14.44 J.-C. Charlier, X. Blase, S. Roche: Electronic and transport properties of nanotubes, Rev. Mod. Phys. **79**, 677–732 (2007)
14.45 S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, P. Avouris: Carbon nanotubes as Schottky barrier transistors, Phys. Rev. Lett. **89**(106801), 1–4 (2002)
14.46 F. Leonard, J. Tersoff: Role of Fermi-level pinning in nanotube Schottky diodes, Phys. Rev. Lett. **84**, 4693–4696 (2000)
14.47 Z.H. Chen, J. Appenzeller, J. Knoch, Y.M. Lin, P. Avouris: The role of metal-nanotube contact in the performance of carbon nanotube field-effect transistors, Nano Lett. **5**, 1497–1502 (2005)
14.48 A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai: Ballistic carbon nanotube field-effect transistors, Nature **424**, 654–657 (2003)
14.49 J. Appenzeller, J. Knoch, V. Derycke, R. Martel, S. Wind, P. Avouris: Field-modulated carrier transport in carbon nanotube transistors, Phys. Rev. Lett. **89**(126801), 1–4 (2002)
14.50 J. Appenzeller, M. Radosavljević, J. Knoch, P. Avouris: Tunneling versus thermionic emission in one-dimensional semiconductors, Phys. Rev. Lett. **92**(048301), 1–4 (2004)
14.51 D. Mann, A. Javey, J. Kong, Q. Wang, H. Dai: Ballistic transport in metallic nanotubes with reliable Pd ohmic contacts, Nano Lett. **3**, 1541–1544 (2003)
14.52 M.P. Anantram, F. Léonard: Physics of carbon nanotube electronic devices, Rep. Prog. Phys. **69**, 507–561 (2006)
14.53 M.J. Biercuk, S. Ilani, C.M. Marcus, P.L. McEuen: Electrical transport in single-wall carbon nanotubes. In: *Carbon Nanotubes*, ed. by A. Jorio, G. Dresselhaus, M.S. Dresselhaus (Springer, Berlin, Heidelberg 2008) pp. 455–493
14.54 S. Roche, E. Akkermans, O. Chauvet, F. Hekking, J.-P. Issi, R. Martel, G. Montambaux, P. Poncharal: Transport properties. In: *Understanding Carbon Nanotubes*, ed. by A. Loiseau, P. Launois, P. Petit, S. Roche, J.-P. Salvetat (Springer, Berlin, Heidelberg 2006) pp. 335–437
14.55 T.S. Cho, K.-J. Lee, J. Kong, A.P. Chandrakasan: The design of a low power carbon nanotube chemical sensor system. In: *Des. Autom. Conf. 2008 (DAC 2008), Anaheim* (2008), pp. 8–13, 84–89
14.56 P.G. Collins, M.S. Fuhrer, A. Zettl: $1/f$ Noise in carbon nanotubes, Appl. Phys. Lett. **76**, 894–896 (2000)
14.57 J. Appenzeller, Y.M. Lin, J. Knoch, Z.H. Chen, P. Avouris: $1/f$ Noise in carbon nanotube devices – On the impact of contacts and device geometry, IEEE Trans. Nanotechnol. **6**, 368–373 (2007)
14.58 J. Tersoff: Low-frequency noise in nanoscale ballistic transistors, Nano Lett. **7**, 194–198 (2007)
14.59 J. Männik, I. Heller, A.M. Janssens, S.G. Lemay, C. Dekker: Charge noise in liquid-gated single-wall carbon nanotube transistors, Nano Lett. **8**, 685–688 (2007)
14.60 P. Chaturvedi, P. Verma, A. Singh, P. Chaudhary, P. Basu: Carbon nanotube – Purification and sorting protocols, Def. Sci. J. **58**, 591–599 (2008)
14.61 N. Komatsu, F. Wang: A comprehensive review on separation methods and techniques for single-walled carbon nanotubes, Materials **3**, 3818–3844 (2010)
14.62 V. Jourdain, C. Bichara: Current understanding of the growth of carbon nanotubes in catalytic chemical vapour deposition, Carbon **58**, 2–39 (2013)
14.63 K.A. Shah, B.A. Tali: Synthesis of carbon nanotubes by catalytic chemical vapour deposition: A review on carbon sources, catalysts and substrates, Mater. Sci. Semicond. Process. **41**, 67–82 (2016)
14.64 C.L. Cheung, A. Kurtz, H. Park, C.M. Lieber: Diameter-controlled synthesis of carbon nanotubes, J. Phys. Chem. B **106**, 2429–2433 (2002)
14.65 Y. Tu, Z.P. Huang, D.Z. Wang, J.G. Wen, Z.F. Ren: Growth of aligned carbon nanotubes with controlled site density, Appl. Phys. Lett. **80**, 4018–4020 (2002)
14.66 M.-F. Fiawoo, A.-M. Bonnot, H. Amara, C. Bichara, J. Thibault-Pénisson, A. Loiseau: Evidence of correlation between catalyst particles and the single-wall carbon nanotube diameter: A first step towards chirality control, Phys. Rev. Lett. **108**(195503), 1–5 (2012)
14.67 Y. Li, W. Kim, Y. Zhang, M. Rolandi, D. Wang, H. Dai: Growth of single-walled carbon nanotubes from discrete catalytic nanoparticles of various sizes, J. Phys. Chem. B **105**, 11424–11431 (2001)
14.68 Y. Yao, C. Feng, J. Zhang, Z. Liu: "Cloning" of single-walled carbon nanotubes via open-end growth mechanism, Nano Lett. **9**, 1673–1677 (2009)
14.69 J. Liu, C. Wang, X. Tu, B. Liu, L. Chen, M. Zheng, C. Zhou: Chirality-controlled synthesis of single-wall carbon nanotubes using vapour-phase epitaxy, Nat. Commun. **3**(1199), 1–7 (2012)
14.70 J.R. Sanchez-Valencia, T. Dienel, O. Gröning, I. Shorubalko, A. Mueller, M. Jansen, K. Amsharov, P. Ruffieux, R. Fasel: Controlled synthesis of single-chirality carbon nanotubes, Nature **512**, 61–64 (2014)
14.71 A. Mueller, K.Y. Amsharov: Synthesis of robust precursors for the controlled fabrication of (6,6), (8,8), (10,10) and (12,12) armchair single-walled carbon nanotubes, Eur. J. Organic Chem. **2015**, 3053–3056 (2015)
14.72 F. Yang, X. Wang, D. Zhang, J. Yang, D. Luo, Z. Xu, J. Wei, J.-Q. Wang, Z. Xu, F. Peng: Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts, Nature **510**, 522–524 (2014)
14.73 F. Yang, X. Wang, D. Zhang, K. Qi, J. Yang, Z. Xu, M. Li, X. Zhao, X. Bai, Y. Li: Growing zigzag (16,0) carbon nanotubes with structure-defined cata-

lysts, J. Am. Chem. Soc. **137**, 8688–8691 (2015)
14.74 R.M. Tromp, A. Afzali, M. Freitag, D.B. Mitzi, Z. Chen: Novel strategy for diameter-selective separation and functionalization of single-walled carbon nanotubes, Nano Lett. **8**, 469–472 (2008)
14.75 M.S. Arnold, S.I. Stupp, M.C. Hersam: Enrichment of single-walled carbon nanotubes by diameter in density gradients, Nano Lett. **5**, 713–718 (2005)
14.76 M.S. Arnold, A.A. Green, J.F. Hulvat, S.I. Stupp, M.C. Hersam: Sorting carbon nanotubes by electronic structure using density differentiation, Nat. Nanotechnol. **1**, 60–65 (2006)
14.77 X. Tu, S. Manohar, A. Jagota, M. Zheng: DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes, Nature **460**, 250–253 (2009)
14.78 R.E. Smalley, Y. Li, V.C. Moore, B.K. Price, R. Colorado, H.K. Schmidt, R.H. Hauge, A.R. Barron, J.M. Tour: Single wall carbon nanotube amplification: En route to a type-specific growth mechanism, J. Am. Chem. Soc. **128**, 15824–15829 (2006)
14.79 Y. Yao, C. Feng, J. Zhang, Z. Liu: "Cloning" of single-walled carbon nanotubes via open-end growth mechanism, Nano Lett. **9**(4), 1673–1677 (2009) doi:10.1021/nl900207v
14.80 J. Liu, C. Wang, X. Tu, B. Liu, L. Chen, M. Zheng, C. Zhou: Chirality-controlled synthesis of single-wall carbon nanotubes using vapour-phase epitaxy, Nat. Commun. **3**, 1199 (2012) doi:10.1038/ncomms2205
14.81 E. Joselevich, H. Dai, J. Liu, K. Hata, A.H. Windle: Carbon nanotube synthesis and organization. In: *Carbon Nanotubes*, ed. by A. Jorio, G. Dresselhaus, M.S. Dresselhaus (Springer, Berlin, Heidelberg 2008) pp. 101–164
14.82 Y. Yan, M.B. Chan-Park, Q. Zhang: Advances in carbon-nanotube assembly, Small **3**, 24–42 (2007)
14.83 S.J. Tans, A.R.M. Verschueren, C. Dekker: Room-temperature transistor based on a single carbon nanotube, Nature **393**, 49–52 (1998)
14.84 R. Martel, T. Schmidt, H.R. Shea, T. Hertel, P. Avouris: Single- and multi-wall carbon nanotube field-effect transistors, Appl. Phys. Lett. **73**, 2447–2449 (1998)
14.85 V. Derycke, R. Martel, J. Appenzeller, P. Avouris: Carbon nanotube inter- and intramolecular logic gates, Nano Lett. **1**, 453–456 (2001)
14.86 J. Liu, M.J. Casavant, M. Cox, D.A. Walters, P. Boul, W. Lu, A.J. Rimberg, K.A. Smith, D.T. Colbert, R.E. Smalley: Controlled deposition of individual single-walled carbon nanotubes on chemically functionalized templates, Chem. Phys. Lett. **303**, 125–129 (1999)
14.87 C. Klinke, J.B. Hannon, A. Afzali, P. Avouris: Field-effect transistors assembled from functionalized carbon nanotubes, Nano Lett. **6**, 906–910 (2006)
14.88 S.G. Rao, L. Huang, W. Setyawan, S. Hong: Nanotube electronics: Large-scale assembly of carbon nanotubes, Nature **425**, 36–37 (2003)
14.89 H. Park, A. Afzali, S.-J. Han, G.S. Tulevski, A.D. Franklin, J. Tersoff, J.B. Hannon, W. Haensch: High-density integration of carbon nanotubes via chemical self-assembly, Nat. Nanotechnol. **7**, 787–791 (2012)
14.90 R. Krupke, F. Hennrich, H.B. Weber, M.M. Kappes, H. v. Löhneysen: Simultaneous deposition of metallic bundles of single-walled carbon nanotubes using AC-dielectrophoresis, Nano Lett. **3**, 1019–1023 (2003)
14.91 R. Krupke, F. Hennrich, H. von Löhneysen, M.M. Kappes: Separation of metallic from semiconducting single-walled carbon nanotubes, Science **301**, 344–347 (2003)
14.92 B.R. Burg, D. Poulikakos: Large-scale integration of single-walled carbon nanotubes and graphene into sensors and devices using dielectrophoresis: A review, J. Mater. Res. **26**, 1561–1571 (2011)
14.93 J. Kong, H.T. Soh, A.M. Cassell, C. Quate, H. Dai: Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers, Nature **395**, 878–881 (1998)
14.94 H.T. Soh, A. Morpurgo, J. Kong, C. Marcus, C. Quate, H. Dai: Integrated nanotube circuits: Controlled growth and ohmic contacting of single-walled carbon nanotubes, Appl. Phys. Lett. **75**, 627–629 (1999)
14.95 A. Javey, H. Dai: Regular arrays of 2 nm metal nanoparticles for deterministic synthesis of nanomaterials, J. Am. Chem. Soc. **127**, 11942–11943 (2005)
14.96 Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, H. Dai: Electric-field-directed growth of aligned single-walled carbon nanotubes, Appl. Phys. Lett. **79**, 3155–3157 (2001)
14.97 E. Joselevich, C.M. Lieber: Vectorial growth of metallic and semiconducting single-wall carbon nanotubes, Nano Lett. **2**, 1137–1141 (2002)
14.98 S. Dittmer, J. Svensson, E.E.B. Campbell: Electric field aligned growth of single-walled carbon nanotubes, Curr. Appl. Phys. **4**, 595–598 (2004)
14.99 S. Huang, X. Cai, J. Liu: Growth of millimeter-long and horizontally aligned single-walled carbon nanotubes on flat substrates, J. Am. Chem. Soc. **125**, 5636–5637 (2003)
14.100 S. Han, X. Liu, C. Zhou: Template-free directional growth of single-walled carbon nanotubes on a- and r-plane sapphire, J. Am. Chem. Soc. **127**, 5294–5295 (2005)
14.101 A. Ismach, L. Segev, E. Wachtel, E. Joselevich: Atomic-step-templated formation of single wall carbon nanotube patterns, Angew. Chem. Int. Ed. **43**, 6140–6143 (2004)
14.102 X. Liu, S. Han, C. Zhou: Novel nanotube-on-insulator (NOI) approach toward single-walled carbon nanotube devices, Nano Lett. **6**, 34–39 (2005)
14.103 B. Liu, C. Wang, J. Liu, Y. Che, C. Zhou: Aligned carbon nanotubes: From controlled synthesis to electronic applications, Nanoscale **5**, 9483–9502 (2013)
14.104 A.M. Cassell, N.R. Franklin, T. Tombler, E.M. Chan, J. Han, H. Dai: Directed growth of free-standing single-walled carbon nanotubes, J. Am. Chem.

Soc. **121**, 7975–7976 (1999)
14.105 N.R. Franklin, H. Dai: An enhanced CVD approach to extensive nanotube networks with directionality, Adv. Mater. **12**, 890–894 (2000)
14.106 N.R. Franklin, Q. Wang, T.W. Tombler, A. Javey, M. Shim, H. Dai: Integration of suspended carbon nanotube arrays into electronic devices and electromechanical systems, Appl. Phys. Lett. **81**, 913–915 (2002)
14.107 A. Jungen, S. Hofmann, J.C. Meyer, C. Stampfer, S. Roth, J. Robertson, C. Hierold: Synthesis of individual single-walled carbon nanotube bridges controlled by support micromachining, J. Micromech. Microeng. **17**, 603–608 (2007)
14.108 M. Muoth, T. Helbling, L. Durrer, S.-W. Lee, C. Roman, C. Hierold: Hysteresis-free operation of suspended carbon nanotube transistors, Nat. Nanotechnol. **5**, 589–592 (2010)
14.109 S.-W. Lee, M. Muoth, T. Helbling, M. Mattmann, C. Hierold: Suppression of resist contamination during photolithography on carbon nanomaterials by a sacrificial layer, Carbon **66**, 295–301 (2014)
14.110 W. Liu, K. Chikkadi, S.-W. Lee, C. Hierold, M. Haluska: Improving non-suspended carbon nanotube FET performance by using an alumina protective layer, Sens. Actuators B **198**, 479–486 (2014)
14.111 T. Zhang, S. Mubeen, N.V. Myung, M.A. Deshusses: Recent progress in carbon nanotube-based gas sensors, Nanotechnology **19**, 332001 (2008)
14.112 S.N. Kim, J.F. Rusling, F. Papadimitrakopoulos: Carbon nanotubes for electronic and electrochemical detection of biomolecules, Adv. Mater. **19**, 3214–3228 (2007)
14.113 N. Sinha, J. Ma, J.T.W. Yeow: Carbon nanotube-based sensors, J. Nanosci. Nanotechnol. **6**, 573–590 (2006)
14.114 T. Helbling, R. Pohle, L. Durrer, C. Stampfer, C. Roman, A. Jungen, M. Fleischer, C. Hierold: Sensing NO_2 with individual suspended single-walled carbon nanotubes, Sens. Actuators B **132**, 491–497 (2008)
14.115 K. Bradley, J.C.P. Gabriel, A. Star, G. Gruner: Short-channel effects in contact-passivated nanotube chemical sensors, Appl. Phys. Lett. **83**, 3821–3823 (2003)
14.116 J. Zhang, A. Boyd, A. Tselev, M. Paranjape, P. Barbara: Mechanism of NO_2 detection in carbon nanotube field effect transistor chemical sensors, Appl. Phys. Lett. **88**(123112), 1–3 (2006)
14.117 X.L. Liu, Z.C. Luo, S. Han, T. Tang, D.H. Zhang, C.W. Zhou: Band engineering of carbon nanotube field-effect transistors via selected area chemical gating, Appl. Phys. Lett. **86**(243501), 1–3 (2005)
14.118 M. Mattmann, T. Helbling, L. Durrer, C. Roman, C. Hierold, R. Pohle, M. Fleischer: Sub-ppm NO_2 detection by Al_2O_3 contact passivated carbon nanotube field effect transistors, Appl. Phys. Lett. **94**(3502), 1–3 (2009)
14.119 K. Chikkadi, M. Muoth, V. Maiwald, C. Roman, C. Hierold: Ultra-low power operation of self-heated, suspended carbon nanotube gas sensors, Appl. Phys. Lett. **103**(223109), 1–3 (2013)
14.120 K. Chikkadi, M. Muoth, C. Roman, M. Haluska, C. Hierold: Advances in NO_2 sensing with individual single-walled carbon nanotube transistors, Beilstein J. Nanotechnol. **5**, 2179–2191 (2014)
14.121 I. Dube, D. Jiménez, G. Fedorov, A. Boyd, I. Gayduchenko, M. Paranjape, P. Barbara: Understanding the electrical response and sensing mechanism of carbon-nanotube-based gas sensors, Carbon **87**, 330–337 (2015)
14.122 M. Kruger, M.R. Buitelaar, T. Nussbaumer, C. Schonenberger, L. Forro: Electrochemical carbon nanotube field-effect transistor, Appl. Phys. Lett. **78**, 1291–1293 (2001)
14.123 R.J. Chen, Y. Zhang, D. Wang, H. Dai: Noncovalent sidewall functionalization of single-walled carbon nanotubes for protein immobilization, J. Am. Chem. Soc. **123**, 3838–3839 (2001)
14.124 K. Balasubramanian, M. Burghard: Chemically functionalized carbon nanotubes, Small **1**, 180–192 (2005)
14.125 I. Heller, A.M. Janssens, J. Mannik, E.D. Minot, S.G. Lemay, C. Dekker: Identifying the mechanism of biosensing with carbon nanotube transistors, Nano Lett. **8**, 591–595 (2008)
14.126 T.-T. Tran, A. Mulchandani: Carbon nanotubes and graphene nano field-effect transistor-based biosensors, Trends Anal. Chem. **79**, 222–232 (2015)
14.127 G. Chen, T.M. Paronyan, E.M. Pigos, A.R. Harutyunyan: Enhanced gas sensing in pristine carbon nanotubes under continuous ultraviolet light illumination, Sci. Rep. **2**(343), 1–7 (2012)
14.128 Y. Choi, I.S. Moody, P.C. Sims, S.R. Hunt, B.L. Corso, I. Perez, G.A. Weiss, P.G. Collins: Single-molecule lysozyme dynamics monitored by an electronic circuit, Science **335**, 319–324 (2012)
14.129 T. Sharf, N.-P. Wang, J.W. Kevek, M.A. Brown, H. Wilson, S. Heinze, E.D. Minot: Single electron charge sensitivity of liquid-gated carbon nanotube transistors, Nano Lett. **14**, 4925–4930 (2014)
14.130 T. Helbling, C. Roman, L. Durrer, C. Stampfer, C. Hierold: Gauge factor tuning, long-term stability, and miniaturization of nanoelectromechanical carbon-nanotube sensors, IEEE Trans. Electron Devices **58**, 4053–4060 (2011)
14.131 B.R. Burg, T. Helbling, C. Hierold, D. Poulikakos: Piezoresistive pressure sensors with parallel integration of individual single-walled carbon nanotubes, J. Appl. Phys. **109**, 064310 (2011)
14.132 M. Muoth, K. Chikkadi, Y. Liu, C. Hierold: Suspended CNT-FET piezoresistive strain gauges: Chirality assignment and quantitative analysis. In: *IEEE 26th Int. Conf. Micro Electro Mech. Syst. (MEMS)* (2013) pp. 496–499
14.133 M.M.J. Treacy, T.W. Ebbesen, J.M. Gibson: Exceptionally high Young's modulus observed for individual carbon nanotubes, Nature **381**, 678–680 (1996)
14.134 D. Garcia-Sanchez, A. San Paulo, M.J. Esplandiu, F. Perez-Murano, L. Forró, A. Aguasca, A. Bachtold: Mechanical detection of carbon nanotube resonator vibrations, Phys. Rev. Lett. **99**(085501),

1–4 (2007)
14.135 J. Moser, J. Güttinger, A. Eichler, M.J. Esplandiu, D. Liu, M. Dykman, A. Bachtold: Ultrasensitive force detection with a nanotube mechanical resonator, Nat. Nanotechnol. **8**, 493–496 (2013)
14.136 B. Lassagne, D. Garcia-Sanchez, A. Aguasca, A. Bachtold: Ultrasensitive mass sensing with a nanotube electromechanical resonator, Nano Lett. **8**, 3735–3738 (2008)
14.137 A.K. Hüttel, G.A. Steele, B. Witkamp, M. Poot, L.P. Kouwenhoven, H.S. van der Zant: Carbon nanotubes as ultrahigh quality factor mechanical resonators, Nano Lett. **9**, 2547–2552 (2009)
14.138 J. Chaste, A. Eichler, J. Moser, G. Ceballos, R. Rurali, A. Bachtold: A nanomechanical mass sensor with yoctogram resolution, Nat. Nanotechnol. **7**, 301–304 (2012)
14.139 Z. Ning, T. Shi, M. Fu, Y. Guo, X. Wei, S. Gao, Q. Chen: Transversally and axially tunable carbon nanotube resonators in situ fabricated and studied inside a scanning electron microscope, Nano Lett. **14**, 1221–1227 (2014)
14.140 S. Truax, S.-W. Lee, M. Muoth, C. Hierold: Axially tunable carbon nanotube resonators using co-integrated microactuators, Nano Lett. **14**, 6092–6096 (2014)

15. Nanomechanical Cantilever Array Sensors

Hans Peter Lang, Martin Hegner, Christoph Gerber

In recent years, mechanics has experienced a revival, as microfabrication technologies and nanotechnology are applied to produce tiny structures. The development of ultraprecise position sensing started three decades ago with a novel imaging technique called atomic force microscopy, which provides ultrahigh topography resolution on the atomic scale by raster-scanning the surface with a microfabricated cantilever beam that has a tiny tip at its free end. The high force sensitivity can not only be used for imaging, but also allows the measurement of surface forces during molecule adsorption processes on the cantilever surface, thus enabling cantilevers to act as chemical sensors. Because of their small size, cantilevers allow fast and reliable detection of small concentrations of molecules in air and solution. In addition to artificial nose and label-free biosensing applications, they have also been employed to measure physical properties of tiny amounts of materials in miniaturized versions of conventional standard techniques such as calorimetry, thermogravimetry, weighing, photothermal spectroscopy and monitoring of chemical reactions. In the past few years, the cantilever-sensor concept has been extended to medical applications and has entered clinics for pilot studies on patients. The small size and scalability of cantilever array sensors might turn out to be advantageous for diagnostic screening applications and disease monitoring, as well as for genomics or proteomics. Using microcantilever arrays allows simultaneous detection of several analytes and solves the inherent problem of thermal drifts often present when using single microcantilever sensors, as some of the cantilevers can be used as sensor cantilevers for detection, and others as passivated reference cantilevers that do not show affinity to the molecules to be detected.

15.1 **Technique** 458
15.1.1 Cantilevers 458
15.1.2 History of Cantilever Sensors 458
15.2 **Cantilever Array Sensors** 459
15.2.1 Concept 459
15.2.2 Compressive and Tensile Stress 459
15.2.3 Disadvantages of Single Microcantilevers 459
15.2.4 Reference and Sensor Cantilevers in an Array 460
15.3 **Modes of Operation** 460
15.3.1 Static mode 460
15.3.2 Dynamic Mode 462
15.3.3 Heat Mode 463
15.3.4 Further Operation Modes 464
15.4 **Microfabrication** 464
15.5 **Measurement Setup** 465
15.5.1 Measurements in Gaseous or Liquid Environment 465
15.5.2 Readout Principles 466
15.5.3 Miniaturized Piezoresistive Arrays 468
15.5.4 Membrane Surface Stress Sensors 468
15.6 **Functionalization Techniques** 468
15.6.1 General Strategy 468
15.6.2 Functionalization Methods 469
15.6.3 Microfluidics 469
15.6.4 Array of Dimension-Matched Capillaries 469
15.6.5 Inkjet Spotting 469
15.7 **Applications** 470
15.7.1 Chemical Detection 470
15.7.2 Biochemical Environment 472
15.7.3 Microcantilever Sensors to Measure Physical Properties 475
15.7.4 Recent Developments 476
15.8 **Conclusions and Outlook** 476
References 477

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_15

15.1 Technique

Sensors are devices that detect, or sense, changes in the environment and produce a measurable output signal. Moreover, a sensor is also a transducer, i. e., it transforms one form of energy into another. Most people will associate sensors with electrical or electronic devices that respond by a signal change when an external physical parameter is changed. Many more types of transducers exist, such as electrochemical (pH probe), electromechanical (piezoelectric actuator, quartz, strain gauge), electroacoustic (gramophone pick-up, microphone), photoelectric (photodiode, solar cell), electromagnetic (antenna), magnetic (Hall-effect sensor, hard-disk head for storage applications), electrostatic (electrometer), thermoelectric (thermocouple, thermoresistors), and electrical (capacitor, resistor). Here we focus on a further type of sensor not yet mentioned: the mechanical sensor. It responds to changes of an external parameter, such as temperature variations or molecule adsorption, by a mechanical response, e.g., by bending or deflection.

15.1.1 Cantilevers

Mechanical sensors consist of a fixed and a movable part. The movable part can be a thin membrane, a plate or a beam, fixed at one or both ends. The structures described here are called cantilevers. A cantilever is defined here as a microfabricated rectangular bar-shaped structure that is longer than it is wide and has a thickness that is much smaller than its length or width. It is a horizontal structural element supported only at one end on a chip body; the other end is free (Fig. 15.1). Most often it is used as a mechanical probe to image the topography of a sample in a technique called atomic force microscopy (AFM) or scanning force microscopy (SFM) [15.1], invented by *Binnig* et al. in the mid 1980s [15.1]. For AFM a microfabricated sharp tip is located at the apex of the cantilever that serves as a local probe to scan the sample surface. The distance between tip and surface is controlled via sensitive measurement of interatomic forces in the piconewton range.

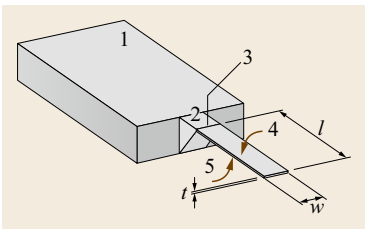


Fig. 15.1 Schematic of a cantilever: (1) rigid chip body, (2) solid cantilever support structure, (3) hinge of cantilever, (4) upper surface of the cantilever that is usually functionalized with a sensor layer for detection of molecules, (5) lower surface of the cantilever, usually passivated in order not to show affinity to the molecules to be detected. The geometrical dimensions, length l, width w and thickness t, are indicated

By scanning the tip across a conductive or nonconductive surface using an x-y-z actuator system (e.g., a piezoelectric scanner), an image of the topography is obtained by recording for all positions x and y the correction signal that has to be applied to the z-actuation drive to keep the interaction between tip and sample surface constant. SFM methods are well established in scientific research, education and, to a certain extent, also in industry. Beyond imaging of surfaces, cantilevers have been used for many other purposes, e.g., structuring of surfaces. Here, we focus on the application as mechanical sensor devices, responding by mechanical bending or change in resonance frequency.

15.1.2 History of Cantilever Sensors

The idea to use beams of silicon as sensors to measure deflections or changes in resonance frequency is actually quite old. First reports go back to 1968, when *Wilfinger* et al. [15.2] investigated silicon cantilever structures of 50 mm × 30 mm × 8 mm, i. e., quite large cantilever beams, for detecting resonances. On the one hand, they took advantage of localized thermal expansion in diffused resistors (piezoresistors) located near the cantilever support to create a temperature gradient to actuate the cantilever at its resonance frequency. On the other hand, the piezoresistors have also been used to sense mechanical deflection of the cantilever. This early report already contains essential concepts for sensing and actuation of cantilevers. In the following years only a few reports were published on the use of cantilevers as sensors, e.g., *Heng* [15.3], who fabricated in 1971 gold cantilevers capacitively coupled to microstrip lines for mechanical trimming of high-frequency oscillator circuits. In 1979, *Petersen* [15.4] constructed cantilever-type micromechanical membrane switches in silicon that should have filled the gap between silicon transistors and mechanical electromagnetic relays. *Kolesar* [15.5] suggested the use of cantilever structures as electronic nerve agent detectors in 1985.

Only with the availability of microfabricated cantilevers for AFM [15.1] did reports on the use of cantilevers as sensors become more frequent. In 1994, *Itoh* et al. [15.6] presented a cantilever coated with a thin film of zinc oxide and proposed a piezoresistive deflection readout as an alternative to the optical beam

deflection readout. *Cleveland* et al. [15.7] reported the tracking of cantilever resonance frequency to detect nanogram changes in mass loading when small particles were deposited onto AFM probe tips. *Thundat* et al. [15.8] showed that the resonance frequency as well as static bending of microcantilevers (MCs) are influenced by ambient conditions, such as moisture adsorption, and that deflection of metal-coated cantilevers can be further influenced by thermal effects (bimetallic effect). First chemical sensing applications were presented by *Gimzewski* et al. [15.9], who used static cantilever bending to detect chemical reactions with very high sensitivity. Later *Thundat* et al. [15.10] observed changes in the resonance frequency of microcantilevers due to adsorption of analyte vapor on exposed surfaces. Furthermore, frequency changes have been observed caused by mass loading or adsorption-induced changes in cantilever spring constant. By coating cantilever surfaces with hygroscopic materials, such as phosphoric acid or gelatin, the cantilever can sense water vapor with picogram mass resolution.

The deflection of individual cantilevers can easily be determined applying the optical beam deflection technique and modified AFM electronics. A single cantilever can be applied as a sensor, but its response can be prone to artifacts such as thermal drifts or unspecific adsorption. For this reason the use of cantilever arrays including passivated reference cantilevers is recommended. The first use of cantilever arrays with sensor and reference cantilevers was reported in 1998 [15.11], and represented significant progress for the understanding of true cantilever responses (calculation of difference responses of sensor and reference cantilevers).

15.2 Cantilever Array Sensors

15.2.1 Concept

For the use of a cantilever as a sensor, neither a sharp tip at the cantilever apex nor a sample surface is required. The upper or lower surface of the cantilever serves as the sensor area and allows the processes taking place on the cantilever beam to be monitored with unprecedented accuracy, in particular the adsorption of molecules. The formation of molecule layers on the cantilever surface will generate surface stress, eventually resulting in a bending of the cantilever, provided the adsorption preferentially occurs on one of the surfaces of the cantilever. Adsorption is controlled by coating one surface (typically the upper one) of a cantilever with a thin layer of a material that shows affinity to molecules in the environment (sensor surface). This surface of the cantilever is referred to as the *functionalized* surface. The other surface of the cantilever (typically the lower one) may be left uncoated or be coated with a passivation layer, i. e., a chemical surface that does not exhibit substantial affinity to the molecules from the environment we would like to detect. To create functionalized surfaces, often a metal layer is evaporated onto the area to be used as the sensor surface. Metal surfaces, e.g., gold, may be used to covalently bind a monolayer representing the chemical surface sensitive to the molecules to be detected from environment. Frequently, a monolayer of thiol molecules is used, which is covalently bound to a gold surface. The gold layer is also favorable for use as a reflection layer if the bending of the cantilever is read out via an optical beam deflection method.

15.2.2 Compressive and Tensile Stress

A cantilever coated with gold on its upper surface for adsorption of alkanethiol molecules and left uncoated on its lower surface (consisting of silicon and silicon oxide) will respond in the following way to the adsorption of thiol molecules: the molecules will covalently bind to the gold layer on the upper surface of the cantilever, resulting in a downward bending of the cantilever due to the formation of surface stress. We will call this process *development of compressive surface stress*, because the surface stress change associated with the formation of the self-assembled monolayer (SAM) produces forces stretching laterally the adsorbed layer, resulting in a downward bending of the cantilever (away from the gold coating). In the opposite situation, i. e., when the cantilever bends upwards, we would speak of tensile stress. If both upper and lower surfaces of the cantilevers are involved in the reaction, then the situation will be much more complex, as a predominant compressive stress formation on the lower cantilever surface might appear as tensile stress on the upper surface. For this reason, it is of utmost importance that the lower cantilever surface is passivated in such a way that ideally no processes will take place on the lower surface of the cantilever.

15.2.3 Disadvantages of Single Microcantilevers

Single microcantilevers are susceptible to parasitic deflections that may be caused by thermal drift or chem-

ical interaction of a cantilever with its environment, in particular if the cantilever is operated in a liquid. Often, a baseline drift is observed while measuring cantilever bending. Moreover, nonspecific physisorption of molecules on the cantilever surface or nonspecific binding to receptor molecules during measurements may contribute to the drift, as well as electrostatic effects.

15.2.4 Reference and Sensor Cantilevers in an Array

To exclude such unwanted influences, simultaneous measurement of reference cantilevers aligned in the same array as the sensing cantilevers is crucial [15.11]. As the difference in signals from the reference and sensor cantilevers represents the net cantilever response, even tiny sensor responses may be extracted from large cantilever deflections without being dominated by undesired effects. When only single microcantilevers are used, no thermal-drift compensation is possible. To obtain useful data under such circumstances, both microcantilever surfaces have to be chemically well defined. One of the surfaces, typically the lower one, has to be passivated; otherwise the cantilever response will be influenced through undesired effects originating from uncontrolled reactions taking place on the lower surface (Fig. 15.2a). However, with a pair of cantilevers, reliable measurements can be obtained. One cantilever is used as the sensor cantilever (coated typically on the upper side with a molecule layer showing substantial affinity to the molecules to be detected), whereas the other cantilever serves as the reference cantilever. Actually, it should be coated with a passivation layer on the upper surface so as not to exhibit much affinity to the molecules to be detected. In this configuration, thermal drifts are canceled out if difference responses are taken, i. e., difference in deflections of sensor and reference cantilevers. Alternatively, both cantilevers may be used as sensor cantilevers (sensor layer on the upper surfaces), provided the lower surface has been passivated (Fig. 15.2b). However, it is best to apply a cantilever array (Fig. 15.2c), in which several cantilevers are used either as sensor or as reference cantilevers so that multiple difference signals can be evaluated simultaneously. Moreover, thermal drift is canceled out if one surface of all cantilevers, typically the lower one, is left uncoated or coated with the same passivation layer.

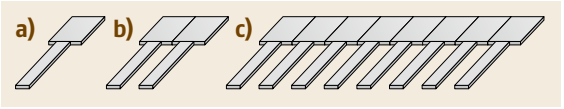


Fig. 15.2a–c Cantilever sensor arrangements. **(a)** Single cantilever; **(b)** pair of cantilevers, one to be used as the sensor cantilever, the other as the reference cantilever, and **(c)** array of cantilevers with several sensor and reference cantilevers

15.3 Modes of Operation

In analogy to AFM, various operating modes for cantilevers are described in the literature. The measurement of static deflection upon formation of surface stress during adsorption of a molecular layer is termed *static mode*. *Ibach* used cantilever-like structures to study adsorbate-induced surface stress [15.12] in 1994. Surface stress-induced bending of cantilevers during the adsorption of alkanethiols on gold was reported by *Berger* et al. in 1997 [15.13]. The mode corresponding to the technique of noncontact AFM is termed *dynamic mode*, where a cantilever is oscillated at its resonance frequency, and was described by *Cleveland* et al. [15.7]. They calculated mass changes from shifts in the cantilever resonance frequency upon the mounting of tiny tungsten particle spheres at the apex of the cantilever. Not only adsorption effects can provoke cantilever bending, but also thermal effects. The so-called *heat mode* was pioneered by *Gimzewski* et al. [15.9], who took advantage of the bimetallic effect that produces bending of a metal-coated cantilever when heat is generated on its surface. Therewith they constructed a miniaturized calorimeter with picojoule sensitivity. Further operating modes exploit other physical effects such as the production of heat from the absorption of light by materials deposited on the cantilever (photothermal spectroscopy) [15.14], or cantilever bending due to electric or magnetic forces.

15.3.1 Static mode

The continuous bending of a cantilever caused by increasing coverage with molecules is referred to as operation in the *static mode* (Fig. 15.3a). Adsorption of molecules onto the functional layer produces stress at the interface between the functional layer and the forming molecular layer. A quick estimation of the weight of the adsorbed layer of molecules clearly implies that the bending of the cantilever is not due to the mass increase by the adsorbed molecular layer. Because the forces within the functional layer try to keep the dis-

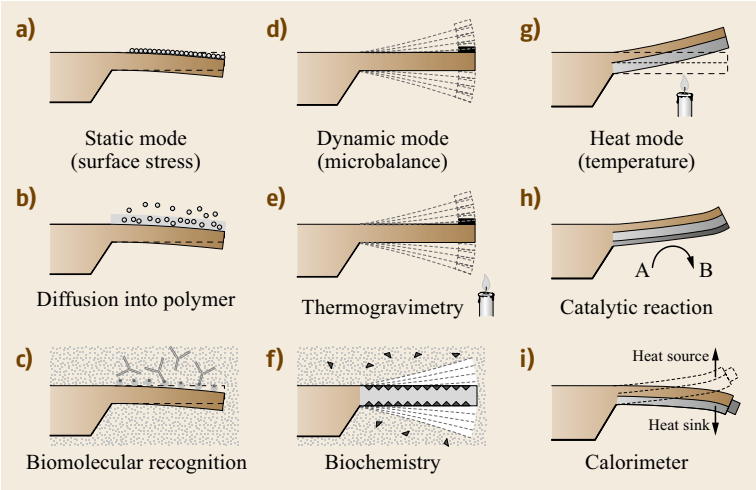


Fig. 15.3a–i Basic cantilever operation modes: **(a)** static bending of a cantilever on adsorption of a molecular layer. **(b)** Diffusion of molecules into a polymer layer leads to swelling of the polymer and eventually to a bending of the cantilever. **(c)** Highly specific molecular recognition of biomolecules by receptors changes surface stress on the upper surface of the cantilever and results in bending. **(d)** Oscillation of a cantilever at its resonance frequency (dynamic mode) allows information on mass changes taking place on the cantilever surface to be obtained (application as a microbalance). **(e)** Changing the temperature while a sample is attached to the apex of the cantilever allows information to be gathered on decomposition or oxidation processes. **(f)** Dynamic mode measurements in liquids yield details on mass changes during biochemical processes. **(g)** In heat mode, a bimetallic cantilever is employed. Here bending is due to the difference in the thermal expansion coefficients of the two materials. **(h)** Catalysic reactions produce heat, resulting in cantilever bending. **(i)** A tiny sample attached to the apex of the cantilever is investigated taking advantage of the bimetallic effect. Tracking the deflection as a function of temperature allows the observation of phase transitions in the sample in a calorimeter mode

tance between molecules constant, the cantilever beam responds by bending because of its extreme flexibility. This property is described by the spring constant k of the cantilever, which for a rectangular microcantilever of length l, thickness t, and width w is calculated as follows

$$k = \frac{Ewt^3}{4l^3} , \tag{15.1}$$

where E is the Young's modulus ($E_{Si} = 1.3 \times 10^{11}\,N/m^2$ for Si(100)).

As a response to surface stress, e.g., owing to adsorption of a molecular layer, the microcantilever bends, and its shape can be approximated as part of a circle with radius R. This radius of curvature is given by [15.15, 16]

$$\frac{1}{R} = \frac{6\,(1-\nu)}{Et^2} . \tag{15.2}$$

The resulting surface stress change is described using Stoney's formula [15.15]

$$\Delta\sigma = \frac{Et^2}{6R\,(1-\nu)} , \tag{15.3}$$

where E is Young's modulus, t the thickness of the cantilever, ν the Poisson's ratio ($\nu_{Si} = 0.24$), and R the bending radius of the cantilever.

Static mode operation has been reported in various environments. In its simplest configuration, molecules from gaseous environments adsorb on the functionalized sensing surface and form a molecular layer (Fig. 15.3a), provided the molecules exhibit at least some affinity to the surface. In case of alkanethiol covalently binding to gold, the affinity is very high, resulting in a fast bending response within minutes [15.13]. A less specific way of detection takes advantage of the penetrable network of polymer layers. Polymer sensing layers only exhibit a partial sensitivity, i. e., polymer-coated cantilevers always respond to the presence of volatile molecules, but the magnitude and temporal behavior are specific to the chemistry of the polymer. Molecules from the environment diffuse into the polymer layer at different rates, mainly depending on the size and solubility of the molecules in the polymer layer (Fig. 15.3b). A wide range of hydrophilic or hydrophobic polymers can be selected as sensing layers, differing in their affinity to polar and unpolar molecules. Thus, polymer coatings can be chosen ac-

cording to the respective requirements of the sensing application.

Static-mode operation in liquids, however, usually requires rather specific sensing layers, and is based on molecular recognition, e.g., in DNA hybridization [15.17] or antigen-antibody recognition (Fig. 15.3c). Cantilevers functionalized by coating with biochemical sensing layers respond very specifically using the biomolecular key-lock principles of molecular recognition. Whether molecular recognition will actually lead to bending of the cantilever will depend on the efficiency of transduction, as the surface stress has to be generated very close to the cantilever surface to produce substantial bending. By just scaling down standard gene chip strategies to cantilever geometry through utilization of long spacer molecules, hybridization takes place at a distance of several nanometers away from the cantilever surface. The original idea to make DNA molecules become more accessible for hybridization in a gene chip configuration, is unfavorable for cantilever-based sensing through surface stress detection, as the magnitude of bending is greatly reduced or not even observable [15.18].

15.3.2 Dynamic Mode

Since the eigenfrequency of an oscillating cantilever depends on its mass, we can determine mass changes accurately by tracking shifts in eigenfrequency. Technically, the resonance frequency of the cantilever can be measured more easily than the eigenfrequency. Often, the resonance frequency is measured by actuating the cantilever actively, e.g., by a piezoelectric actuator. The resonance frequency represents the frequency at which the oscillation amplitude is largest. The eigenfrequency is equal to the resonance frequency of an oscillating cantilever if the elastic properties of the cantilever remain unchanged and damping effects are insignificant during the molecular adsorption process. This mode of operation is called dynamic mode (e.g., the use of cantilevers as a microbalance, Fig. 15.3d). Owing to mass addition on the cantilever surface, the eigenfrequency of the cantilever will shift to a lower value. The frequency change per mass change on a rectangular cantilever is calculated [15.19] according to

$$\frac{\Delta f}{\Delta m} = \frac{1}{4\pi n_1 l^3 w}\sqrt{\frac{E}{\rho^3}}\,, \tag{15.4}$$

where $\rho = m/lwt$ is the mass density of the microcantilever and the deposited mass, and $n_1 \approx 1$ is a geometrical factor.

The mass change is calculated [15.8] from the frequency shift using

$$\Delta m = \frac{k}{4\pi^2}\left(\frac{1}{f_1^2} - \frac{1}{f_0^2}\right)\,, \tag{15.5}$$

where f_0 is the eigenfrequency before the mass change occurs, and f_1 the eigenfrequency after the mass change. Please note that this formula is only accurate for eigenfrequencies and not for resonance frequencies, if damping cannot be neglected.

Mass-change determination can be combined with varying conditions of environmental temperature (Fig. 15.3e) to obtain a method introduced in the literature as *micromechanical thermogravimetry* [15.20]. For investigation of thermal properties, a tiny piece of sample has to be mounted at the apex of the cantilever. Its mass should not exceed several hundred nanograms. Adsorption, desorption and decomposition processes, occurring while changing the temperature, produce mass changes in the picogram range that can be observed in real time by tracking the resonance-frequency shift.

Dynamic mode operation in a liquid environment is more difficult than in air, because of significant damping of the cantilever oscillation due to the high viscosity of the surrounding media (Fig. 15.3f). Significant damping results in a low quality factor Q of the oscillation, and thus the resonance frequency shift is difficult to track with high resolution. The quality factor is defined as

$$Q = 2\frac{\Delta f}{f_0}\,. \tag{15.6}$$

Whereas in air the resonance frequency can easily be determined with a resolution of below 1 Hz, only a frequency resolution of about 20 Hz is expected for measurements in liquid environments.

The damping or altered elastic properties of the cantilever during the experiment, e.g., by stiffening or softening of the spring constant caused by adsorption of a molecule layer, result in the fact that the measured resonance frequency will not be exactly equal to the eigenfrequency of the cantilever, and therewith the mass derived from the frequency shift will be inaccurate. In a medium, the vibration of a cantilever is described by the model of a driven damped harmonic oscillator

$$m^*\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}x}{\mathrm{d}t} + kx = F\cos(2\pi ft)\,, \tag{15.7}$$

where $m^* = const\,(m_c + m_l)$ is the effective mass of the cantilever (for a rectangular cantilever the constant is

0.25). Especially in liquids, the mass of the co-moved liquid m_l adds significantly to the mass of the cantilever m_c. The term $\gamma dx/dt$ is the drag force due to damping, $F\cos(2\pi ft)$ is the driving force executed by the piezoelectric actuator, and k is the spring constant of the cantilever.

If no damping is present, the eigenfrequencies of the various oscillation modes of a bar-shaped cantilever are calculated according to

$$f_n = \frac{\alpha_n^2}{2\pi}\sqrt{\frac{k}{2\,(m_c + m_l)}}\,, \tag{15.8}$$

where f_n are the eigenfrequencies in n-th mode, α_n are constants depending on the mode: $\alpha_1 = 1.8751$, $\alpha_2 = 4.6941$, $\alpha_n = \pi\,(n - 0.5)$. k is the spring constant of the cantilever, m_c the mass of the cantilever, and m_l the mass of the medium surrounding the cantilever, e.g., liquid [15.21].

Addition of mass to the cantilever due to adsorption will change the effective mass as follows

$$m^* = const\,(m_c + m_l + \Delta m)\,, \tag{15.9}$$

where Δm is the additional mass adsorbed. Typically, the co-moved mass of the liquid is much larger than the adsorbed mass.

Figure 15.4 clearly shows that the resonance frequency is only equal to the eigenfrequency if no damping is present. With damping, the frequency at which the peak of the resonance curve occurs is no longer identical with the one at which the turning point of the phase curve occurs. For example, resonance curve 2 with damping γ_2 has its maximum amplitude at frequency f_2. The corresponding phase would be $\phi_{res}(\gamma_2)$, which is not equal to $\pi/2$, as would be expected in the undamped case. If direct resonance frequency tracking or a phase-locked loop is used to determine the frequency of the oscillating cantilever, then only its resonance frequency is detected, but not its eigenfrequency. Remember that the eigenfrequency, and not the resonance frequency, is required to determine mass changes.

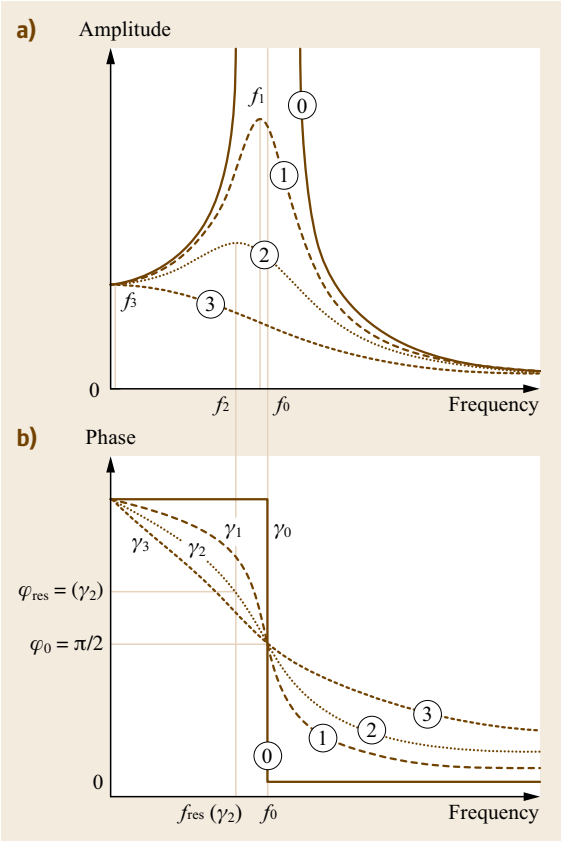


Fig. 15.4 **(a)** Resonance curve with no damping (0), and increasing damping (1)–(3). The undamped curve with resonance frequency f_0 exhibits a very high amplitude, whereas the resonance peak amplitude decreases with damping. This also involves a shift in resonance frequencies f_1 to f_3 to lower values. **(b)** Corresponding phase curves showing no damping (0), and increasing damping (1)–(3). The step-like phase jump for the undamped resonance gradually broadens with increased damping

15.3.3 Heat Mode

If a cantilever is coated with metal layers, thermal expansion differences in the cantilever and the coating layer will further influence cantilever bending as a function of temperature. This mode of operation is referred to as the *heat mode* and causes cantilever bending because of differing thermal expansion coefficients in the sensor layer and cantilever materials [15.9] (Fig. 15.3g)

$$\Delta z = 1.25\,(\alpha_1 - \alpha_2)\,\frac{(t_1 + t_2)}{t_2^2\kappa}\,\frac{l^3}{(\lambda_1 t_1 + \lambda_2 t_2)\,w}P\,. \tag{15.10}$$

Here α_1, α_2 are the thermal expansion coefficients of the cantilever and coating materials, respectively, t_1, t_2 the material thicknesses, P is the total power generated on the cantilever, and κ is a geometry parameter of the cantilever device. λ_1 and λ_2 are the thermal conductivities of layers 1 and 2.

Heat changes are either caused by external influences (change in temperature, Fig. 15.3g), occur directly on the surface by exothermal, e.g., catalytic,

reactions (Fig. 15.3h), or are due to material properties of a sample attached to the apex of the cantilever (micromechanical calorimetry, Fig. 15.3i). The sensitivity of the cantilever heat mode is orders of magnitude higher than that of traditional calorimetric methods performed on milligram samples, as it only requires nanogram amounts of sample and achieves nanojoule [15.20], picojoule [15.22], and femtojoule [15.23] sensitivity.

These three measurement modes have established cantilevers as versatile tools to perform experiments in nanoscale science with very small amounts of material.

15.3.4 Further Operation Modes

Photothermal Spectroscopy

When a material adsorbs photons, a fraction of the energy is converted into heat. This photothermal heating can be measured as a function of the light wavelength to provide optical absorption data of the material. The interaction of light with a bimetallic microcantilever creates heat on the cantilever surface, resulting in a bending of the cantilever [15.14]. Such bimetallic-cantilever devices are capable of detecting heat flows due to an optical heating power of 100 pW, being two orders of magnitude better than in conventional photothermal spectroscopy.

Electrochemistry

A cantilever coated with a metallic layer (measurement electrode) on one side is placed in an electrolytic medium, e.g., a salt solution, together with a metallic reference electrode, usually made of a noble metal. If the voltage between the measurement and the reference electrode is changed, electrochemical processes on the measurement electrode (cantilever) are induced, such as adsorption or desorption of ions from the electrolyte solution onto the measurement electrode. These processes lead to a bending of the cantilever due to changes in surface stress and in the electrostatic forces [15.24].

Detection of Electrostatic and Magnetic Forces

The detection of electrostatic and magnetic forces is possible if charged or magnetic particles are deposited on the cantilever [15.25, 26]. If the cantilever is placed in the vicinity of electrostatic charges or magnetic particles, attractive or repulsion forces occur according to the polarity of the charges or magnetic particles present on the cantilever. These forces will result in upward or downward bending of the cantilever. The magnitude of bending depends on the distribution of charged or magnetic particles on both the cantilever and in the surrounding environment according to the laws of electrostatics and magnetism.

15.4 Microfabrication

Several methods have been applied for microfabrication of cantilevers out of various materials, using either wet or dry etching techniques. The silicon cantilever sensor arrays described here have been microfabricated using a dry-etching silicon-on-insulator (SOI) fabrication technique developed in the Micro-/Nanomechanics Department at IBM's Zurich Research Laboratory. One chip comprises eight cantilevers, having a length of 500 μm, a width of 100 μm, and a thickness of 0.5 μm, and arranged at a pitch of 250 μm. For dynamic mode, the cantilever thickness may amount up to 7 μm to obtain higher resonance frequencies and better mass resolution. The resonance frequencies of the cantilevers vary by 0.5% only, demonstrating the high reproducibility and precision of cantilever fabrication. A scanning electron microscopy image of a cantilever sensor array chip is shown in Fig. 15.5.



Fig. 15.5 Scanning electron micrograph of a cantilever-sensor array. Image courtesy of Viola Barwich, University of Basel, Switzerland ►

15.5 Measurement Setup

15.5.1 Measurements in Gaseous or Liquid Environment

A measurement setup for cantilever arrays consists of four major parts:

1. The measurement chamber containing the cantilever array.
2. An optical or electrical system to detect the cantilever deflection (e.g., laser sources, collimation lenses and a position-sensitive detector (PSD), or piezoresistors and Wheatstone bridge detection electronics).
3. Electronics to amplify, process and acquire the signals from the detector.
4. A gas- or liquid-handling system to reproducibly inject samples into the measurement chamber and purge the chamber.

The following description applies to our measurement setup.

Figure 15.6 shows the schematic setup for experiments performed in (Fig. 15.6a) a gaseous and (Fig. 15.6b) a liquid (biochemical) environment for the optical beam deflection embodiment of the measurement setup. The cantilever sensor array is placed in an analysis chamber 3–90 µl in volume, which has inlet and outlet ports for gases or liquids. The cantilever deflection is determined by means of an array of eight vertical-cavity surface-emitting lasers (VCSELs) arranged at a linear pitch of 250 µm that emit at a wavelength of 760 nm into a narrow cone of 5–10°.

The light of each VCSEL is collimated and focused onto the apex of the corresponding cantilever by a pair of achromatic doublet lenses, 12.5 mm in diameter. This size has to be selected in such a way that all eight laser beams pass through the lens close to its center to minimize scattering, chromatic and spherical aberration artifacts. The light is then reflected off the gold-coated surface of the cantilever and hits the surface of a PSD. PSDs are light-sensitive photo-potentiometer-like devices that produce photocurrents at two opposing electrodes. The magnitude of the photocurrents depends linearly on the distance of the impinging light spot from the electrodes. Thus, the position of an incident light

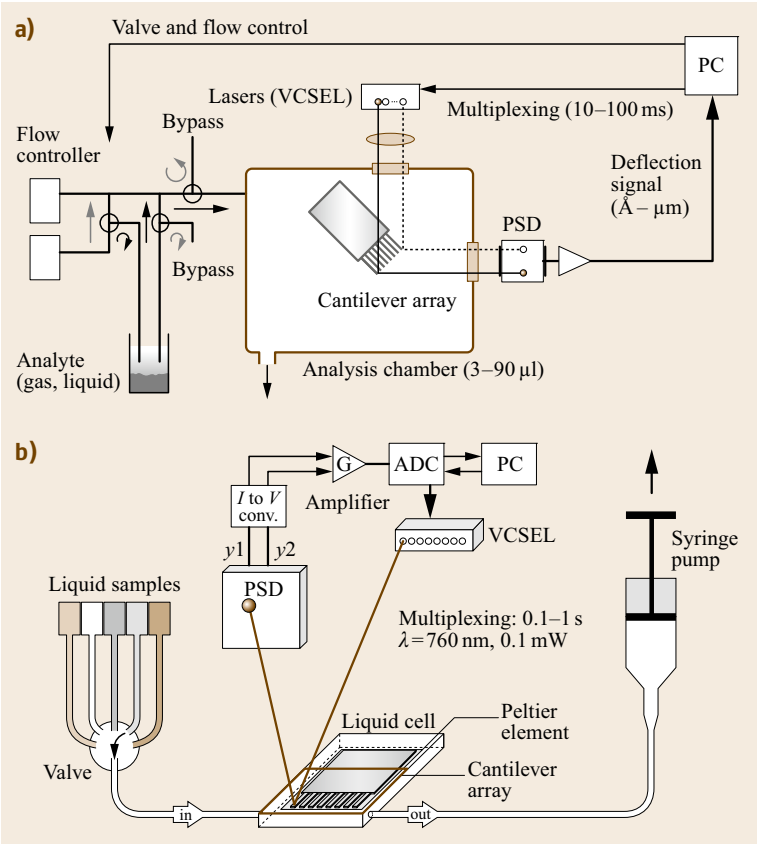


Fig. 15.6a,b Schematic of measurement setups for **(a)** a gaseous (artificial nose) and **(b)** a liquid environment (biochemical sensor)

beam can easily be determined with micrometer precision. The photocurrents are transformed into voltages and amplified in a preamplifier. As only one PSD is used, the eight laser light sources cannot be switched on simultaneously. Therefore, a time-multiplexing procedure is used to switch the lasers on and off sequentially at typical intervals of 10–100 ms. The resulting deflection signal is digitized and stored together with time information on a personal computer (PC), which also controls the multiplexing of the VCSELs as well as the switching of the valves and mass flow controllers used for setting the composition ratio of the analyte mixture.

The measurement setup for liquids (Fig. 15.6b) consists of a poly-etheretherketone (PEEK) liquid cell, which contains the cantilever array and is sealed by a Viton O-ring and a glass plate. The VCSELs and the PSD are mounted on a metal frame around the liquid cell. After preprocessing the position of the deflected light beam in a current-to-voltage converter and amplifier stage, the signal is digitized in an analog-to-digital converter and stored on a PC. The liquid cell is equipped with inlet and outlet ports for liquids. They are connected via 0.18 mm of i.d. Teflon tubing to individual thermally equilibrated glass containers, in which the biochemical liquids are stored. A six-position valve allows the inlet to the liquid chamber to be connected to each of the liquid-sample containers separately. The liquids are pulled (or pushed) through the liquid chamber by means of a syringe pump connected to the outlet of the chamber. A Peltier element is situated very close to the lumen of the chamber to allow temperature regulation within the chamber. The entire experimental setup is housed in a temperature-controlled box regulated with an accuracy of 0.01 K to the target temperature.

15.5.2 Readout Principles

This section describes various ways to determine the deflection of cantilever sensors. They differ in sensitivity, effort for alignment and setup, robustness, and ease of readout as well as in potential for miniaturization.

Piezoresistive Readout

Piezoresistive cantilevers [15.6, 20] are usually U-shaped having diffused piezoresistors in both of the legs close to the hinge (Fig. 15.7a). The resistance in the piezoresistors is measured by a Wheatstone bridge technique employing three reference resistors, one of which is adjustable. The current flowing between the two branches of the Wheatstone bridge is initially nulled by changing the resistance of the adjustable resistor. If the cantilever bends, the piezoresistor changes its value and a current will flow between the two branches of the Wheatstone bridge. This current is converted via a differential amplifier into a voltage for static mode measurement. For dynamic mode measurement, the piezoresistive cantilever is externally actuated via a frequency generator connected to a piezocrystal. The ac actuation voltage is fed as reference voltage into a lock-in amplifier and compared with the response of the Wheatstone bridge circuit. This technique allows one to sweep resonance curves and to determine shifts in resonance frequency.

Piezoelectric Readout

Piezoelectric cantilevers [15.27] are actuated by applying an electric ac voltage via the inverse piezoelectric effect (self-excitation) to the piezoelectric material (PZT or ZnO). Sensing of bending is performed by recording the piezoelectric current change due to the fact that the PZT layer may produce a sensitive field response to weak stress through the direct piezoelectric effect. Such cantilevers are multilayer structures consisting of an SiO_2 cantilever and the PZT piezoelectric layer. Two electrode layers, insulated from each other, provide electrical contact. The entire structure is protected using passivation layers (Fig. 15.7b). An identical structure is usually integrated into the rigid chip body to provide a reference for the piezoelectric signals from the cantilever.

Capacitive Readout

For capacitive readout (Fig. 15.7c), a rigid beam with an electrode mounted on the solid support and a flexible cantilever with another electrode layer are used [15.28, 29]. Both electrodes are insulated from each other. Upon bending of the flexible cantilever the capacitance between the two electrodes changes and allows the deflection of the flexible cantilever to be determined. Both static and dynamic measurement modes are possible.

Optical (Interferometric) Readout

Interferometric methods [15.30, 31] are most accurate for the determination of small movements. A laser beam passes through a polarizer plate (polarization 45°) and is partially transmitted by a nonpolarized beam splitter (Fig. 15.7d). The transmitted beam is divided in a Wollaston prism into a reference and an object beam. These mutually orthogonally polarized beams are then focused onto the cantilever. Both beams (the reference beam from the hinge region and the object beam from the apex region of the cantilever) are reflected back to the objective lens, pass the Wollaston prism, where they are re-combined into one beam, which is then reflected into the other arm of the interferometer, where after the $\lambda/4$ plate a phase shift of a quarter wavelength

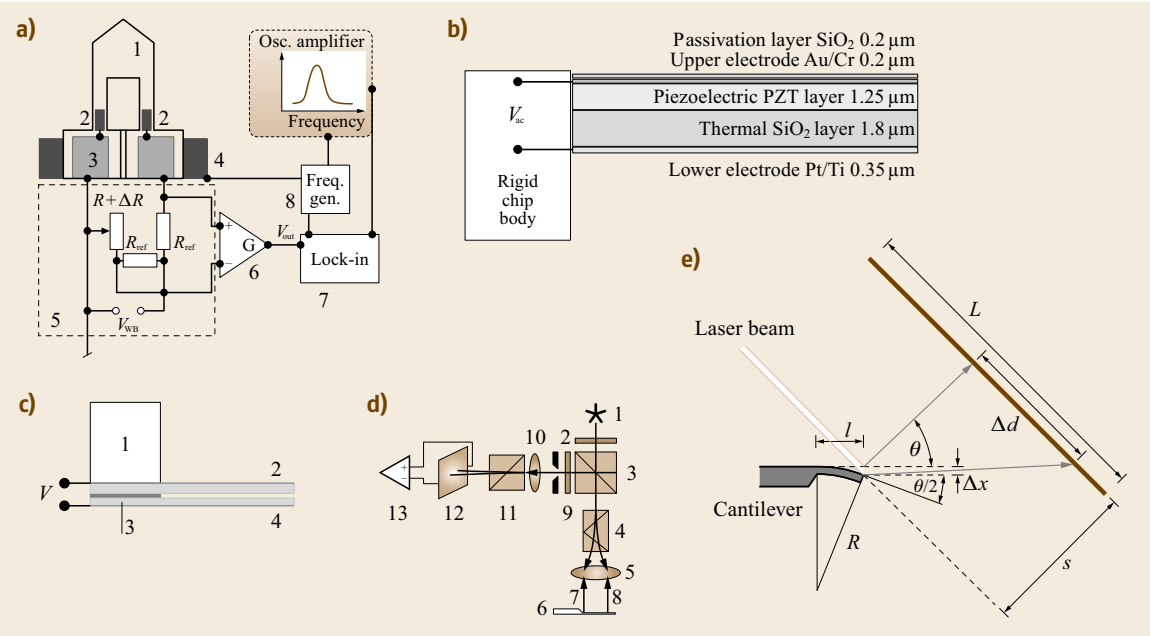


Fig. 15.7 **(a)** Piezoresistive readout: 1 cantilever, 2 piezoresistors, 3 Au contact pads, 4 external piezocrystal for actuation, 5 Wheatstone bridge circuit, 6 differential amplifier, 7 lock-in amplifier, 8 function generator. **(b)** Piezoelectric readout. **(c)** Capacitive readout: 1 solid support, 2 rigid beam with counterelectrode, 3 insulation layer (SiO_2), 4 flexible cantilever with electrode. **(d)** Interferometric readout: 1 laser diode, 2 polarizer, 3 nonpolarizing beam splitter, 4 Wollaston prism, 5 focusing lens, 6 cantilever, 7 reference beam (near cantilever hinge), 8 object beam (near cantilever apex), 9 diaphragm and $\lambda/4$ plate, 10 focusing lens, 11 Wollaston prism, 12 quadrant photodiode, 13 differential amplifier. **(e)** Beam-deflection readout

between object and reference beam is established. Another Wollaston prism separates reference and object beams again for analysis with a four-quadrant photodiode. A differential amplifier is used to obtain the cantilever deflection with high accuracy. However, the interferometry setup is quite bulky and difficult to handle.

Optical (Beam-Deflection) Readout

The most frequently used approach to read out cantilever deflections is optical beam deflection [15.32], because it is a comparatively simple method with an excellent lateral resolution. The schematic is shown in Fig. 15.7e.

The actual cantilever deflection Δx scales with the cantilever dimensions. Therefore, the surface stress $\Delta\sigma$ in N/m is a convenient quantity to measure and compare cantilever responses. It takes into account the cantilever material properties, such as the Poisson ratio ν, the Young's modulus E, and the cantilever thickness t. The radius of curvature R of the cantilever is a measure of bending, see (15.2). As shown in the drawing in Fig. 15.7e, the actual cantilever displacement Δx is transformed into a displacement Δd on the PSD. The position of a light spot on a PSD is determined by measuring the photocurrents from the two facing electrodes. The movement of the light spot on the linear PSD is calculated from the two currents I_1 and I_2 and the size L of the PSD by

$$\Delta d = \frac{I_1 - I_2}{I_1 + I_2}\frac{L}{2} \,. \tag{15.11}$$

As all angles are very small, it can be assumed that the bending angle of the cantilever is equal to half of the angle θ of the deflected laser beam, i. e., $\theta/2$. Therefore, the bending angle of the cantilever can be calculated to be

$$\frac{\theta}{2} = \frac{\Delta d}{2s} \,, \tag{15.12}$$

where s is the distance between the PSD and the cantilever. The actual cantilever deflection Δx is calculated from the cantilever length l and the bending angle $\theta/2$ by

$$\Delta x = \frac{\theta/2}{2} l \,. \tag{15.13}$$

Combination of (15.12) and (15.13) relates the actual cantilever deflection Δx to the PSD signal

$$\Delta x = \frac{l\Delta d}{4s} . \tag{15.14}$$

The relation between the radius of curvature and the deflection angle is

$$\frac{\theta}{2} = \frac{l}{R} , \tag{15.15}$$

and after substitution becomes

$$R = \frac{2ls}{\Delta d} , \tag{15.16}$$

or

$$R^{-1} = \frac{2\Delta x}{l^2} .$$

15.5.3 Miniaturized Piezoresistive Arrays

One of the drawbacks of cantilever array sensors is the bulky size of equipment required for optical cantilever deflection readout. Especially for the development of handheld devices one would prefer a more compact type of measurement electronics. This disadvantage can be overcome by employing another method for deflection detection involving the use of piezoresistor elements to determine bending. Since no alignment optics are required, the complete readout electronics for piezoresistive sensors easily fits in a portable box of typically 10 cm × 10 cm × 16 cm, including data acquisition and gas handling.

15.5.4 Membrane Surface Stress Sensors

Membrane-type surface stress sensors (MSS) were first described by *Yoshikawa* et al. [15.33]. Their application for medical sensing was reported in a publication by *Loizeau* et al. [15.34, 35]. MSS are also arranged in arrays and are mainly used for molecular detection in the gaseous phase. MSS have been microfabricated from silicon-on-insulator substrates and structured by deep reactive ion etching. The sensors consists of a round membrane with a diameter of typically 500 µm and a thickness of 2.5 µm, which serves as a platform for applying the sensing layer (typically a polymer layer similar to that used in cantilever array-based electronic noses). The preferred way of coating the membranes is inkjet spotting of polymer solution, but also spray-coating or spin-coating is applicable. The membrane is suspended by four sensing beams with integrated p-type piezoresistors, representing a full Wheatstone bridge (Fig. 15.8). The four piezoresistor bridges are wired in such a way that their responses add up, increasing the sensitivity by a factor of 4 as compared to piezoresistive cantilevers. Further geometrical enhancements reported by *Yoshikawa* et al. [15.36] allow one to increase the sensitivity of MSS by a factor of 100 compared the previous generation of piezoresistive cantilevers. p-Doped piezoresistors were fabricated using two subsequent doping processes (ion diffusion through Boron Silica Glass and implantation). The latter method features shallow resistors, which are very sensitive to surface stress changes. The measured signal pattern consists of the bending responses of an array of piezoresistive polymer-coated membranes due to exposure to volatile organic compounds (VOCs). Recent applications of MSS are described in [15.37–42].

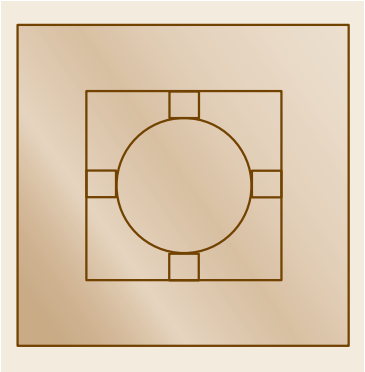


Fig. 15.8 Schematic representation of a membrane-type surface stress sensor (MSS). The actual diameter of the round membrane is 500 µm and its thickness is 2.5 µm. The membrane is suspended by four sensing beams with integrated p-type piezoresistors, representing a full Wheatstone bridge. A solid supporting frame holds the sensor. Several MSS can be arranged in an array

15.6 Functionalization Techniques

15.6.1 General Strategy

To serve as sensors, cantilevers have to be coated with a sensor layer that is either highly specific, i. e., is able to recognize target molecules in a key-lock process, or partially specific, meaning the sensor information from several cantilevers yields a pattern that is characteristic of the target molecules.

To provide a platform for specific functionalization, the upper surface of these cantilevers is typically coated

with 2 nm of titanium and 20 nm of gold, which yields a reflective surface as well as an interface for attaching functional groups of probe molecules, e.g., for anchoring molecules with a thiol group to the gold surface of the cantilever. Such thin metal layers are believed not to contribute significantly to bimetallic bending, particularly because the temperature is kept constant.

15.6.2 Functionalization Methods

There are numerous ways to coat a cantilever with material, both simple and more advanced ones. The method of choice should be fast, reproducible, reliable, and allow one or both of the surfaces of a cantilever to be coated separately.

Simple Methods

Obvious methods to coat a cantilever are thermal or electron beam-assisted evaporation of material, electrospray, or other standard deposition methods. The disadvantage of these methods is that they are only suitable for coating large areas, but not individual cantilevers in an array, unless shadow masks are used. Such masks need to be accurately aligned with the cantilever structures, which is a time-consuming process.

Other methods to coat cantilevers use manual placement of particles onto the cantilever [15.9, 20, 43–45], which requires skillful handling of tiny samples. Cantilevers can also be coated by directly pipetting solutions of the probe molecules onto the cantilevers [15.46] or by employing air-brush spraying and shadow masks to coat the cantilevers individually [15.47]. All these methods have only limited reproducibility and are very time-consuming if a larger number of cantilever arrays has to be coated.

15.6.3 Microfluidics

Microfluidic networks (μFN) [15.48] are structures of channels and wells, etched several tens to hundreds of micrometers deep into silicon wafers or glass. Also polymer replicas, typically from poly(dimethylsiloxane) (PDMS) may be used. The wells can be filled easily using a laboratory pipette, so that the fluid with the probe molecules for coating the cantilever is guided through the channels towards openings at a pitch matched to the distance between individual cantilevers in the array (Fig. 15.9a).

The cantilever array is then introduced into the open channels of the μFN that are filled with a solution of the probe molecules. The incubation of the cantilever array in the channels of the μFN takes from a few seconds (self-assembly of alkanethiol monolayers) to several tens of minutes (coating with protein solutions).

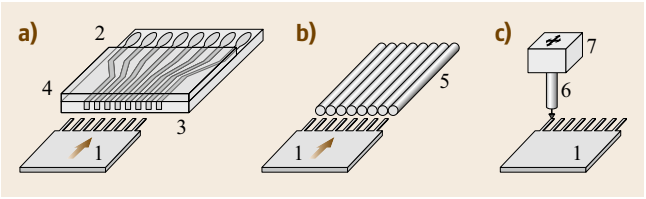


Fig. 15.9 (a) Cantilever functionalization in microfluidic networks. (b) Incubation in dimension-matched microcapillaries. (c) Coating with an inkjet spotter: 1 cantilever array, 2 reservoir wells, 3 microfluidic network with channels, 4 PDMS cover to avoid evaporation, 5 microcapillaries, 6 inkjet nozzle, 7 inkjet *x*-*y*-*z* positioning unit

To prevent evaporation of the solutions, the channels are covered by a slice of PDMS. In addition, the microfluidic network may be placed in an environment filled with saturated vapor of the solvent used for the probe molecules.

15.6.4 Array of Dimension-Matched Capillaries

A similar approach is insertion of the cantilever array into an array of dimension-matched disposable glass capillaries. The outer diameter of the glass capillaries used here is 240 μm so that they can be placed neatly next to each other to accommodate the pitch of the cantilevers in the array (250 μm). Their inner diameter is 150 μm, providing sufficient room to insert the cantilevers (width: 100 μm) safely (Fig. 15.9b). This method has been successfully applied for the deposition of a variety of materials onto cantilevers, such as polymer solutions [15.47], self-assembled monolayers [15.49], thiol-functionalized single-stranded DNA oligonucleotides [15.50], and protein solutions [15.51].

15.6.5 Inkjet Spotting

All of the above techniques require manual alignment of cantilever array and functionalization tool, and are therefore not ideal for coating a large number of cantilever arrays. The inkjet-spotting technique, however, allows rapid and reliable coating of cantilever arrays [15.52, 53]. An *x*-*y*-*z* positioning system allows a fine nozzle (capillary diameter: 70 μm) to be positioned with an accuracy of approx. 10 μm over a cantilever. Individual droplets (diameter: 60–80 μm, volume 0.1–0.3 nl) can be dispensed by means of a piezo-driven ejection system in the inkjet nozzle. When the droplets are spotted with a pitch smaller than 0.1 mm, they merge and form continuous films. By adjusting the number of droplets deposited on cantilevers,

the resulting film thickness can be controlled precisely. The inkjet-spotting technique allows a cantilever to be coated within seconds and yields very homogeneous, reproducibly deposited layers of well-controlled thickness. Successful coating of self-assembled alkanethiol monolayers, polymer solutions, self-assembled DNA single-stranded oligonucleotides [15.53], and protein layers has been demonstrated. In conclusion, inkjet spotting has turned out to be a very efficient and versatile method for functionalization, which can even be used to coat arbitrarily shaped sensors reproducibly and reliably [15.54, 55].

15.7 Applications

15.7.1 Chemical Detection

Hydrogen

Early reports on detection of gases such as hydrogen involved nanomechanical detection of catalytic reactions of bimetallic microcantilevers coated with aluminum and a top layer of platinum in thermal mode [15.9]. The catalytic reaction of oxygen present in a reaction chamber with hydrogen being introduced into the chamber produces oscillatory chemical reactions resulting in mechanical oscillations of the cantilever due to heat formation related to catalytic conversion of H_2 and O_2 to form H_2O. By use of an array of four platinum-coated and four uncoated microcantilevers, a change of the deflection signal due to bending of the platinum-coated cantilever relative to the uncoated cantilevers can be observed upon hydrogen adsorption in the presence of oxygen [15.11]. Similar responses were obtained with Pd-coated glass cantilevers [15.56] and with Pd-coated silicon microcantilevers using dynamic mode [15.57], capacitive readout [15.58] or beam-deflection readout in static mode [15.59].

Water Vapor

The first observation of microcantilever resonance frequency detuning was reported in [15.8]. A dependence on relative humidity of ZFM5 zeolites attached to resonating microcantilevers was observed in [15.60]. Relative humidity was measured with an accuracy of 1% using piezoresistive sensors embedded in polymer [15.61]. A detection limit of 10 ppm was achieved using Al_2O_3-coated microcantilevers [15.62].

Other Vapors

ZFM5 zeolites have been used to detect vapor of *p*-nitroaniline dye in dynamic mode with picogram sensitivity [15.60]. A freon gas sensor using a piezoelectric microcantilever coated with MFI zeolite was described in [15.63]. Ethanol vapor detection in dynamic mode was described in [15.64].

Alkane Thiol Vapors

Surface stress changes and kinetics were measured in situ during the self-assembly of alkanethiols on gold by means of a micromechanical sensor, observing scaling of compressive surface stress with the length of the alkane chain [15.13, 65]. 65 ppb of 2-mercaptoethanol have been measured evaluating the response of gold-coated silicon nitride microcantilevers [15.66]. The mechanism of stress formation upon adsorption of thiol layers has been studied by exposing monolayers of alkanethiols on gold to low-energy Ar ions, resulting in formation of a large tensile stress [15.67]. The influence of surface morphology and thickness of the gold coating of the cantilever was discussed in [15.68, 69]. A multiple-point deflection technique has been used to investigate stress evolution during the adsorption of dodecanethiol on microcantilever sensors, allowing one to assess the cantilever bending profile [15.70]. Using gold-coated, piezoelectric-excited, millimeter-sized cantilevers exposed to 1-hexadecanethiol (HDT) in ethanol, a detection range between 1 fM to 1 mM has been claimed [15.71]. The formation of alkanedithiol (HS-(CH)SH) monolayers on gold in solution has been monitored using microcantilever sensors [15.72]. The nanomechanical bending of microfabricated cantilevers during the immobilization of alkanethiols of different chain lengths has been investigated in the liquid phase [15.73].

Metal Vapors

Detection of mercury vapor was one of the first applications of microcantilever sensors in dynamic mode [15.10]. 20 ppb of Hg vapor was detected using a microcantilever with an integrated piezoelectric film [15.74]. A monolayer of 1,6-hexanedithiol has been identified as an unusually specific recognition agent for CH_3Hg^+ [15.75]. A 15 ppb detection limit for mercury was reported using microcantilevers that were thermally excited at the fundamental and first three higher order modes [15.76, 77]. Cs ion concentrations in the range of 10^{-11}–10^{-7} M were detected using a 1,3-alternate 25,27-bis(11-mercapto-1-undecanoxy)-26,28-calix[4]benzo-crown-6 caesium recognition agent bound to a gold-coated microcantilever [15.78]. The crown cavity is highly selective to Cs, as compared to K or Na. An atomic force microscope cantilever has been used as a bending-beam

sensor to measure surface stress changes which occur during electrochemical processes, such as the formation of a Pb layer on Au [15.79].

HF and HCN

Microcantilevers have been used as a gas sensor to detect hydrofluoric acid (HF) at a threshold of 0.2 ppm [15.80]. Femtomolar HF concentrations, which are also a decomposition component of nerve agents, were detected using an SiO_2 microcantilever. The high sensitivity was considered to be due to the reaction of HF with SiO_2 [15.81]. The etching rate was determined to be 0.05 nm/min for SiO_2 and 0.7 nm/min for Si_3N_4 [15.82]. Sensor responses towards HCN at a concentration of 150 ppm within seconds were reported [15.83].

Ion Sensing

Using microcantilevers coated with a self-assembled monolayer of triethyl-12-mercaptododecylammonium bromide on gold CrO_4^{2-} ions were detected at a concentration of 10^{-9} M. Other anions, such as Cl^-, Br^-, CO_3^{2-}, HCO_3^-, and SO_4^{2-} did not deflect such modified cantilevers significantly [15.84]. Hg^{2+} has been measured down to a concentration of 10^{-11} M using a microcantilever coated with gold. Almost no affinity to other cations exists, such as K^+, Na^+ Pb^{2+}, Zn^{2+}, Ni^{2+}, Cd^{2+}, Cu^{2+}, and Ca^{2+} [15.85]. Adsorption characteristics of Ca^{2+} ions as a function of concentration in aqueous $CaCl_2$ solution were investigated in static and dynamic mode [15.86]. Microcantilevers functionalized with metal-binding protein, AgNt84-6, have been shown to detect heavy metal ions like Hg^{2+} and Zn^{2+}, but were insensitive to Mn^{2+} [15.87]. Hydrogels containing benzo-18-crown-6 have been used to modify microcantilevers for measurements of the concentration of Pb^{2+} in aqueous solutions [15.88]. Using different thiolated ligands as self-assembled monolayers (SAMs) functionalized on silicon microcantilevers (MCs) coated with gold allows one to detect Cs^+, Co^{2+}, and Fe^{3+} [15.89]. A gold-coated microcantilever has been utilized as the working electrode to detect Cr(VI) [15.90]. Others used 11-undecenyltriethylammonium bromide [15.91] or sol-gel layers [15.92] for detection of Cr(VI). On the basis of the EDTA-Cd(II) complex and its binding capability to bovine serum albumine (BSA) an antibody-based Cd(II) sensor using microcantilevers has been presented [15.93].

Volatile Organic Compounds (VOC)

A microcantilever-based alcohol vapor sensor was described in [15.94]. It used the piezoresistive technique and a polymer coating. The authors also presented a simple evaporation model that allowed determination of concentration. The detection limit found was 10 ppm for methanol, ethanol, and 2-propanol. In [15.95], an integrated complementary metal oxide semiconductor (CMOS) chemical microsensor with piezoresistive detection (Wheatstone bridge configuration) using poly(etherurethane) (PEUT) as the sensor layer was presented. The authors were able to reversibly detect volatile organic compounds (VOCs) such as toluene, *n*-octane, ethyl acetate, and ethanol with a sensitivity level down to 200 ppm. An improved version of that device was described in [15.96]. The sensitivity was increased to 5 ppm for *n*-octane. Later, the technique was refined using electromagnetic rather than electrothermal actuation and a transistor-based readout reducing power dissipation on the cantilever [15.97]. Piezoelectric readout in dynamic mode and electromagnetic actuation of cantilevers spray-coated with PEUT was reported in [15.98], which achieved a sensitivity of 14 ppm for ethanol. In [15.99], a study on how to prepare polyethylene glycol (PEG)-coated microcantilever sensors using a microcapillary pipette-assisted method was reported. The PEG coating is suitable for ethanol sensing as ethanol quickly forms hydrogen bonds with the −OH groups of the PEG. Sensor operation was reported to be reversible and reproducible. In [15.100], artificial neural networks were used for analyte species and concentration identification with polymer-coated optically read-out microcantilevers. The analytes detected were carbon dioxide, dichloromethane, diisopropylmethylphosphonate (DIMP), dioxane, ethanol, water, 2-propanol, methanol, trichloroethylene, and trichloromethylene. In [15.101], the chemical sensing performance of a silicon resonant microcantilever sensor was investigated in dependence on the thickness of the sensitive coating. For a coating thickness of 1, 4, and 21 μm of PEUT a limit of detection of 30 ppm was found for ethanol. A new concept of a parylene micromembrane array for chemical sensing was presented in [15.102] using the capacitive method. The parylene membrane was suspended over a metal pad patterned on the substrate. The pad and part of the membrane that was metal-coated served as electrodes for capacitive measurement. The top electrode located on the membrane was chemically modified by application of a gold layer and self-assembled thiol monolayers (−COOH, $-CH_3$ and −OH) for detection of analyte molecules. Successful detection of 2-propanol and toluene was reported. In [15.103], a sensitive self-oscillating cantilever array was described for quantitative and qualitative analysis of organic vapor mixtures. The cantilevers were electromagnetically actuated and the resonance frequency was measured using a frequency counter. Sensor response was reproducible and reversible. Using

a PEUT coating the smallest measured concentration was 400 ppm, but the limit of detection was well below 1 ppm. In [15.104], a combination of gas chromatography with a microcantilever sensor array for enhanced selectivity was reported. Test VOC mixtures composed of acetone, ethanol, and trichloroethylene in pentane, as well as methanol with acetonitrile in pentane were first separated in a gas chromatography column and then detected using microcantilevers coated with responsive phases such as 3-aminopropyltriethoxy silane, copper phtalocyanine, and methyl-β-cyclodextrin. Analytes detected included pentane, methanol, acetonitrile, acetone, ethanol, and trichloroethylene. In [15.105], results were presented on independent component analysis (ICA) of ethanol, propanol, and DIMP using a cantilever coated with molecular recognition phases (MRP), whereby ICA proved its feature extraction ability for components in mixtures.

Toxins

Detection of the organochlorine insecticide compound dichlorodiphenyltrichloroethane (DDT) has been reported using a synthetic hapten of the pesticide as a recognition site conjugated with bovine serum albumin (BSA) covalently immobilized on the gold-coated side of the cantilever by using thiol self-assembled monolayers [15.106].

Explosives, Chemical Warfare, and Biohazards

Security measures require inexpensive, highly selective and very sensitive small sensors that can be mass-produced and microfabricated. Such low-cost sensors could be arranged as a sensor grid for large area coverage of sensitive infrastructure, such as airports, public buildings, or traffic infrastructure. Threats may be of chemical, biological, radioactive, or explosive nature. Microcantilever sensors are reported to offer very high sensitivities for explosives detection. Photomechanical chemical microsensors based on adsorption-induced and photo-induced stress changes due to the presence of diisopropyl methyl phosphonate (DIMP), which is a model compound for phosphorous-containing chemical warfare agents, and trinitrotoluene (TNT), an explosive, were reported [15.107]. Further explosives frequently used include pentaerythritol tetranitrate (PETN) and hexahydro-1,3,5-triazine (RDX), often also with plastic fillers [15.108]. These compounds are very stable, if no detonator is present. Their explosive power, however, is very large, and moreover, the vapor pressures of PETN and RDX are very low, in the range of ppb and ppt. By functionalizing microcantilevers with self-assembled monolayers of 4-mercaptobenzoic acid (4-MBA) PETN was detected at a level of 1400 ppt and RDX at a level of 290 ppt [15.109]. TNT was found to readily stick to Si surfaces, suggesting the use of microcantilevers for TNT detection, taking advantage of the respective adsorption/desorption kinetics [15.110, 111]. Detection of TNT via deflagration on a microcantilever was described by *Pinnaduwage* et al. [15.111]. They used piezoresistive microcantilevers where the cantilever deflection was measured optically via beam deflection. TNT vapor from a generator placed 5 mm away from the microcantilever was observed to adsorb on its surface resulting in a decrease of resonance frequency. Application of an electrical pulse (10 V, 10 ms) to the piezoresistive cantilever resulted in deflagration of the TNT vapor and a bump in the cantilever bending signal. This bump was found to be related to the heat produced during deflagration. The amount of heat released is proportional to the area of the bump in the time versus bending signal diagram of the process. The deflagration was found to be complete, as the same resonance frequency as before the experiment was observed. The amount of TNT mass involved was determined as 50 pg. The technique was later extended to the detection of PETN and RDX, where much slower reaction kinetics was observed [15.109, 112]. Traces of 2,4-dinitrotoluene (DNT) in TNT can also be used for detection of TNT, because it is the major impurity in production-grade TNT. Furthermore, DNT is a decomposition product of TNT. The saturation concentration of DNT in air at 20 °C is 25 times higher than that of TNT. DNT was reported detectable at the 300 ppt level using polysiloxane polymer layers [15.113]. Microfabrication of electrostatically actuated resonant microcantilever beams in CMOS technology for detection of the nerve agent stimulant dimethylmethylphosphonate (DMMP) using polycarbosilane-coated beams [15.114] is an important step towards an integrated platform based on silicon microcantilevers, which besides compactness might also include telemetry [15.115]. Cu^{2+}/l-cysteine bilayer-coated microcantilever demonstrated high sensitivity and selectivity toward organo-phosphorus compounds in aqueous solution. The microcantilever was found to undergo bending upon exposure to nerve agent simulant DMMP at concentrations as low as 10^{-15} M due to the complexation of the phosphonyl group and the Cu^{2+}/l-cysteine bilayer on the microcantilever surface [15.116, 117].

15.7.2 Biochemical Environment

pH

Control of pH is often important in biochemical reactions. Therefore, this section concerns measurement of pH using microcantilevers. The interfacial stress of self-assembled monolayers of mercaptohexadecanoic

acid and hexadecanethiol depends on pH values and ionic strength [15.49]. SiO_2 and silicon nitride microcantilevers were also found to exhibit a deflection dependency with pH when coated with 4-aminobutyltriethoxysilane, 11-mercaptoundecanoic acid and Au/Al-coated over a pH range between 2 and 12. Aminosilane-modified SiO_2/Au cantilevers performed robustly over a pH range between 2 and 8 (49 nm deflection/pH unit), while Si_3N_4/Au cantilevers performed well at a pH between 2 and 6 as well as between 8 and 12 (30 nm deflection/pH unit) [15.118]. Microcantilevers with poly(methacrylic acid) (PMAA) and poly(ethylene glycol) dimethacrylate coating were shown to be sensitive to pH changes [15.119]. Also hydrogel coatings were found to be sensitive to pH [15.120]. The dependence of the micromechanical responses to different ionic strength and ion species present in the aqueous environment was discussed in [15.121], highlighting the critical role of counter- and co-ions on surface stress.

Glucose

Glucose sensing via microcantilevers is achieved by coating the cantilevers with the enzyme glucose oxidase on gold [15.122] or via polyethyleneimine (PEI) conjugation [15.123]. Glucose concentrations between 0.2 and 20 mM could be detected [15.124]. In another study a detection range between 2 and 50 mM was reported for glucose. No signal was observed for fructose, mannose, and galactose [15.125].

Hydrogen Peroxide (H_2O_2)

Hydrogen peroxide is detected at the nM level using multilayer modified microcantilevers functionalized through a layer-by-layer nanoassembly technique via intercalation of the enzyme horseradish peroxidase. The magnitudes of bending were found to be proportional to the concentrations of hydrogen peroxide [15.126].

DNA, RNA

Specific DNA hybridization detection was observed via surface stress changes related to transduction of receptor-ligand binding into a direct nanomechanical response of microfabricated cantilevers without the need for external labeling or amplification. The differential deflection of the cantilevers was found to provide a true molecular recognition signal despite large responses of individual cantilevers. Hybridization of complementary oligonucleotides shows that a single base mismatch between two 12-mer oligonucleotides is clearly detectable [15.17]. The findings were confirmed or modeled by several groups [15.127, 128]. Hybridization in a complex nonspecific background was observed in a complement concentration range between 75 nM and 2 μM [15.50], following Langmuir model kinetics [15.129]. Enzymatic processes were directly performed on a microcantilever functionalized with DNA incorporating a Hind III restriction endonuclease site, followed by digestion with Hind III to produce DNA comprising a single-stranded end on the cantilever surface. Ligase was used to couple a second DNA molecule with a compatible end to the DNA on the cantilever [15.130]. Using gold nanoparticle-labeled DNA, microcantilevers have been used to detect DNA strands with a specific sequence in dynamic mode, whereby a concentration of 23 pM could still be detected, as well as a single base pair mismatch [15.131]. Whereby adsorption of thiol-functionalized single-stranded DNA is easily observed, hybridization cannot be observed if long hydrocarbon spacer molecules between single-strand DNA and the thiol anchor are used [15.132]. DNA hybridization was also observed using piezoresistive cantilevers [15.129, 133]. A different technique to read out the microcantilever deflections in an array was reported in [15.134]. There the optical beam deflection technique was combined with the scanning of a laser beam illuminating the cantilevers of an array sequentially. DNA hybridization has also been reported using polymer SU-8 cantilevers [15.135]. *Mukhopadhyay* et al. reported 20 nM hybridization sensitivity using piezoresistive cantilevers and DNA sequences with an overhang extension distal to the surface [15.136]. A larger array comprising 20 microcantilevers was described in [15.137]. Moreover, the authors presented integration of the array with microfluidics. Surface stress changes in response to thermal dehybridization, or melting, has been reported [15.138]. The dependence of salt concentration and hybridization efficiency was discussed in [15.139]. Two different DNA-binding proteins, the transcription factors SP1 and NF-kappa B were investigated in [15.140]. Phase transition and stability issues of DNA were discussed in [15.141]. Differential gene expression of the gene 1-8U, a potential marker for cancer progression or viral infections, has been observed in a complex background. The measurements provided results within minutes at the picomolar level without target amplification, and were sensitive to base mismatches [15.142].

Proteins and Peptides

Microfabricated cantilevers were utilized to detect adsorption of low-density lipoproteins (LDL) and their oxidized form (oxLDL) on heparin, and to detect adsorption of bovine serum albumine and Immunoglobuline G (IgG) [15.143]. In [15.144] the activity, stability, lifetime, and re-usability of monoclonal antibodies to myoglobin covalently immobilized onto microfabricated cantilever surfaces was investigated. Using

piezoresistive microcantilevers the interaction of anti-bovine serum albumin (a-BSA) with bovine serum albumin (BSA) was studied [15.145]. Continuous label-free detection of two cardiac biomarker proteins (creatin kinase and myoglobin) has been demonstrated using an array of microfabricated cantilevers functionalized with covalently anchored anti-creatin kinase and anti-myoglobin antibodies [15.51]. Label-free protein detection has been reported using a microcantilever functionalized with DNA aptamer receptors for Taq DNA polymerase [15.146]. Label-free detection of C-reactive protein (CRP) using resonant frequency shift in piezoresistive cantilevers was described in [15.147], utilizing the specific binding characteristics of CRP antigen to its antibody, which is immobilized with Calixcrown SAMs on Au. In [15.148], the authors discuss receptors on microcantilevers for serotonin that are insensitive to its biological precursor with a similar tryptophan structure. Using single-chain fragment antibodies instead of complete antibodies allowed the limit of detection to be lowered to concentrations of about 1 nM [15.149]. In [15.150], detection of prostate-specific antigen (PSA) and C-reactive protein was reported. Detection of the human oestrogen receptor in free and oestradiol-bound conformation can be distinguished [15.151]. The Ca^{2+} binding protein calmodulin changes its conformation in presence or absence of Ca^{2+} resulting in a microcantilever deflection change [15.152]. No effect was observed upon exposure to K^{+} and Mg^{2+}. Detection of activated cyclic adenosine monophosphate (cyclic AMP)-dependent protein kinase (PKA) was performed in dynamic mode employing a peptide derived from the heat-stable protein kinase inhibitor (PKI) in [15.153]. Detection of streptavidin at 1−10 nM concentration was reported using biotin-coated cantilevers in [15.154]. Using GST (glutathione-S-transferase) for detection of GST antibodies, a sensitivity of 40 nM was obtained [15.155]. A two-dimensional multiplexed real-time, label-free antibody-antigen binding assay by optically detecting nanoscale motions of two-dimensional arrays of microcantilever beams was presented in [15.156]. Prostate-specific antigen was detected at 1 ng/mL using antibodies covalently bound to one surface of the cantilevers. Conformational changes in membrane protein patches of bacteriorhodopsin proteoliposomes were observed with microcantilevers through prosthetic retinal removal (bleaching) [15.157]. Using an analog of the myc-tag decapeptide, binding of anti-myc-tag antibodies was reported [15.158].

Lipid Bilayers, Liposomes, Cells

Cantilever array sensors can sense the formation by vesicle fusion of supported phospholipid bilayers of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) on their surface and can monitor changes in mechanical properties of lipid bilayers [15.159]. Liposomes are detected based on their interaction with protein C2A which recognizes the phosphatidylserine (PS) exposed on the surface of the liposome [15.160]. Individual *Escherichia coli* (*E. coli*) O157:H7 cell-antibody binding events using microcantilevers operated in dynamic mode have been reported [15.161]. The contractile force of self-organized cardiomyocytes was measured on biocompatible poly(dimethylsiloxane) cantilevers, representing a microscale cell-driven motor system [15.162]. Resonating cantilevers were used to detect individual phospholipid vesicle adsorption in liquid. A resonance frequency shift corresponding to an added mass of 450 pg has been measured [15.163].

Spores, Bacteria, and Viruses

Micromechanical cantilever arrays have been used for quantitative detection of vital fungal spores of *Aspergillus niger* and *Saccharomyces cerevisiae*. The specific adsorption and growth on concanavalin A, fibronectin, or immunoglobulin G cantilever surfaces was investigated. Maximum spore immobilization, germination, and mycelium growth was observed on the immunoglobulin G-functionalized cantilever surfaces, as measured from shifts in resonance frequency within a few hours, being much faster than with standard petri dish cultivation [15.164]. Short peptide ligands can be used to efficiently capture *Bacillus subtilis* (a simulant of *Bacillus anthracis*) spores in liquids. Fifth-mode resonant frequency measurements were performed before and after dipping microcantilever arrays into a static *B. subtilis* solution showing a substantial decrease in frequency for binding-peptide-coated microcantilevers as compared to that for control peptide cantilevers [15.165].

Medical

A bioassay of prostate-specific antigen (PSA) using microcantilevers has been presented in [15.166], covering a wide range of concentrations from 0.2 ng/ml to 60 µg/ml in a background of human serum albumin (HSA). Detection has been confirmed by another group using microcantilevers in resonant mode [15.167, 168]. The feasibility of detecting severe acute respiratory syndrome-associated coronavirus (SARS-CoV) using microcantilever technology was studied in [15.169] by showing that the feline coronavirus (FIP) type I virus can be detected by a microcantilever modified by feline coronavirus (FIP) type I antiviral antiserum. A method for quantification of a prostate cancer biomarker in urine without sample preparation using monoclonal antibodies was described in [15.170].

15.7.3 Microcantilever Sensors to Measure Physical Properties

Besides chemical and biochemical sensing, microcantilevers can also detect changes in physical properties of surrounding media, such as gas or liquid, or of layers deposited on the cantilever itself.

Density and Viscosity

A piezoelectric unimorph cantilever as a liquid viscosity-and-density sensor was tested using water-glycerol solutions of different compositions, whereby the resonance frequency decreased while the width of the resonance peak increased with increasing glycerol content [15.171]. The viscosity of complex organic liquids with non-Newtonian behavior was studied in [15.172] over a wide range from 10 to 500 mm^2/s. Simultaneous determination of density and viscosity of water-ethanol mixtures based on resonance curves of microcantilevers was reported in [15.173]. A detailed theoretical study of viscoelastic effects on the frequency shift of microcantilever chemical sensors was presented in [15.174]. Microcantilever deflection as a function of flow speed of viscous fluids was investigated in [15.175]. Viscosity of sugar solutions has been tested using microcantilevers [15.176].

Gas and Flow Sensing

Gas sensing not only involves chemical detection, but also pressure and flow sensing. *Brown* et al. [15.177] studied the behavior of magnetically actuated oscillating microcantilevers at large deflections and found hysteresis behavior at resonance. The amplitude at the actuation frequency changes, depending on pressure due to damping. The authors used cantilever-in-cantilever (CIC) structures, and observed changes in deflection as gas pressure was varied. At atmospheric pressure, damping is large and the oscillation amplitude is relative small and hysteresis effects are absent. At lower pressure, abrupt changes in the oscillation amplitude occur with changes in the driving frequency. Since the change of amplitude and driving frequency, at which they occur is pressure dependent, these quantities can be used for accurate determination of gas pressure, demonstrated in the range between 10^{-3} and 10^2 mbar. *Brown* et al. [15.178] emphasized that microelectromechanical system pressure sensors will have a wide range of applications, especially in the automotive industry. Piezoresistive-cantilever-based deflection measurement has major advantages over diaphragms. The pressure range has been extended to 15−1450 mbar by means of design geometry adaptation. *Su* et al. [15.179] presented highly sensitive ultrathin piezoresistive silicon microcantilevers for gas velocity sensing, whereby the deflection increased with airflow distribution in a steel pipe. The detection principle was based on normal pressure drag producing bending of the cantilever. The minimum flow speed measured was 0.07 m/s, which is comparable to classical hot-wire anemometers. *Mertens* et al. [15.180] have investigated the effects of temperature and pressure on microcantilever resonance response in helium and nitrogen. Resonance response as a function of pressure showed three different regimes, which correspond to molecular flow, transition regimes and viscous flow, whereby the frequency variation of the cantilever is mainly due to changes in the mean free path of gas molecules. Effects observed allow measurement of pressures between 10^{-2}−10^6 Pa. *Mortet* et al. [15.181] presented a pressure sensor based on a piezoelectric bimorph microcantilever with a measurement range between 0.1 and 8.5 bar. The resonance frequency shift was constant for pressures below 0.5 bar. For higher pressures the sensitivity was typically a few ppm/mbar, but depended on the mode number. *Sievilä* et al. [15.182] presented a cantilever paddle within a frame that operated like a moving mirror to detect displacements in the oscillating cantilever using a He/Ne laser in a Michelson interferometer configuration, whereby the cantilever acted as a moving mirror element in one path of the interferometer. A fixed mirror served as a reference in the other arm of the interferometer.

Thermal Expansion

The thermal expansion of TaO_xN_y thin films deposited on a microcantilever was measured to examine the residual stress and the thermal expansion coefficient by observing changes in radius of curvature [15.183]. Thermal drift issues of resonating microcantilevers were discussed in detail in [15.184].

Infrared

Microcantilevers can also be used as uncooled, microcantilever-based infrared (IR) imaging devices by monitoring the bending of the microcantilever as a function of the IR radiation intensity incident on the cantilever surface. The infrared (thermal) image of the source is obtained by rastering a single microfabricated cantilever across the image formed at the focal plane of a concave mirror [15.185–187]. The method was later refined such that photons are detected using the stress caused by photoelectrons emitted from a Pt film surface in contact with a semiconductor microstructure, which forms a Schottky barrier. The photo-induced bending of the Schottky barrier microstructure is due to electronic stress produced by photoelectrons diffusing into the microstructure [15.188]. The performance of IR imaging via microcantilevers has been enhanced by one-fold

leg and two-fold leg beam structures with absorber plates [15.189–191].

15.7.4 Recent Developments

More than two decades after the first reports on the use of cantilevers as sensors the field is still active with a substantial number of research groups working and publishing. A new trend is the integration of novel materials as sensing layers or the direct use of them as nanomechanical structures. Mass sensing in ultrahigh vacuum down to the yoctogram (10^{-24} g) range by employing a frequency-modulation mixing technique for the readout of a carbon nanotube resonator has recently been demonstrated by *Chaste* et al. [15.192]. Electrochemistry is reported to be a way to control the surface of a microcantilever biosensor [15.193], as the topography of the functionalized cantilever surface seems to influence the efficiency of detection [15.194]. In biochemical measurements the work by *Mukhopadhyay* et al. [15.195] represents major progress in reliable recognition of DNA by using modified versions of DNA like PNA (peptide nucleic acid) and LNA (locked nuclei acid) having different sugar backbones. Especially for detection of single base mismatches PNA and LNA produce more reliable results [15.196]. The effects of hydration might play an important role in DNA hybridization [15.197]. *Ndieyira* et al. [15.198] report on nanomechanical detection of antibiotic resistance in bacteria by observing antibiotic-mucopeptide binding, even in blood serum [15.199]. Significant progress in high-throughput screening with cantilever sensors combined with a DVD-ROM-based optical readout technique has been reported by the Boisen group [15.200]. Detection of femtomolar concentrations of oligonucleotides in a complex background of total cellular RNA extracts from cell lines without labeling or amplification has been achieved by *Mishra* et al. [15.201] giving the perspective of RNA pharmacokinetics, nucleic acid biomarker assays, and miRNA quantification for disease monitoring and diagnosis. *Huber* et al. [15.202, 203] demonstrated application of cantilever sensor arrays for detection of melanoma, based on the presence of a predominant mutation in the BRAF gene. Detection of this single-point mutation allows one to screen melanoma patients for eligibility for treatment with a specific drug which only works if this point mutation is present. Detection and monitoring of *Aspergillus niger* growth directly on an oscillating cantilever using the first two flexural resonance modes has been observed in [15.204]. Even single spore growth has been observed within 10 h and is reported in [15.205]. A very ingenious way to detect motility of bacteria has been described by *Longo* et al. [15.206]. They show that fluctuations (noise) in the *dc* signal of cantilevers onto which low numbers of bacteria have been attached, characterize their metabolism. The technique quantitatively screens (within minutes) the response of bacteria such as *Escherichia coli* and *Staphylococcus aureus* to antibiotics. Living bacteria produced larger cantilever fluctuations than bacteria exposed to antibiotics [15.207]. A major contribution to the problem of chemical sensitivity of cantilever sensors has been published by the Thundat group. By using photothermal spectroscopy molecules adsorbing on a bimetallic cantilever can be qualitatively identified [15.208]. The technique is even capable of distinguishing both anti and gauche conformers of EtOH based on wavelength-dependent cantilever deflections [15.209]. Deeper theoretical understanding of viscoelastic phenomena in vibrating nanostructures is described in [15.210].

15.8 Conclusions and Outlook

Cantilever-sensor array techniques have turned out to be a very powerful and highly sensitive tool to study physisorption and chemisorption processes, as well as to determine material-specific properties such as heat transfer during phase transitions. Experiments in liquids have provided new insights into such complex biochemical reactions as the hybridization of DNA or molecular recognition in antibody-antigen systems or proteomics.

Future developments must go towards technological applications, in particular to find new ways to characterize real-world samples such as clinical samples. The development of medical diagnosis tools requires an improvement of the sensitivity of a large number of genetic tests to be performed with small amounts of single donor-blood or body-fluid samples at low cost. From a scientific point of view, the challenge lies in optimizing cantilever sensors to improve their sensitivity to the ultimate limit: the detection of individual molecules.

Several fundamentally new concepts in microcantilever sensing are described in the recent literature, and these could help to achieve these goals: the issue of low quality factor of resonating microcantilevers in liquid has been elegantly solved by fabrication of a hollow cantilever that can be filled with biochemical liquids. Confining the fluid to the inside of a hollow cantilever also allows direct integration with conventional microfluidic systems, and significantly increases sensitivity by eliminating high damping and viscous drag

[15.211]. Biochemical selectivity can be enhanced by using enantioselective receptors [15.212]. Other shapes for micromechanical sensors like microspirals could be advantageous for biochemical detection [15.213]. Miniaturization of microcantilevers down to *true* nanometric dimensions, for example by using single-wall carbon nanotubes [15.214] or graphene sheets [15.215] will further increase sensitivity. Reports on optomechanical coupling to graphene-based oscillators via microwaves are already available [15.216]. Single-layer graphene films have also been applied as pressure sensors [15.217].

What is still lacking after two decades of development is the ease of use of cantilever array sensors. Complicated alignment and calibration procedures should be avoided and preparatory steps before measurement should be automatized. Functionalization and measurement protocols must be standardized, so that every user gets the same results when doing the same experiment. Array-to-array variability must be reduced and evaluation of results must be done in a transparent and reproducible way. We should still keep in mind that we are measuring with cantilever sensors an indirect mechanically transduced response and not a primary physical quantity. Nevertheless, nanomechanical cantilever array sensors offer possibilities to monitor molecular processes that are inaccessible or barely measurable using traditional techniques.

Acknowledgments. We thank R. McKendry (University College London, London, U.K.), J. Zhang, A. Bietsch, V. Barwich, M. Ghatkesar, T. Braun, F. Huber, N. Backmann, J.-P. Ramseyer, P. Oliva, H.R. Hidber, E. Meyer, and the late H.-J. Güntherodt (University of Basel, Basel, Switzerland) for valuable contributions and discussions, as well as U. Drechsler, M. Despont, H. Schmid, E. Delamarche, H. Wolf, R. Stutz, R. Allenspach, and P.F. Seidler (IBM Research, Zurich Research Laboratory, Rüschlikon, Switzerland). We are grateful for MSS collaboration with G. Yoshikawa, T. Akiyama, F. Loizeau, S. Gautsch, P. Vettiger, and Nico de Rooij. We acknowledge the continuous technical support from A. Tonin, R. Maffiolini (electronics workshop), H. Breitenstein, S. Martin and his crew (mechanical workshop) in the Department of Physics, University of Basel. We thank the Swiss Nanoscience Institute (SNI), NanoTera Program (projects 853_406, 20NA21_150957 and MSSdevices), the Cleven Foundation, and the Swiss National Science Foundation for financial support.

References

15.1 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56**, 930–933 (1986)
15.2 R.J. Wilfinger, P.H. Bardell, D.S. Chhabra: The Resonistor, a frequency sensitive device utilizing the mechanical resonance of a silicon substrate, IBM J. Res. Div. **12**, 113–118 (1968)
15.3 T.M.S. Heng: Trimming of microstrip circuits utilizing microcantilever air gaps, IEEE Trans. Microw. Theory Technol. **19**, 652–654 (1971)
15.4 K.E. Petersen: Micromechanical membrane switches on silicon, IBM J. Res. Develop. **23**, 376–385 (1979)
15.5 E.S. Kolesar: Electronic Nerve Agent Detector, United States Patent 4549427 (1983), filed Sep. 19
15.6 T. Itoh, T. Suga: Force sensing microcantilever using sputtered zinc-oxide thin-film, Appl. Phys. Lett. **64**, 37–39 (1994)
15.7 J.P. Cleveland, S. Manne, D. Bocek, P.K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64**, 403–405 (1993)
15.8 T. Thundat, R.J. Warmack, G.Y. Chen, D.P. Allison: Thermal and ambient-induced deflections of scanning force microscope cantilevers, Appl. Phys. Lett. **64**, 2894–2896 (1994)
15.9 J.K. Gimzewski, Ch. Gerber, E. Meyer, R.R. Schlittler: Observation of a chemical reaction using a micromechanical sensor, Chem. Phys. Lett. **217**, 589–594 (1994)
15.10 T. Thundat, G.Y. Chen, R.J. Warmack, D.P. Allison, E.A. Wachter: Vapor detection using resonating microcantilevers, Anal. Chem. **67**, 519–521 (1995)
15.11 H.P. Lang, R. Berger, C. Andreoli, J. Brugger, M. Despont, P. Vettiger, Ch. Gerber, J.K. Gimzewski, J.-P. Ramseyer, E. Meyer, H.-J. Güntherodt: Sequential position readout from arrays of micromechanical cantilever sensors, Appl. Phys. Lett. **72**, 383–385 (1998)
15.12 H. Ibach: Adsorbate-induced surface stress, J. Vac. Sci. Technol. A **12**, 2240–2245 (1994)
15.13 R. Berger, E. Delamarche, H.P. Lang, C. Gerber, J.K. Gimzewski, E. Meyer, H.-J. Güntherodt: Surface stress in the self-assembly of alkanethiols on gold, Science **276**, 2021–2024 (1997)
15.14 J.R. Barnes, R.J. Stephenson, M.E. Welland, C. Gerber, J.K. Gimzewski: Photothermal spectroscopy with femtojoule sensitivity based on micromechanics, Nature **372**, 79–81 (1994)
15.15 G.G. Stoney: The tension of metallic films deposited by electrolysis, Proc. R. Soc. London, Ser. A **82**, 172–177 (1909)
15.16 F.J. von Preissig: Applicability of the classical curvature-stress relation for thin films on plate substrates, J. Appl. Phys. **66**, 4262–4268 (1989)
15.17 J. Fritz, M.K. Baller, H.P. Lang, H. Rothuizen, P. Vettiger, E. Meyer, H.-J. Güntherodt, C. Gerber,

J.K. Gimzewski: Translating biomolecular recognition into nanomechanics, Science **288**, 316–318 (2000)

15.18 M. Alvarez, L.G. Carrascosa, M. Moreno, A. Calle, A. Zaballos, L.M. Lechuga, C. Martinez-A, J. Tamayo: Nanomechanics of the formation of DNA self-assembled monolayers and hybridization on microcantilevers, Langmuir **20**, 9663–9668 (2004)

15.19 D. Sarid: *Scanning Force Microscopy with Applications to Electric, Magnetic, and Atomic Forces* (Oxford Univ. Press, New York 1991)

15.20 R. Berger, H.P. Lang, C. Gerber, J.K. Gimzewski, J.H. Fabian, L. Scandella, E. Meyer, H.-J. Güntherodt: Micromechanical thermogravimetry, Chem. Phys. Lett. **294**, 363–369 (1998)

15.21 J.E. Sader: Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope, J. Appl. Phys. **84**, 64–76 (1998)

15.22 T. Bachels, F. Tiefenbacher, R. Schafer: Condensation of isolated metal clusters studied with a calorimeter, J. Chem. Phys. **110**, 10008–10015 (1999)

15.23 J.R. Barnes, R.J. Stephenson, C.N. Woodburn, S.J. O'Shea, M.E. Welland, T. Rayment, J.K. Gimzewski, C. Gerber: A femtojoule calorimeter using micromechanical sensors, Rev. Sci. Instrum. **65**, 3793–3798 (1994)

15.24 T.A. Brunt, T. Rayment, S.J. O'Shea, M.E. Welland: Measuring the surface stresses in an electrochemically deposited monolayer: Pb on Au(111), Langmuir **12**, 5942–5946 (1996)

15.25 R. Puers, D. Lapadatu: Electrostatic forces and their effects on capacitive mechanical sensors, Sens. Actuat. A **56**, 203–210 (1996)

15.26 J. Fricke, C. Obermaier: Cantilever beam accelerometer based on surface micromachining technology, J. Micromech. Microeng. **3**, 190–192 (1993)

15.27 C. Lee, T. Itoh, T. Ohashi, R. Maeda, T. Suga: Development of a piezoelectric self-excitation and self-detection mechanism in PZT microcantilevers for dynamic scanning force microscopy in liquid, J. Vac. Sci. Technol. B **15**, 1559–1563 (1997)

15.28 T. Göddenhenrich, H. Lemke, U. Hartmann, C. Heiden: Force microscope with capacitive displacement detection, J. Vac. Sci. Technol. A **8**, 383–387 (1990)

15.29 J. Brugger, R.A. Buser, N.F. de Rooij: Micromachined atomic force microprobe with integrated capacitive read-out, J. Micromech. Microeng. **2**, 218–220 (1992)

15.30 C. Schönenberger, S.F. Alvarado: A differential interferometer for force microscopy, Rev. Sci. Instrum. **60**, 3131–3134 (1989)

15.31 M.J. Cunningham, S.T. Cheng, W.W. Clegg: A differential interferometer for scanning force microscopy, Meas. Sci. Technol. **5**, 1254–1350 (1994)

15.32 G. Meyer, N.M. Amer: Novel optical approach to atomic force microscopy, Appl. Phys. Lett. **53**, 2400–2402 (1988)

15.33 G. Yoshikawa, T. Akiyama, S. Gautsch, P. Vettiger, H. Rohrer: Nanomechanical membrane-type surface stress sensor, Nano Lett **11**, 1044–1048 (2011)

15.34 F. Loizeau, T. Akiyama, S. Gautsch, P. Vettiger, G. Yoshikawa, N. de Rooij: Membrane-type surface stress sensor with piezoresistive readout, Procedia Eng **47**, 1085–1088 (2012)

15.35 F. Loizeau, H.P. Lang, T. Akiyama, S. Gautsch, S.P. Vettiger, A. Tonin, G. Yoshikawa, Ch. Gerber, N. de Rooij: Piezoresistive membrane-type surface stress sensors arranged in arrays for cancer diagnosis through breath analysis. In: *26th IEEE Int. Confe. Micro Electro Mech. Syst. (MEMS)* (2013) pp. 621–624

15.36 G. Yoshikawa, T. Akiyama, F. Loizeau, K. Shiba, S. Gautsch, T. Nakayamam, P. Vettiger, N. de Rooij, M. Aono: Two dimensional array of piezoresistive nanomechanical membrane-type surface stress sensor (MSS) with improved sensitivity, Sensors **12**, 15873–15887 (2012)

15.37 G. Yoshikawa: Nanomechanical sensors and membrane-type surface stress sensor (MSS) for medical, security and environmental applications. In: *Manipulation of Nanoscale Materials: An Introduction to Nanoarchitectonics, RSC Nanoscience and Nanotechnology*, ed. by K. Ariga (The Royal Society of Chemistry, Cambridge 2012) pp. 428–448

15.38 G. Yoshikawa, F. Loizeau, C.J.Y. Lee, T. Akiyama, K. Shiba, S. Gautsch, T. Nakayama, P. Vettiger, N.F. de Rooij, M. Aono: Double-side-coated nanomechanical membrane-type surface stress sensor (MSS) for one-chip-one-channel setup, Langmuir **29**, 7551–7556 (2013)

15.39 G. Yoshikawa, C.J.Y. Lee, K. Shiba: Effects of coating materials on two dimensional stress-induced deflection of nanomechanical sensors, J. Nanosci. Nanotechnol. **14**, 2908–2912 (2014)

15.40 F. Loizeau, T. Akiyama, S. Gautsch, P. Vettiger, G. Yoshikawa, N.F. de Rooij: Comparing membrane- and cantilever-based surface stress sensors for reproducibility, Sens. Act. A **228**, 9–15 (2015)

15.41 G. Imamura, K. Shiba, G. Yoshikawa: Finite element analysis on nanomechanical detection of small particles: Toward virus detection, Front. Microbiol. **7**, 488 (2016)

15.42 H.P. Lang, F. Loizeau, A. Hiou, J.-P. Rivals, P. Romero, T. Akiyama, C. Gerber, E. Meyer: Piezoresistive membrane surface stress sensors for characterization of breath samples of head & neck cancer patients. In: *Proc. 2nd Int. Electron. Conf. Sens. Appl.*, Sciforum Electronic Conference Series, Vol. 2 (2015)

15.43 R. Berger, C. Gerber, J.K. Gimzewski, E. Meyer, H.-J. Güntherodt: Thermal analysis using a micromechanical calorimeter, Appl. Phys. Lett. **69**, 40–42 (1996)

15.44 L. Scandella, G. Binder, T. Mezzacasa, J. Gobrecht, R. Berger, H.P. Lang, C. Gerber, J.K. Gimzewski, J.H. Koegler, J.C. Jansen: Combination of single crystal zeolites and microfabrication: Two appli-

cations towards zeolite nanodevices, Microporous Mesoporous Mater **21**, 403–409 (1998)
15.45 R. Berger, C. Gerber, H.P. Lang, J.K. Gimzewski: Micromechanics: a toolbox for femtoscale science: Towards a laboratory on a tip, Microelectron. Eng. **35**, 373–379 (1997)
15.46 H.P. Lang, R. Berger, F.M. Battiston, J.-P. Ramseyer, E. Meyer, C. Andreoli, J. Brugger, P. Vettiger, M. Despont, T. Mezzacasa, L. Scandella, H.-J. Güntherodt, C. Gerber, J.K. Gimzewski: A chemical sensor based on a micromechanical cantilever array for the identification of gases and vapors, Appl. Phys. A **66**, S61–S64 (1998)
15.47 M.K. Baller, H.P. Lang, J. Fritz, Ch. Gerber, J.K. Gimzewski, U. Drechsler, H. Rothuizen, M. Despont, P. Vettiger, F.M. Battiston, J.-P. Ramseyer, P. Fornaro, E. Meyer, H.-J. Güntherodt: A cantilever array based artificial nose, Ultramicroscopy **82**, 1–9 (2000)
15.48 S. Cesaro-Tadic, G. Dernick, D. Juncker, G. Buurman, H. Kropshofer, B. Michel, C. Fattinger, E. Delamarche: High-sensitivity miniaturized immunoassays for tumor necrosis factor α using microfluidic systems, Lab Chip **4**, 563–569 (2004)
15.49 J. Fritz, M.K. Baller, H.P. Lang, T. Strunz, E. Meyer, H.-J. Güntherodt, E. Delamarche, Ch. Gerber, J.K. Gimzewski: Stress at the solid-liquid interface of self-assembled monolayers on gold investigated with a nanomechanical sensor, Langmuir **16**, 9694–9696 (2000)
15.50 R. McKendry, J. Zhang, Y. Arntz, T. Strunz, M. Hegner, H.P. Lang, M.K. Baller, U. Certa, E. Meyer, H.-J. Güntherodt, C. Gerber: Multiple label-free biodetection and quantitative DNA-binding assays on a nanomechanical cantilever array, Proc. Nat. Acad. Sci. USA **99**, 9783–9787 (2002)
15.51 Y. Arntz, J.D. Seelig, H.P. Lang, J. Zhang, P. Hunziker, J.-P. Ramseyer, E. Meyer, M. Hegner, C. Gerber: Label-free protein assay based on a nanomechanical cantilever array, Nanotechnology **14**, 86–90 (2003)
15.52 A. Bietsch, M. Hegner, H.P. Lang, C. Gerber: Inkjet deposition of alkanethiolate monolayers and DNA oligonucleotides on gold: Evaluation of spot uniformity by wet etching, Langmuir **20**, 5119–5122 (2004)
15.53 A. Bietsch, J. Zhang, M. Hegner, H.P. Lang, C. Gerber: Rapid functionalization of cantilever array sensors by inkjet printing, Nanotechnology **15**, 873–880 (2004)
15.54 D. Lange, C. Hagleitner, A. Hierlemann, O. Brand, H. Baltes: Complementary metal oxide semiconductor cantilever arrays on a single chip: Mass-sensitive detection of volatile organic compounds, Anal. Chem. **74**, 3084–3085 (2002)
15.55 C.A. Savran, T.P. Burg, J. Fritz, S.R. Manalis: Microfabricated mechanical biosensor with inherently differential readout, Appl. Phys. Lett. **83**, 1659–1661 (2003)
15.56 S. Okuyama, Y. Mitobe, K. Okuyama, K. Matsushita: Hydrogen gas sensing using a Pd-coated cantilever, Jpn. J. Appl. Phys. **39**, 3584–3590 (2000)
15.57 A. Fabre, E. Finot, J. Demoment, S. Contreras: Monitoring the chemical changes in Pd induced by hydrogen absorption using microcantilevers, Ultramicroscopy **97**, 425–432 (2003)
15.58 D.R. Baselt, B. Fruhberger, E. Klaassen, S. Cemalovic, C.L. Britton, S.V. Patel, T.E. Mlsna, D. McCorkle, B. Warmack: Design and performance of a microcantilever-based hydrogen sensor, Sens. Act. B Chem. **88**, 120–131 (2003)
15.59 Y.-I. Chou, H.-C. Chiang, C.-C. Wang: Study on Pd functionalization of microcantilever for hydrogen detection promotion, Sens. Act. B Chem. **129**, 72–78 (2008)
15.60 L. Scandella, G. Binder, T. Mezzacasa, J. Gobrecht, R. Berger, H.P. Lang, C. Gerber, J.K. Gimzewski, J.H. Koegler, J.C. Jansen: Combination of single crystal zeolites and microfabrication: Two applications towards zeolite nanodevices, Microporous Mesoporous Mat **21**, 403–409 (1998)
15.61 R.L. Gunter, W.D. Delinger, T.L. Porter, R. Stewart, J. Reed: Hydration level monitoring using embedded piezoresistive microcantilever sensors, Med. Eng. Phys. **27**, 215–220 (2005)
15.62 X. Shi, Q. Chen, J. Fang, K. Varahramyan, H.F. Ji: Al_2O_3-coated microcantilevers for detection of moisture at ppm level, Sens. Act. B Chem. **129**, 241–245 (2008)
15.63 J. Zhou, P. Li, S. Zhang, Y.C. Long, F. Zhou, Y.P. Huang, P.Y. Yang, M.H. Bao: Zeolite-modified microcantilever gas sensor for indoor air quality control, Sens. Act. B Chem. **94**, 337–342 (2003)
15.64 A. Vidic, D. Then, C. Ziegler: A new cantilever system for gas and liquid sensing, Ultramicroscopy **97**, 407–416 (2003)
15.65 R. Berger, E. Delamarche, H.P. Lang, C. Gerber, J.K. Gimzewski, E. Meyer, H.J. Guntherodt: Surface stress in the self-assembly of alkanethiols on gold probed by a force microscopy technique, Appl. Phys. A Mater. **66**, S55–S59 (1998)
15.66 P.G. Datskos, I. Sauers: Detection of 2-mercaptoethanol using gold-coated micromachined cantilevers, Sens. Act. B Chem. **61**, 75–82 (1999)
15.67 A.N. Itakura, R. Berger, T. Narushima, M. Kitajima: Low-energy ion-induced tensile stress of self-assembled alkanethiol monolayers, Appl. Phys. Lett. **80**, 3712–3714 (2002)
15.68 M. Godin, P.J. Williams, V. Tabard-Cossa, O. Laroche, L.Y. Beaulieu, R.B. Lennox, P. Grutter: Surface stress, kinetics, and structure of alkanethiol self-assembled monolayers, Langmuir **20**, 7090–7096 (2004)
15.69 R. Desikan, I. Lee, T. Thundat: Effect of nanometer surface morphology on surface stress and adsorption kinetics of alkanethiol self-assembled monolayers, Ultramicroscopy **106**, 795–799 (2006)
15.70 S. Jeon, N. Jung, T. Thundat: Nanomechanics of a self-assembled monolayer on microcantilever sensors measured by a multiple-point deflection technique, Sens. Act. B Chem. **122**, 365–368 (2007)

15.71 K. Rijal, R. Mutharasan: Method for measuring the self-assembly of alkanethiols on gold at femtomolar concentrations, Langmuir **23**, 6856–6863 (2007)

15.72 S. Kohale, S.M. Molina, B.L. Weeks, R. Khare, L.J. Hope-Weeks: Monitoring the formation of self-assembled monolayers of alkanedithiols using a micromechanical cantilever sensor, Langmuir **23**, 1258–1263 (2007)

15.73 R. Desikan, S. Armel, H.M. Meyer, T. Thundat: Effect of chain length on nanomechanics of alkanethiol self-assembly, Nanotechnology **18**, 424028 (2007)

15.74 B. Rogers, L. Manning, M. Jones, T. Sulchek, K. Murray, B. Beneschott, J.D. Adams, Z. Hu, T. Thundat, H. Cavazos, S.C. Minne: Mercury vapor detection with a self-sensing, resonating piezoelectric cantilever, Rev. Sci. Instrum. **74**, 4899–4901 (2003)

15.75 H.F. Ji, Y.F. Zhang, V.V. Purushotham, S. Kondu, B. Ramachandran, T. Thundat, D.T. Haynie: 1,6-Hexanedithiol monolayer as a receptor for specific recognition of alkylmercury, Analyst **130**, 1577–1579 (2005)

15.76 A.R. Kadam, G.P. Nordin, M.A. George: Use of thermally induced higher order modes of a microcantilever for mercury vapor detection, J. Appl. Phys. **99**, 094905 (2006)

15.77 A.R. Kadam, G.P. Nordin, M.A. George: Comparison of microcantilever Hg sensing behavior with thermal higher order modes for as-deposited sputtered and thermally evaporated Au films, J. Vac. Sci. Technol. B **24**, 2271–2275 (2006)

15.78 H.F. Ji, E. Finot, R. Dabestani, T. Thundat, G.M. Brown, P.F. Britt: A novel self-assembled monolayer (SAM) coated microcantilever for low level caesium detection, Chem. Commun. **6**, 457–458 (2000)

15.79 T.A. Brunt, T. Rayment, S.J. O'Shea, M.E. Welland: Measuring the surface stresses in an electrochemically deposited metal monolayer: Pb on Au(111), Langmuir **12**, 5942–5946 (1996)

15.80 J. Mertens, E. Finot, M.H. Nadal, V. Eyraud, O. Heintz, E. Bourillot: Detection of gas trace of hydrofluoric acid using microcantilever, Sens. Act. B Chem. **99**, 58–65 (2004)

15.81 Y.J. Tang, J. Fang, X.H. Xu, H.F. Ji, G.M. Brown, T. Thundat: Detection of femtomolar concentrations of HF using an SiO_2 microcantilever, Anal. Chem. **76**, 2478–2481 (2004)

15.82 J. Mertens, E. Finot, O. Heintz, M.-H. Nadal, V. Eyraud, A. Cathelat, G. Legay, E. Bourillot, A. Dereux: Changes in surface stress, morphology and chemical composition of silica and silicon nitride surfaces during the etching by gaseous HF acid, Appl. Surf. Sci. **253**, 5101–5108 (2007)

15.83 T.L. Porter, T.L. Vail, M.P. Eastman, R. Stewart, J. Reed, R. Venedam, W. Delinger: A solid-state sensor platform for the detection of hydrogen cyanide gas, Sens. Act. B Chem. **123**, 313–317 (2007)

15.84 H.F. Ji, T. Thundat, R. Dabestani, G.M. Brown, P.F. Britt, P.V. Bonnesen: Ultrasensitive detection of CrO_4^{2-} using a microcantilever sensor, Anal. Chem. **73**, 1572–1576 (2001)

15.85 X.H. Xu, T.G. Thundat, G.M. Brown, H.F. Ji: Detection of Hg^{2+} using microcantilever sensors, Anal. Chem. **74**, 3611–3615 (2002)

15.86 S. Cherian, A. Mehta, T. Thundat: Investigating the mechanical effects of adsorption of Ca^{2+} ions on a silicon nitride microcantilever surface, Langmuir **18**, 6935–6939 (2002)

15.87 S. Cherian, R.K. Gupta, B.C. Mullin, T. Thundat: Detection of heavy metal ions using protein-functionalized microcantilever sensors, Biosens. Bioelectron. **19**, 411–416 (2003)

15.88 K. Liu, H.F. Ji: Detection of Pb^{2+} using a hydrogel swelling microcantilever sensor, Anal. Sci. **20**, 9–11 (2004)

15.89 P. Dutta, P.J. Chapman, P.G. Datskos, M.J. Sepaniak: Characterization of ligand-functionalized microcantilevers for metal ion sensing, Anal. Chem. **77**, 6601–6608 (2005)

15.90 F. Tian, V.I. Boiadjiev, L.A. Pinnaduwage, G.M. Brown, T. Thundat: Selective detection of Cr(VI) using a microcantilever electrode coated with a self-assembled monolayer, J. Vac. Sci. Technol. A **23**, 1022–1028 (2005)

15.91 V.I. Boiadjiev, G.M. Brown, L.A. Pinnaduwage, G. Goretzki, P.V. Bonnesen, T. Thundat: Photochemical hydrosilylation of 11-undecenyltriethylammonium bromide with hydrogen-terminated Si surfaces for the development of robust microcantilever sensors for Cr(VI), Langmuir **21**, 1139–1142 (2005)

15.92 N.A. Carrington, L. Yong, Z.-L. Xue: Electrochemical deposition of sol-gel films for enhanced chromium(VI) determination in aqueous solutions, Anal. Chim. Acta **572**, 17–24 (2006)

15.93 S. Velanki, S. Kelly, T. Thundat, D.A. Blake, H.F. Ji: Detection of Cd(II) using antibody-modified microcantilever sensors, Ultramicroscopy **107**, 1123–1128 (2007)

15.94 H. Jensenius, J. Thaysen, A.A. Rasmussen, L.H. Veje, O. Hansen, A. Boisen: A microcantilever-based alcohol vapor sensor-application and response model, Appl. Phys. Lett. **76**, 2815–2817 (2000)

15.95 A. Hierlemann, D. Lange, C. Hagleitner, N. Kerness, A. Koll, O. Brand, H. Baltes: Application-specific sensor systems based on CMOS chemical microsensors, Sens. Act. B Chem. **70**, 2–11 (2000)

15.96 D. Lange, C. Hagleitner, A. Hierlemann, O. Brand, H. Baltes: Complementary metal oxide semiconductor cantilever arrays on a single chip: Mass-sensitive detection of volatile organic compounds, Anal. Chem. **74**, 3084–3095 (2002)

15.97 C. Vancura, M. Rüegg, Y. Li, C. Hagleitner, A. Hierlemann: Magnetically actuated complementary metal oxide semiconductor resonant cantilever gas sensor systems, Anal. Chem. **77**, 2690–2699 (2005)

15.98 L. Fadel, F. Lochon, I. Dufour, O. Français: Chemical sensing: millimeter size resonant microcantilever performance, J. Micromech. Microeng. **14**, S23–S30

(2004)
15.99 Y.I. Wright, A.K. Kar, Y.W. Kim, C. Scholz, M.A. George: Study of microcapillary pipette-assisted method to prepare polyethylene glycol-coated microcantilever sensors, Sens. Act. B **107**, 242–251 (2005)
15.100 L.R. Senesac, P. Dutta, P.G. Datskos, M.J. Sepaniak: Analyte species and concentration identification using differentially functionalized microcantilever arrays and artificial neural networks, Anal. Chim. Acta **558**, 94–101 (2006)
15.101 F. Lochon, L. Fadel, I. Dufour, D. Rebière, J. Pistré: Silicon made resonant microcantilever: Dependence of the chemical sensing performances on the sensitive coating thickness, Mat. Sci. Eng. C **26**, 348–353 (2006)
15.102 S. Satyanarayana, D.T. McCormick, A. Majumdar: Parylene micro membrane capacitive sensor array for chemical and biological sensing, Sens. Act. B **115**, 494–502 (2006)
15.103 D. Then, A. Vidic, C. Ziegler: A highly sensitive self-oscillating cantilever array for the quantitative and qualitative analysis of organic vapor mixtures, Sens. Act. B **117**, 1–9 (2006)
15.104 P.J. Chapman, F. Vogt, P. Dutta, P.G. Datskos, G.L. Devault, M.J. Sepaniak: Facile hyphenation of gas chromatography and a microcantilever array sensor for enhanced selectivity, Anal. Chem. **79**, 364–370 (2007)
15.105 R. Archibald, P. Datskos, G. Devault, V. Lamberti, N. Lavrik, D. Noid, M. Sepianiak, P. Dutta: Independent component analysis of nanomechanical responses of cantilever arrays, Anal. Chim. Acta **584**, 101–105 (2007)
15.106 M. Alvarez, A. Calle, J. Tamayo, L.M. Lechuga, A. Abad, A. Montoya: Development of nanomechanical biosensors for detection of the pesticide DDT, Biosens. Bioelectron. **18**, 649–653 (2003)
15.107 P.G. Datskos, M.J. Sepaniak, C.A. Tipple, N. Lavrik: Photomechanical chemical microsensors, Sens. Act. B **76**, 393–402 (2001)
15.108 L.A. Pinnaduwage, V. Boiadjiev, J.E. Hawk, T. Thundat: Sensitive detection of plastic explosives with self-assembled monolayer-coated microcantilevers, Appl. Phys. Lett. **83**, 1471–1473 (2003)
15.109 L.A. Pinnaduwage, A. Gehl, D.L. Hedden, G. Muralidharan, T. Thundat, R.T. Lareau, T. Sulchek, L. Manning, B. Rogers, M. Jones, J.D. Adams: A microsensor for trinitrotoluene vapour, Nature **425**, 474–474 (2003)
15.110 L.A. Pinnaduwage, D. Yi, F. Tian, T. Thundat, L.T. Lareau: Adsorption of trinitrotoluene on uncoated silicon micro-cantilever surfaces, Langmuir **20**, 2690–2694 (2004)
15.111 L.A. Pinnaduwage, A. Wig, D.L. Hedden, A. Gehl, D. Yi, T. Thundat, R.T. Lareau: Detection of trinitrotoluene via deflagration on a microcantilever, J. Appl. Phys. **95**, 5871–5875 (2004)
15.112 L.A. Pinnaduwage, T. Thundat, A. Gehl, S.D. Wilson, D.L. Hedden, R.T. Lareau: Desorption characteristics, of un-coated silicon microcantilever surfaces for explosive and common nonexplosive vapors, Ultramicroscopy **100**, 211–216 (2004)
15.113 L.A. Pinnaduwage, T. Thundat, J.E. Hawk, D.L. Hedden, R. Britt, E.J. Houser, S. Stepnowski, R.A. McGill, D. Bubb: Detection of 2,4-dinitrotoluene using microcantilever sensors, Sens. Act. B **99**, 223–229 (2004)
15.114 I. Voiculescu, M.E. Zaghloul, R.A. McGill, E.J. Houser, G.K. Fedder: Electrostatically actuated resonant microcantilever beam in CMOS technology for the detection of chemical weapons, IEEE Sens. J. **5**, 641–647 (2005)
15.115 L.A. Pinnaduwage, H.F. Ji, T. Thundat: Moore's law in homeland defense: An integrated sensor platform based on silicon microcantilevers, IEEE Sensors J **5**, 774–785 (2005)
15.116 Y.M. Yang, H.F. Ji, T. Thundat: Nerve agents detection using a Cu^{2+}/L-cysteine bilayer-coated microcantilever, J. Am. Chem. Soc. **125**, 1124–1125 (2003)
15.117 H.F. Ji, X.D. Yan, J. Zhang, T. Thundat: Molecular recognition of biowarfare agents using micromechanical sensors, Expert Rev. Mol. Diagn. **4**, 859–866 (2004)
15.118 H.F. Ji, K.M. Hansen, Z. Hu, T. Thundat: Detection of pH variation using modified microcantilever sensors, Sens. Act. B Chem. **72**, 233–238 (2001)
15.119 R. Bashir, J.Z. Hilt, O. Elibol, A. Gupta, N.A. Peppas: Micromechanical cantilever as an ultrasensitive pH microsensor, Appl. Phys. Lett. **81**, 3091–3093 (2002)
15.120 Y.F. Zhang, H.F. Ji, D. Snow, R. Sterling, G.M. Brown: A pH sensor based on a microcantilever coated with intelligent hydrogel, Instrum. Sci. Technol. **32**, 361–369 (2004)
15.121 M. Watari, J. Galbraith, H.P. Lang, M. Sousa, M. Hegner, Ch. Gerber, M.A. Horton, R.A. McKendry: Investigating the molecular mechanisms of in-plane mechanochemistry on cantilever arrays, J. Am. Chem. Soc. **129**, 601–609 (2007)
15.122 A. Subramanian, P.I. Oden, S.J. Kennel, K.B. Jacobson, R.J. Warmack, T. Thundat, M.J. Doktycz: Glucose biosensing using an enzyme-coated microcantilever, Appl. Phys. Lett. **81**, 385–387 (2002)
15.123 X.D. Yan, H.F. Ji, Y. Lvov: Modification of microcantilevers using layer-by-layer nanoassembly film for glucose measurement, Chem. Phys. Lett. **396**, 34–37 (2004)
15.124 J.H. Pei, F. Tian, T. Thundat: Glucose biosensor based on the microcantilever, Anal. Chem. **76**, 292–297 (2004)
15.125 X.D. Yan, X.H.K. Xu, H.F. Ji: Glucose oxidase multilayer modified microcantilevers for glucose measurement, Anal. Chem. **77**, 6197–6204 (2005)
15.126 X.D. Yan, X.L. Shi, K. Hill, H.F. Ji: Microcantilevers modified by horseradish peroxidase intercalated nano-assembly for hydrogen peroxide detection, Anal. Sci. **22**, 205–208 (2006)
15.127 K.M. Hansen, H.F. Ji, G.H. Wu, R. Datar, R. Cote, A. Majumdar, T. Thundat: Cantilever-based optical deflection assay for discrimination of DNA single-nucleotide mismatches, Anal. Chem. **73**,

1567–1571 (2001)
15.128 M.F. Hagan, A. Majumdar, A.K. Chakraborty: Nanomechanical forces generated by surface grafted DNA, J. Phys. Chem. B **106**, 10163–10173 (2002)
15.129 R. Marie, H. Jensenius, J. Thaysen, C.B. Christensen, A. Boisen: Adsorption kinetics and mechanical properties of thiol-modified DNA-oligos on gold investigated by microcantilever sensors, Ultramicroscopy **91**, 29–36 (2002)
15.130 K.A. Stevenson, A. Mehta, P. Sachenko, K.M. Hansen, T. Thundat: Nanomechanical effect of enzymatic manipulation of DNA on microcantilever surfaces, Langmuir **18**, 8732–8736 (2002)
15.131 M. Su, S.U. Li, V.P. Dravid: Microcantilever resonance-based DNA detection with nanoparticle probes, Appl. Phys. Lett. **82**, 3562–3564 (2003)
15.132 M. Alvarez, L.G. Carrascosa, M. Moreno, A. Calle, A. Zaballos, L.M. Lechuga, C. Martinez, J. Tamayo: Nanomechanics of the formation of DNA self-assembled monolayers and hybridization on microcantilevers, Langmuir **20**, 9663–9668 (2004)
15.133 R.L. Gunter, R. Zhine, W.G. Delinger, K. Manygoats, A. Kooser, T.L. Porter: Investigation of DNA sensing using piezoresistive microcantilever probes, IEEE Sens. J. **4**, 430–433 (2004)
15.134 M. Alvarez, J. Tamayo: Optical sequential readout of microcantilever arrays for biological detection, Sens. Act. B Chem. **106**, 687–690 (2005)
15.135 M. Calleja, M. Nordstrom, M. Alvarez, J. Tamayo, L.M. Lechuga, A. Boisen: Highly sensitive polymer-based cantilever-sensors for DNA detection, Ultramicroscopy **105**, 215–222 (2005)
15.136 R. Mukhopadhyay, M. Lorentzen, J. Kjems, F. Besenbacher: Nanomechanical sensing of DNA sequences using piezoresistive cantilevers, Langmuir **21**, 8400–8408 (2005)
15.137 L.M. Lechuga, J. Tamayo, M. Alvarez, L.G. Carrascosa, A. Yufera, R. Doldan, E. Peralias, A. Rueda, J.A. Plaza, K. Zinoviev, C. Dominguez, A. Zaballos, M. Moreno, C. Martinez, D. Wenn, N. Harris, C. Bringer, V. Bardinal, T. Camps, C. Vergnenegre, C. Fontaine, V. Diaz, A. Bernad: A highly sensitive microsystem based on nanomechanical biosensors for genomics applications, Sens. Act. B Chem. **118**, 2–10 (2006)
15.138 S.L. Biswal, D. Raorane, A. Chaiken, H. Birecki, A. Majumdar: Nanomechanical detection of DNA melting on microcantilever surfaces, Anal. Chem. **78**, 7104–7109 (2006)
15.139 J.C. Stachowiak, M. Yue, K. Castelino, A. Chakraborty, A. Majumdar: Chemomechanics of surface stresses induced by DNA hybridization, Langmuir **22**, 263–268 (2006)
15.140 F. Huber, M. Hegner, C. Gerber, H.J. Guntherodt, H.P. Lang: Label free analysis of transcription factors using microcantilever arrays, Biosens. Bioelectron. **21**, 1599–1605 (2006)
15.141 S.L. Biswal, D. Raorane, A. Chaiken, A. Majumdar: Using a microcantilever array for detecting phase transitions and stability of DNA, Clin. Lab. Med. **27**, 163–163 (2007)
15.142 J. Zhang, H.P. Lang, F. Huber, A. Bietsch, W. Grange, U. Certa, R. McKendry, H.-J. Guentherodt, M. Hegner, C. Gerber: Rapid and label-free nanomechanical detection of biomarker transcripts in human RNA, Nat. Nanotechnol. **1**, 214–220 (2006)
15.143 A.M. Moulin, S.J. O'Shea, M.E. Welland: Microcantilever-based biosensors, Ultramicroscopy **82**, 23–31 (2000)
15.144 C. Grogan, R. Raiteri, G.M. O'Connor, T.J. Glynn, V. Cunningham, M. Kane, M. Charlton, D. Leech: Characterisation of an antibody coated microcantilever as a potential immuno-based biosensor, Biosens. Bioelectron. **17**, 201–207 (2002)
15.145 A. Kooser, K. Manygoats, M.P. Eastman, T.L. Porter: Investigation of the antigen antibody reaction between anti-bovine serum albumin (a-BSA) and bovine serum albumin (BSA) using piezoresistive microcantilever based sensors, Biosens. Bioelectron. **19**, 503–508 (2003)
15.146 C.A. Savran, S.M. Knudsen, A.D. Ellington, S.R. Manalis: Micromechanical detection of proteins using aptamer-based receptor molecules, Anal. Chem. **76**, 3194–3198 (2004)
15.147 J.H. Lee, K.H. Yoon, K.S. Hwang, J. Park, S. Ahn, T.S. Kim: Label free novel electrical detection using micromachined PZT monolithic thin film cantilever for the detection of C-reactive protein, Biosens. Bioelectron. **20**, 269–275 (2004)
15.148 Y.F. Zhang, S.P. Venkatachalan, H. Xu, X.H. Xu, P. Joshi, H.F. Ji, M. Schulte: Micromechanical measurement of membrane receptor binding for label-free drug discovery, Biosens. Bioelectron. **19**, 1473–1478 (2004)
15.149 N. Backmann, C. Zahnd, F. Huber, A. Bietsch, A. Pluckthun, H.P. Lang, H.J. Güntherodt, M. Hegner, C. Gerber: A label-free immunosensor array using single-chain antibody fragments, Proc. Natl. Acad. Sci. USA **102**, 14587–14592 (2005)
15.150 K.W. Wee, G.Y. Kang, J. Park, J.Y. Kang, D.S. Yoon, J.H. Park, T.S. Kim: Novel electrical detection of label-free disease marker proteins using piezoresistive self-sensing micro-cantilevers, Biosens. Bioelectron. **20**, 1932–1938 (2005)
15.151 R. Mukhopadhyay, V.V. Sumbayev, M. Lorentzen, J. Kjems, P.A. Andreasen, F. Besenbacher: Cantilever sensor for nanomechanical detection of specific protein conformations, Nano Lett **5**, 2385–2388 (2005)
15.152 X. Yan, K. Hill, H. Gao, H.F. Ji: Surface stress changes induced by the conformational change of proteins, Langmuir **22**, 11241–11244 (2006)
15.153 H.S. Kwon, K.C. Han, K.S. Hwang, J.H. Lee, T.S. Kim, D.S. Yoon, E.G. Yang: Development of a peptide inhibitor-based cantilever sensor assay for cyclic adenosine monophosphate-dependent protein kinase, Anal. Chim. Acta **585**, 344–349 (2007)
15.154 W. Shu, E.D. Laue, A.A. Seshia: Investigation of biotin-streptavidin binding interactions using microcantilever sensors, Biosens. Bioelectron. **22**,

2003–2009 (2007)
15.155 V. Dauksaite, M. Lorentzen, F. Besenbacher, J. Kjems: Antibody-based protein detection using piezoresistive cantilever arrays, Nanotechnology **18**, 125503 (2007)
15.156 M. Yue, J.C. Stachowiak, H. Lin, R. Datar, R. Cote, A. Majumdar: Label-free protein recognition two-dimensional array using nanomechanical sensors, Nano Lett **8**, 520–524 (2008)
15.157 T. Braun, N. Backmann, M. Vogtli, A. Bietsch, A. Engel, H.P. Lang, C. Gerber, M. Hegner: Conformational change of bacteriorhodopsin quantitatively monitored by microcantilever sensors, Biophys. J. **90**, 2970–2977 (2006)
15.158 B.H. Kim, O. Mader, U. Weimar, R. Brock, D.P. Kern: Detection of antibody peptide interaction using microcantilevers as surface stress sensors, J. Vac. Sci. Technol. B **21**, 1472–1475 (2003)
15.159 I. Pera, J. Fritz: Sensing lipid bilayer formation and expansion with a microfabricated cantilever array, Langmuir **23**, 1543–1547 (2007)
15.160 S.J. Hyun, H.S. Kim, Y.J. Kim, H.I. Jung: Mechanical detection of liposomes using piezoresistive cantilever, Sens. Act. B Chem. **117**, 415–419 (2006)
15.161 B. Ilic, D. Czaplewski, M. Zalalutdinov, H.G. Craighead, P. Neuzil, C. Campagnolo, C. Batt: Single cell detection with micromechanical oscillators, J. Vac. Sci. Technol. B **19**, 2825–2828 (2001)
15.162 J. Park, J. Ryu, S.K. Choi, E. Seo, J.M. Cha, S. Ryu, J. Kim, B. Kim, S.H. Lee: Real-time measurement of the contractile forces of self-organized cardiomyocytes on hybrid biopolymer microcantilevers, Anal. Chem. **77**, 6571–6580 (2005)
15.163 S. Ghatnekar-Nilsson, J. Lindahl, A. Dahlin, T. Stjernholm, S. Jeppesen, F. Hook, L. Montelius: Phospholipid vesicle adsorption measured in situ with resonating cantilevers in a liquid cell, Nanotechnology **16**, 1512–1516 (2005)
15.164 N. Nugaeva, K.Y. Gfeller, N. Backmann, H.P. Lang, M. Duggelin, M. Hegner: Micromechanical cantilever array sensors for selective fungal immobilization and fast growth detection, Biosens. Bioelectron. **21**, 849–856 (2005)
15.165 B. Dhayal, W.A. Henne, D.D. Doorneweerd, R.G. Reifenberger, P.S. Low: Detection of Bacillus subtilis spores using peptide-functionalized cantilever arrays, J. Am. Chem. Soc. **128**, 3716–3721 (2006)
15.166 G. Wu, R.H. Datar, K.M. Hansen, T. Thundat, R.J. Cote, A. Majumdar: Bioassay of prostate-specific antigen (PSA) using microcantilevers, Nature Biotechnol **19**, 856–860 (2001)
15.167 K.S. Hwang, J.H. Lee, J. Park, D.S. Yoon, J.H. Park, T.S. Kim: In-situ quantitative analysis of a prostate-specific antigen (PSA) using a nanomechanical PZT cantilever, Lab Chip **4**, 547–552 (2004)
15.168 J.H. Lee, K.S. Hwang, J. Park, K.H. Yoon, D.S. Yoon, T.S. Kim: Immunoassay of prostate-specific antigen (PSA) using resonant frequency shift of piezoelectric nanomechanical microcantilever, Biosens. Bioelectron. **20**, 2157–2162 (2005)
15.169 S. Velanki, H.-F. Ji: Detection of feline coronavirus using microcantilever sensors, Meas. Sci. Technol. **17**, 2964–2968 (2006)
15.170 D. Maraldo, F.U. Garcia, R. Mutharasan: Method for quantification of a prostate cancer biomarker in urine without sample preparation, Anal. Chem. **79**, 7683–7690 (2007)
15.171 W.Y. Shih, X.P. Li, H.M. Gu, W.H. Shih, I.A. Aksay: Simultaneous liquid viscosity and density determination with piezoelectric unimorph cantilevers, J. Appl. Phys. **89**, 1497–1505 (2001)
15.172 A. Agoston, F. Keplinger, B. Jakoby: Evaluation of a vibrating micromachined cantilever sensor for measuring the viscosity of complex organic liquids, Sens. Act. A Phys. **123–24**, 82–86 (2005)
15.173 N. McLoughlin, S.L. Lee, G. Hahner: Simultaneous determination of density and viscosity of liquids based on resonance curves of uncalibrated microcantilevers, Appl. Phys. Lett. **89**, 184106 (2006)
15.174 U. Sampath, S.M. Heinrich, F. Josse, F. Lochon, I. Dufour, D. Rebiere: Study of viscoelastic effect on the frequency shift of microcantilever chemical sensors, IEEE T. Ultrason. Ferr. **53**, 2166–2173 (2006)
15.175 A. Jana, A. Raman, B. Dhayal, S.L. Tripp, R.G. Reifenberger: Microcantilever mechanics in flowing viscous fluids, Appl. Phys. Lett. **90**, 114110 (2007)
15.176 M. Hennemeyer, S. Burghardt, R. Stark: Cantilever micro-rheometer for the characterization of sugar solutions, Sensors **8**, 10–22 (2008)
15.177 K.B. Brown, Y. Ma, W. Allegretto, R.P.W. Lawson, F.E. Vermeulen, A.M. Robinson: Microstructural pressure sensor based on an enhanced resonant mode hysteresis effect, J. Vac. Sci. Technol. B **19**, 1628–1632 (2001)
15.178 K.B. Brown, W. Allegretto, F.E. Vermeulen, A.M. Robinson: Simple resonating microstructures for gas pressure measurement, J. Micromech. Microeng. **12**, 204–210 (2002)
15.179 Y. Su, A.G.R. Evans, A. Brunnschweiler, G. Ensell: Characterization of a highly sensitive ultrathin piezoresistive silicon cantilever probe and its application in gas flow velocity sensing, J. Micromech. Microeng. **12**, 780–785 (2002)
15.180 J. Mertens, E. Finot, T. Thundat, A. Fabre, M.H. Nadal, V. Eyraud, E. Bourillot: Effects of temperature and pressure on microcantilever resonance response, Ultramicroscopy **97**, 119–126 (2003)
15.181 V. Mortet, R. Petersen, K. Haenen, M. D'Olieslaeger: Wide range pressure sensor based on a piezoelectric bimorph microcantilever, Appl. Phys. Lett. **88**, 133511 (2006)
15.182 P. Sievilä, V.P. Rytkönen, O. Hahtela, N. Chekurov, J. Kauppinen, I. Tittonen: Fabrication and characterization of an ultrasensitive acousto-optical cantilever, J. Micromech. Microeng. **17**, 852–859 (2007)
15.183 C.A. Jong, T.S. Chin, W.L. Fang: Residual stress and thermal expansion behavior of TaO_xN_y films by the micro-cantilever method, Thin Solid Films

401, 291–297 (2001)
15.184 F. Shen, P. Lu, S.J. O'Shea, K.H. Lee, T.Y. Ng: Thermal effects on coated resonant microcantilevers, Sens. Act. A Phys. **95**, 17–23 (2001)
15.185 P.G. Datskos, P.I. Oden, T. Thundat, E.A. Wachter, R.J. Warmack, S.R. Hunter: Remote infrared radiation detection using piezoresistive microcantilevers, Appl. Phys. Lett. **69**, 2986–2988 (1996)
15.186 P.I. Oden, P.G. Datskos, T. Thundat, R.J. Warmack: Uncooled thermal imaging using a piezoresistive microcantilever, Appl. Phys. Lett. **69**, 3277–3279 (1996)
15.187 E.A. Wachter, T. Thundat, P.I. Oden, R.J. Warmack, P.G. Datskos, S.L. Sharp: Remote optical detection using microcantilevers, Rev. Sci. Instrum. **67**, 3434–3439 (1996)
15.188 P.G. Datskos, S. Rajic, I. Datskou: Detection of infrared photons using the electronic stress in metal-semiconductor cantilever interfaces, Ultramicroscopy **82**, 49–56 (2000)
15.189 C.B. Li, B.B. Jiao, S.L. Shi, D.P. Chen, T.C. Ye, Q.C. Zhang, Z.Y. Guo, F.L. Dong, Z.Y. Miao: A novel uncooled substrate-free optical-readable infrared detector: design, fabrication and performance, Meas. Sci. Technol. **17**, 1981–1986 (2006)
15.190 X. Yu, Y. Yi, S. Ma, M. Liu, X. Liu, L. Dong, Y. Zhao: Design and fabrication of a high sensitivity focal plane array for uncooled IR imaging, J. Micromech. Microeng. **18**, 057001 (2008)
15.191 B. Jiao, C. Li, D. Chen, T. Ye, Y. Ou, L. Dong, Q. Zhang: An Optical Readout Method Based Uncooled Infrared Imaging System, Int. J. Infrared Milli. Waves **29**, 261–271 (2008)
15.192 J. Cheste, A. Eichler, J. Moser, C. Ceballos, R. Rurali, A. Bachtold: A nanomechanical mass sesnor with yoctogram resolution, Nat. Nanotechnol. **7**, 301–304 (2012)
15.193 Y. Nagai, J.D. Carbajal, J.H. White, R. Sladek, P. Grutter, R.B. Lennox: An electrochemically controlled microcantilever biosensor, Langmuir: ACS J. Surf. Colloids **29**, 9951–9957 (2013)
15.194 A.-L. Haag, Y. Nagai, R.B. Lennox, P. Grütter: Characterization of a gold coated cantilever surface for biosensing applications, EPJ Tech. Instrum. **2**, 1–12 (2015)
15.195 S. Mishra, S. Ghosh, R. Mukhopadhyay: Ordered self-assembled Locked Nucleic Acid (LNA) structures on gold(111) surface with enhanced single base mismatch recognition capability, Langmuir **28**, 4325–4333 (2012)
15.196 S. Mishra, S. Ghosh, R. Mukhopadhyay: Regulating on-surface LNA probe density for the highest target recognition efficiency, Langmuir **30**, 10389–10397 (2014)
15.197 C.M. Domínguez, P.M. Kosaka, G. Mokry, V. Pini, O. Malvar, M. del Rey, D. Ramos, Á. San Paulo, J. Tamayo, M. Calleja: Hydration induced stress on DNA monolayers grafted on microcantilevers, Langmuir **30**, 10962–10969 (2014)
15.198 J.W. Ndieyira, M. Watari, A.D. Barrera, D. Zhou, M. Vögtli, M. Batchelor, M.A. Cooper, T. Strunz, M.A. Horton, C. Abell, T. Rayment, G. Aeppli, R.A. McKendry: Nanomechanical detection of antibiotic–mucopeptide binding in a model for superbug drug resistance, Nat. Nanotech. **3**, 691–696 (2008)
15.199 J.W. Ndieyira, N. Kappeler, S. Logan, M.A. Cooper, C. Abell, R.A. McKendry, G. Aeppli: Surface-stress sensors for rapid and ultrasensitive detection of active free drugs in human serum, Nat. Nanotech. **9**, 225–232 (2014)
15.200 M. Bache, F.G. Bosco, A.L. Brogger, K.B. Frohling, T.S. Alstrom, E.T. Hwu, C.H. Chen, J. Eugen-Olsen, I.S. Hwang, A. Boisen: Nanomechanical recognition of prognostic biomarker suPAR with DVD-ROM optical technology, Nanotechnology **24**, 444011 (2013)
15.201 R. Mishra, M. Hegner: Effect of non-specific species competition from total RNA on the static mode hybridization response of nanomechanical assays of oligonucleotides, Nanotechnology **25**, 225501 (2014)
15.202 F. Huber, H.P. Lang, N. Backmann, D. Rimoldi, C. Gerber: Direct detection of a BRAF mutation in total RNA from melanoma cells using cantilever arrays, Nat. Nanotechnol. **8**, 125–129 (2013)
15.203 F. Huber, H.P. Lang, J.Y. Zhang, D. Rimoldi, C. Gerber: Nanosensors for cancer detection, Swiss Med. Wkly. **145**, w14092 (2015) doi:10.4414/smw.2015.14092
15.204 N. Maloney, G. Lukacs, J. Jensen, M. Hegner: Nanomechanical sensors for single microbial cell growth monitoring, Nanoscale **6**, 8242–8249 (2014)
15.205 N. Maloney, G. Lukacs, S.L. Ball, M. Hegner: Device for filamentous fungi growth monitoring using the multimodal frequency response of cantilevers, Rev. Sci. Instrum. **85**, 015003 (2014)
15.206 G. Longo, L. Alonso-Sarduy, L. Marques Rio, A. Bizzini, A. Trampuz, J. Notz, G. Dietler, S. Kasas: Rapid detection of bacterial resistance to antibiotics using AFM cantilevers as nanomechanical sensors, Nat. Nanotech. **8**, 522–526 (2013)
15.207 S. Kasas, F.S. Ruggeri, C. Benadiba, C. Maillard, P. Stupar, H. Tournu, G. Dietler, G. Longo: Detecting nanoscale vibrations as signature of life, Proc. Natl. Acad. Sci. (USA) PNAS **112**, 378–381 (2015)
15.208 E. Finot, V. Rouger, L. Markey, R. Seigneuric, M.H. Nadal, T. Thundat: Visible photothermal deflection spectroscopy using microcantilevers, Sens. Act. B **169**, 222–228 (2012)
15.209 M.S. Ghoraishi, J.E. Hawk, A. Phani, M.F. Khan, T. Thundat: Clustering mechanism of ethanol-water mixtures investigated with photothermal microfluidic cantilever deflection spectroscopy, Sci. Rep. **6**, 23966 (2016) doi:10.1038/srep23966
15.210 M. Pelton, D. Chakraborty, E. Malachosky, P. Guyot-Sionnest, J.E. Sader: Viscoelastic flows in simple liquids generated by vibrating nanostructures, Phys. Rev. Lett. **111**, 244502 (2013)
15.211 T.P. Burg, S.R. Manalis: Suspended microchannel resonators for biomolecular detection, Appl. Phys. Lett. **83**, 2698–2700 (2003)

15.212 P. Dutta, C. Tipple, N. Lavrik, P. Datskos: Enantioselective sensors based on antibody-mediated nanomechanics, Anal. Chem. **75**, 2342–2348 (2003)
15.213 H.F. Ji, Y.Q. Lu, H.W. Du, X.H. Xu, T. Thundat: Spiral springs and microspiral springs for chemical and biological sensing, Appl. Phys. Lett. **88**, 063504 (2006)
15.214 G. Singh, P. Rice, R.L. Mahajan: Fabrication and mechanical characterization of a force sensor based on an individual carbon nanotube, Nanotechnology **18**, 475501 (2007)
15.215 A. Sakhaee-Pour, M.T. Ahmadian, A. Vafai: Applications of single-layered graphene sheets as mass sensors and atomistic dust detectors, Solid State Commun **145**, 168–172 (2008)
15.216 X. Song, M. Oksanen, J. Li, P.J. Hakonen, M.A. Sillanpää: Graphene optomechanics realized at microwave frequencies, Phys. Rev. Lett. **113**, 027404 (2014)
15.217 R.J. Dolleman, D. Davidovikj, S.J. Cartamil-Bueno, H.S. van der Zant, P.G. Steeneken: Graphene squeeze-film pressure sensors, Nano Lett **16**, 568–571 (2016)

16. Microfluidic Devices and Their Applications

Aditya Aryasomayajula, Pouriya Bayat, Pouya Rezai, P. Ravi Selvaganapathy

Microfluidics and nanofluidics is a field of science that operates in the micrometer and nanometer scale. A microfluidic–nanofluidic device consists of components such as valves, pumps and mixers for manipulating and transporting the fluid at this scale. In this chapter we review the history, physics, fabrication methods and applications of microfluidics and nanofluidics. This interdisciplinary field has a wide range of application areas including environmental sensing, medical diagnostics, drug discovery, drug delivery, microscale chemical production, combinatorial synthesis and assays, artificial organs, and micropropulsion, microscale energy systems. The global market for microfluidic devices was estimated at around $3.1 billion dollars in 2015 and is expected to rise to $7.5 billion dollars by 2020. In the future, microfluidics and nanofluidics will see miniaturization and development of novel microfabrication techniques along with more sensitive detection methods and diagnosis of diseases in a point-of-care platform. Developments in the fundamental physics of fluid flow and its control, microfabrication methods, microfluidic components, and applications in new and emerging areas are all anticipated.

16.1 **Preface** 487
16.1.1 Why Miniaturize? 488
16.1.2 Advantages of Microfluidics and Nanofluidics 488
16.2 **Historical Developments** 489
16.3 **Dominant Phenomenon and Micro–Nanofluidic Design** 491
16.3.1 Laminar Flow 492
16.3.2 Diffusion in Microfluidics 492
16.3.3 Rapid Heating–Cooling 492
16.3.4 Surface Tension 493
16.3.5 Electrokinetics 494
16.3.6 Geometric Control of Two Phase Flow 495
16.3.7 Nanoscale Phenomena 495
16.4 **Fabrication of Micro–Nanofluidic Devices** 496
16.4.1 Silicon and Glass 496
16.4.2 Polymers 497
16.4.3 Paper 499
16.4.4 Threads 501
16.4.5 Hydrogels 502
16.5 **Applications** 504
16.5.1 Commercial Applications and Their Working Mechanism 504
16.5.2 Research Applications 519
16.6 **Outlook and Future Directions** 521
16.6.1 Additive Manufacturing 521
16.6.2 Artificial Organs 521
16.6.3 Biopharmaceutical Production 521
16.6.4 Droplets in Food and Cosmetics 522
16.6.5 Continuous Health Monitoring 522
References 522

16.1 Preface

Microfluidics and nanofluidics are fields of study dedicated to the understanding of fluid flow in the micrometer (μm) to nanometer (nm) scales and the development of devices to manipulate liquids or gases in the microliter to femtoliter range. In order to put the dimensions in context, a human hair on average is about 100 μm thick and a strand in our genetic code, deoxyribonucleic acid (DNA) is about 2.5 nm thick. Unlike the macroscale, where inertial (bulk) forces and momentum dominate, fluid flow at the micro- and nanoscale is highly influenced by fluid drag, surface forces and interactions as well as other molecular-level effects [16.1]. The understanding of the fluid flow phenomena at these scale and the ability to fabricate micromechanical struc-

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_16

tures can be applied to develop components such as valves, pumps and mixers to handle such small volumes of fluid and process them in a defined sequence. These components can be combined to perform a sequence of unit operations such as filtration, mixing, heating–cooling, dilutions and others required for an assay or chemical production in what is called a micro total analysis system [16.2] (μTAS) or lab-on-chip (LoC) system. A well-known example of such an LoC system is the glucometer. The glucometer performs separation of plasma from whole blood, mixing of the plasma with a dried glucose oxidase enzyme on-chip, reaction of the glucose oxidase enzyme with glucose in the plasma and its subsequent coupling with redox compounds to electrochemically transduce the signal into an electrical current. All of these unit operations are performed in an automated way inside a flow cell coupled with the electrochemical sensing electrode and can analyze a single drop of blood from a patient with a high degree of reliability [16.3].

A more recent example of a nanofluidic device is a DNA analyzer called minION Mkl (Fig. 16.1). This device uses nanopore technology to sequence a single strand of DNA. It consists of a nanopore (≈ 2–20 nm diameter) that is electrically conducting. The presence of the DNA strand inside the nanopores, which can be pulled in by electric field, displaces some of the ions, which reduces the current. Since the pore is small, small stretches of DNA can have a unique modulation of the current through the pore, which can be sensed [16.4].

Apart from these specific examples, microfluidic devices have had significant impact in various bioanalytical and biomedical systems like microarrays, capillary electrophoresis, nucleic acid amplification devices, immunoassays, cell culture, biosorting, hemoanalysis, metabolic analysis, drug discovery assays, *in-vivo* and *in-vitro* drug delivery, artificial organs, medical and environmental sensors, microdispensers and inkjet print heads. The market for microfluidic devices was about $3.1 billion in 2015 and is expected to grow to $7.5 billion by 2020 [16.5].



Fig. 16.1 Oxford Nanopore Technologies Ltd minION Mkl device to sequence the single DNA strand. This device uses a nanopore (≈ 2–20 nm diameter) to measure the current of a single nucleotide for sequencing. (Courtesy of Oxford Nanopore Technologies Ltd)

16.1.1 Why Miniaturize?

There are several important reasons for the drive towards miniaturization of unit operations into a microfluidic format. The foremost among them is that through miniaturization all the unit operations can be packaged into a small form factor that could be completely automated and made portable. Automation can enable the use of these systems by the user without any specific training in the assay, which will then enable a broader and more widespread use. For instance, the Cepheid GeneXpert device [16.6] automates the process of cell lysis, DNA extraction, its amplification and detection so that assays for disease identification can be performed at any location, from hospital laboratories to rural regions and resource-poor settings, with equal ease and without skilled laboratory personnel.

The other main reason is the ability to parallelize many operations on one device thereby expanding the capabilities offered by the device by orders of magnitude. The advent of Affymetrix microarray technology [16.7], which used a photolithographic process employed in the electronic industry to synthesize a large number of unique nucleic acid strands closely spaced to each other in a microarray, has revolutionized biological analysis by analyzing protein compositional signatures of cells and other biological matrices.

Another unique advantage offered by miniaturization is the similar size of sensors and reaction chambers to that of a cell, which makes it possible to manipulate and sense proteins or nucleic acids at single-molecule resolutions. The example of the nanopore sensor [16.4] and its capability in reading a single strand of DNA is illustrative of this reason.

Miniaturized devices also consume significantly lower volumes of reagents, which can reduce the cost of assays and mean they can be manufactured at lower prices. They generally operate in the laminar flow regime, which is more convenient for handling, transporting and aliquoting small volumes of fluids and small objects such as cells. Furthermore, small operating volumes also offer faster processing and analysis times in assays that require heating or cooling such as nucleic acid amplification.

16.1.2 Advantages of Microfluidics and Nanofluidics

The microfluidic technology exploits the physical and chemical properties of fluids at the micrometer scale.

The advantages offered by this technology have led to the development of techniques with improved capabilities compared to the conventional methods. Chemistry and biology are the two primary areas where microfluidics has had a significant impact. Chromatography and electrophoresis were one of the earliest areas of research where microfluidic technology was applied. Microfluidics not only improved the sensitivity of these techniques but also reduced the overall reaction times significantly. This is due to the ability to inject extremely small sample volumes for separation through the tight integration of the sample injector with the separation column, which was not possible earlier. In addition, the low consumption of chemical compounds has reduced the cost of the overall method drastically. In biology, improved control of the microenvironment has made way for studying three-dimensional (3-D) cell culture, which was previously not possible. The operating scale in microfluidic technology makes it particularly suitable for studying cells, organisms, and bacteria. The high throughput and easy handling of microliter volumes has helped researchers to develop alternative methods like polymerase chain reaction (PCR) on a chip, and DNA sequencing and microfluidic sorting methods that are faster, more sensitive and consume fewer reagents when compared to the conventional methods.

One of the unique advantages of micro- and nanofluidic technology is the physical and chemical phenomenon that applies at this microscale. On-chip PCR for DNA amplification exploits the fast diffusion of chemical compounds and transfer of heat at the micrometer level. Microfluidic devices have a high surface area-to-volume ratio because surface tension dominates and reduces inertial and body forces. In digital microfluidic technology, droplets are formed using the surface tension property of liquids and droplet production is precisely manipulated using electric fields. This property is also useful in chip-based electrophoresis devices where Joule heating can be reduced, thereby increasing the performance, sensitivity and resolution of separation of compounds. High viscous forces inside microfluidic channels lead to laminar flow characteristics. This property is exploited to control particle location precisely and for defining electric field strength inside microfluidic channels. Cell sorting and electrotaxis devices are examples that make use of this property. The small size of the devices and channels means lower consumption of reagents, which is particularly useful in drug screening tests as it can lower the overall cost of the experiment. The ability to form thousands and millions of droplets in oil using droplet microfluidics technology where each droplet acts as a separate reaction chamber increases the throughput to a great extent and also makes it possible to carry out single-cell analysis. The rapid prototyping technique using low-cost soft lithography gives a high flexibility in designing devices. The high parallelization or multiplexing is useful for DNA sequencers. Small sample volumes along with high sensitivity enable this technology to detect trace amounts of target species efficiently. Microfluidics also offers the advantage of integrating different technologies onto one platform quickly.

Microfluidics is one of the fastest growing research areas with over 15 300 publications since 2015 (number of publications with the word *microfluidics* searched in Google Scholar). The broad application areas, including physics, chemistry, biology, environmental sciences and engineering, make it a very attractive area to investigate. Microfluidics also fosters a large interdisciplinary community, bringing people in different fields for the first time to interact and carry out research, thereby pushing the frontier of science forward. Microfluidics also has generated a billion-dollar revenue in the area of LoC, diagnostics and the drug screening/pharmaceutical industry, which is likely to grow bigger in the coming years. As is the norm in a fast-growing field, the central challenge in microfluidics is the lack of standardizationof techniques and components due to the diversity of users and applications. In the next sections of this chapter, the history of this field, the physical principles that govern in this technology, different fabrication methodologies, some examples of applications of this area and the outlook and future direction of microfluidics will be presented.

16.2 Historical Developments

Although flows in the micro- and nanoscale have been studied significantly far back in history, such as electroosmosis by *Reuss* in the early 1800s [16.8], they have been investigated as an aggregate phenomenon. The advent of technologies for miniaturization and its application to the fabrication of micromechanical structures such as microchannels kick-started the study of fundamental phenomena of fluid flow in these scales as well as their application in various domains. One of the early works was the application of the emerging silicon microfabrication technology in the early 1970s to fabricate high surface area microchannels for gas chro-

matography applications. *Terry* et al. [16.9] developed a microscale gas chromatography device with an inbuilt sensor, which was sensitive but not specific to gas detection, and used the same principle of a conventional gas chromatographer to separate gaseous hydrocarbon mixtures in less than 10 s. They demonstrated that high surface area columns can be fabricated and extremely long lengths can be fitted into a small form factor on a chip. They also proposed that miniaturization and integration of all the other components including preconcentrators and pump would make the device portable compared to the bulky conventional equipment.

Around the same time, inkjet printers were introduced [16.10] in the market that used microfluidic technology to print ink on paper. The inkjet cartridges have a tiny nozzle ($\approx 30-100\,\mu m$) through which the ink is forced out. The nozzles are fabricated in silicon using photolithographic approaches used in the electronic industry so that up to 600 nozzles can be placed in a single printhead in close proximity to each other, with each one individually connected to the ink reservoir through microfluidic channels. Today the inkjet printing market is worth $\approx$ \$900 million and projected to be over \$1 billion by 2021 [16.11].

Then in the early 1990s, a group of scientists led by *Stephen Fodor* invented the Affymetrix gene chip technology [16.12]. They used semiconductor manufacturing technologies to fabricate DNA microarrays that have millions of unique DNA recognition elements on a single glass chip. These microarrays massively expanded the capability of biologists to analyze gene expression of thousands to millions of proteins in a cell or tissue, a capability that was not available before. This technology is currently used widely and has become a standard in the field. The microarray market is $\approx$ \$1.5 billion in 2015 and is likely to grow to \$2 billion by 2020 [16.13]. Around the same time, *Manz* et al. [16.2] introduced the idea of using miniaturized integrated systems that used electrophoresis or chromatography for separation and integrated sensors for detection of a variety of chemicals and biomolecules. Their pioneering analysis started a new field called micro total analysis systems (μTAS) or lab-on-chip (LoC). They compared the theoretical performance of such a system for flow injection analysis, chromatography, and electrophoresis with the existing chemical sensors and analysis systems and noted that miniaturization made devices separations more efficient and faster with shorter transport times. They also found that μTAS consumed much fewer reagents and had higher sensitivity. Since then research and development in LoC systems for medical, environmental and other biochemical analysis has taken off and excellent review articles on the theory and technology [16.14], applications in chemistry and biology [16.15–17] and miniaturization for POC applications [16.18] are available.

A little later, in 1993, *Northrup* et al. [16.19] developed the first polymerase chain reaction (PCR) chip for DNA amplification and later went on to develop an integrated PCR and electrophoresis system on a single silicon chip [16.20]. The chip was fabricated using a photolithography technique and included a PCR chip to amplify the DNA and later separate the DNA using capillary electrophoresis. Their work opened the field of miniaturization to biological applications. They showed that miniaturization of the PCR analysis reduced the overall amplification time due to faster heating and cooling rates obtained in their device.

One of the first microfluidic fully integrated DNA sample analysis devices was developed by *Burns* et al. [16.21]. The device consisted of microchambers, microchannels, heaters, temperature sensors, and a fluorescence detector all integrated on a single chip. All the components of the system were made from similar materials and fabricated together using photolithography-based microfabrication methods. It was capable of aliquoting sample and reagents, metering them, mixing, performing an amplification reaction, transporting the sample to a separation channel and measuring the sequence of arrival of DNA from the sample on the integrated detector.

Until the late 1990s, researchers mainly used silicon and glass microfabrication technology to develop microfluidic devices. This technology was useful for fabricating microfeatures but was expensive and time consuming. In the late 1990s significant effort was devoted to developing microfabrication methods for LoC devices in plastics and polymers [16.22]. One of the techniques that emerged out of this polymer revolution was a rapid prototyping method known as soft lithography [16.23] where casting of polymers on lithographically fabricated molds was used to fabricate microfluidic devices. The material used was an elastomeric silicone rubber known as polydimethylsiloxane (PDMS), which was capable of replicating mold features down to a few 100 nm and was able to bond with glass or other PDMS substrates using a simple plasma exposure. The material was optically clear, had little autofluorescence, was flexible, had high gas permeability and had excellent elastomeric properties for the fabrication of membranes and diaphragms. The low cost, simple setup and the ability to rapidly prototype made this method popular and accessible to many labs around the world that were interested in developing microfluidic devices. Further development of valving and pumping technology for PDMS-based microfluidic devices using pneumatics by *Quake* et al. in the early 2000s [16.24] and the demonstration of scalability of

this approach and integration of hundreds of valves on a single chip made PDMS and soft lithography well established and the standard method for rapid prototyping in the microfluidic and nanofluidic research community. The PDMS valves that were developed are widely used in many microfluidic systems such as those to study PCR [16.25], protein separations [16.26] and cell sorting [16.27, 28]. Extensive reviews on various aspects of soft lithography including its application in nanofabrication [16.29], microfluidics [16.30], patterning [16.31] and application in the field of biology and biochemistry [16.32] are available.

Towards the middle of the 2000s, a new subfield called droplet-digital microfluidics emerged based on the science of emulsions. This technology exploited the fact that precisely fabricated microscale geometries can be used bring together two or more immiscible phases (such as water and oil) to form tiny droplets (in micrometer scale) precisely. Since then this subfield has grown enormously and large number of studies have been performed to gain deeper understanding of the physics of droplet formation [16.33, 34] and to develop methods for generating droplets [16.35–38] and their manipulation [16.39]. Furthermore, this technology has seen applications in single-cell analysis [16.40], molecular detection [16.41, 42], high-throughput imaging [16.43], drug delivery [16.44, 45] and diagnostics [16.46, 47].

Towards the end of the 2000s, several groups extended the application of microfluidic devices to performing biologically significant assays on small biological organisms in what was termed organism-on-chip. Organisms like *Caenorhabditis elegans*, *Drosophila melanogaster* and zebrafish are biologically important as they are considered to be rudimentary but significant models of human biological systems. In addition, disease models have been developed in these organisms to mimic the physiological characteristics of diseases such as Parkinson's and Alzheimer's. Some of the stages of these organisms, especially the embryo and the larval stages, are at the same size range as microfluidic devices. Several innovative microfluidic devices were developed for *Caenorhabditis elegans* [16.48–50], *Drosophila melanogaster* [16.51–53] and zebrafish [16.54–56] primarily to perform imaging and behavioral studies in these organisms in response to drugs and external stimuli. More recently, microfluidic devices have been used to recreate the biological and functional aspects of human organs for drug testing. These microfluidic chips are engineered to mimic the systems that produce the key functions of a living organ, for example, lungs, kidneys etc. These devices have proven to be more accurate in mimicking the organ functions compared to conventional cell cultures, which has led to an increased interest from the pharmaceutical industry. Recent reviews on organ-on-chip provide a detailed understanding of the state of this research as well as the future prospects [16.57, 58].

Although nanofluidics emerged later than microfluidics, it has made rapid progress in the past decade especially in advancing biological measurements at single-molecule resolution. Use of nanopores [16.59] for DNA sequencing and understanding protein folding are examples of applications where nanofluidic devices are used. Excellent, in-depth reviews on single-molecule detection [16.60] and biological measurement at single-molecule resolution [16.61] have been recently published. Nanopatterned structures have also been developed as artificial sieving matrices to interact with biomolecules at their size range in order to separate them [16.62]. Nanofluidic devices have also been used for improving the performance of microscale polymer electrolyte membrane fuel cells [16.63].

As the historical retrospective shows, micro- and nanofluidic research and technology is driven by exploiting the physics of fluid flow in these scales to perform assays, reactions and analysis faster, at low cost, in parallel but at a single-molecule, -cell and organism resolution. Several of these technologies have also developed large markets and commercially viable products.

16.3 Dominant Phenomenon and Micro–Nanofluidic Design

As an object, structure or device is miniaturized its surface area increases in proportion to its volume, which can be quantified through the surface-to-volume (S/V) ratio. Consider a chamber that is $1\,\mathrm{mm} \times 1\,\mathrm{mm} \times 1\,\mathrm{mm}$ in length, breath and height. The volume of that chamber is $1\,\mathrm{mm}^3$, its surface area is $1\,\mathrm{mm}^2$ and its S/V ratio is $1\,\mathrm{mm}^{-1}$. Now when its size is proportionately reduced to $1\,\mu\mathrm{m} \times 1\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$ its volume reduces to $10^{-9}\,\mathrm{mm}^3$, its surface area reduces to $10^{-6}\,\mathrm{mm}^2$ but its S/V ratio increases to $10^3\,\mathrm{mm}^{-1}$. This increase in S/V ratio is the reason for all the interesting and dominant phenomena that are observed at the micro- and the nanoscale. Forces that depend on the surface gain in prominence over forces that depend on the volume, which typically tend to dominate at the macroscale. For instance, gravity and inertia, which are dominant in the macroscale, are not significant compared to surface tension in the micro- and nanoscale. The following section

details some of the dominant phenomena observed in micro- and nanofluidics that should be considered in design of devices.

16.3.1 Laminar Flow

One of the consequences of miniaturization and small-sized channels is that the walls of the channels are very close to each other. Since the walls are immobile, they impede the flow of fluid layers adjacent to them thereby reducing the overall velocity of flow in a microfluidic channel. The resistance that the channel provides to fluid flow is dependent on the viscosity of the fluid and channel dimensions. For a high aspect ratio channel ($w \ll h$ or $h \ll w$) the fluid resistance R, is given by

$$R = \frac{12\mu L}{wh^3}\,, \tag{16.1}$$

where μ is the viscosity of the fluid, L is the length of the channel, and w and h are the width and height of the channel respectively. This equation demonstrates the nonlinear effect of miniaturization of the channel on the fluidic resistance and consequently on the flow velocity. Therefore, microfluidic flows are usually laminar in nature. For instance, the Reynolds number (Re), which is a dimensionless parameter that is used to identify the significance of inertial forces in a fluid flow in comparison with the viscous drag forces, is usually below 100 and most often close to 1 or less [16.64]. The Reynolds number can be calculated using the following

$$\mathrm{Re} = \frac{\rho \upsilon D_\mathrm{h}}{\mu}\,, \tag{16.2}$$

where ρ is fluid density, υ is the velocity of the fluid stream, D_h is the hydraulic diameter or cross-sectional diameter of the channel and μ is the viscosity of the fluid. In macroscale flows, when $\mathrm{Re} < 2300$, the flow is considered laminar and as Re approaches 2300 the flow transforms from laminar to turbulent. The Reynolds number has a direct dependence on the dimension of the channel through the hydraulic diameter and an indirect dependence through the velocity of the fluid stream, which is dependent on the dimension due to the drag effect induced by the walls. At high Reynolds numbers, which are normally encountered in macroscale flows, the inertial forces are dominant and this leads to turbulent mixing. However, in the case of microfluidic flows the viscous forces dominate over inertial forces and the layers of fluid that flow adjacent to each other along the axis of the channel do not convectively mix. The only mixing that happens is because of diffusion of molecules in the fluid between the individual layers that are flowing adjacent to each other.

16.3.2 Diffusion in Microfluidics

Diffusion is the process of movement of particles from a higher concentration to a lower concentration due to Brownian motion so that over time the average concentration of the particles in the entire volume is equal. It is a time-dependent process and is characterized by the equation

$$d^2 = 2Dt\,, \tag{16.3}$$

where d is the distance traveled by a particle in time t, and D is the diffusion coefficient of the particle. In the above (16.3), it can be seen that the diffusion varies as a square power of distance, which is very significant in the microscale. To understand the significance of diffusion at the microscale, consider diffusion of hemoglobin (blood) in water. The diffusion coefficient of hemoglobin is $7 \times 10^{-7}\,\mathrm{cm}^2/\mathrm{s}$ [16.65]. Hemoglobin takes 106 s to diffuse into 1 cm of water but only 1 s to diffuse through 10 μm. So diffusion, which is negligible at the macroscale, becomes important at the microscale, especially when convective effects are absent.

Therefore, in the design of micro- and nanofluidic devices, one has to contend with low Reynolds number flows, recognize its consequence on the operation of the devices and potentially take advantage of it. This property of low Reynolds number flow in microfluidics allows one to treat the fluid almost like a solid, where it can be structurally positioned and its flow exquisitely controlled. For instance, laminar flow and the dominance of diffusion has been used in order to make a gradient generator (Fig. 16.2), which has been used to study chemotaxis behavior of cells in different pH zones [16.66]. Others have used this property to separate particles of different size at the microscale [16.67] or immunoassays [16.68] that can detect small molecules in the nanomolar concentration in less than a minute. One should also recognize that if mixing of reagents is required, it may not happen spontaneously and therefore special structures that can introduce chaotic advection should be incorporated [16.69, 70].

16.3.3 Rapid Heating–Cooling

As mentioned earlier, the S/V ratio becomes important at the microscale. This means that the scaling down results in faster heat transfer and better isolation. Thermal time constant is a parameter that is used as a measure of the thermal responsiveness of a system. It can be calculated using the following equation

$$\tau = \frac{\rho C_\mathrm{p} V}{\lambda A} \propto [l^1], \tag{16.4}$$

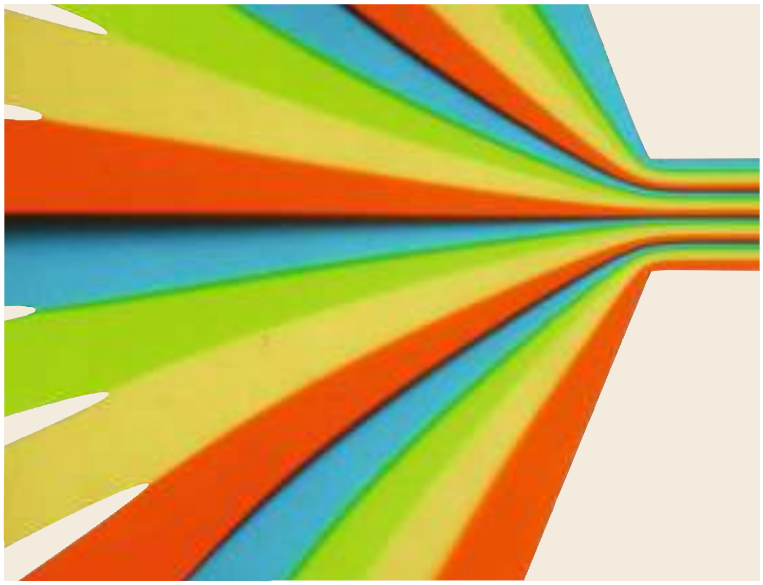


Fig. 16.2 A microfluidic gradient generator. The consequence of laminar flow results in diffusion-based mixing and is a function of time. Each zone can be clearly distinguished, and they do not mix. (Courtesy of Greg Cooksey and Albert Folch)

where τ is the thermal time constant, ρ is the density, C_p is the specific heat, V is the volume, λ is the heat transfer and A is the surface area. From (16.4) we can see that the thermal time constant is directly proportional to the length scale. So in microfluidics, where the length scale is small the thermal time constant is correspondingly smaller and the chambers or channels could be heated or cooled faster. This is particularly advantageous in the case of capillary electrophoresis where the microchannels can quickly remove heat and reduce the detrimental effects of Joule heating on resolution [16.71]. Also the sample volumes are much lower in capillary electrophoresis, which has an added advantage. Similarly, in the case of DNA amplification using a process like polymerase chain reaction (PCR) where the temperature of the chamber has to be rapidly cycled between 55 and 95 °C, the small channels help in faster heating and cooling, which reduces overall time to complete the assay [16.72]. This feature of micro- and nanofluidic devices can be used in the design of devices that require fast heat dissipation or accumulation. For instance, microchannels can be used as reactors to perform flow through exothermic chemical reactions for high-yield specialty chemical production, as the heat generated can be quickly dissipated.

16.3.4 Surface Tension

Another significant force that increases in prominence in micro- and nanofluidics is surface tension. Surface tension is due to the cohesion of fluid molecules at the surface of the liquid and air interface. For example, when a needle is placed gently on the surface of the water it does not sink due to surface tension, even though the density of the needle is much larger than that of water. However, as soon as the needle is pushed down the surface tension is overcome, and the needle immediately sinks. While cohesive forces act on the surface of the fluid, adhesive forces act at the walls, which cause capillary action. The absorption of water into a paper without any external force is due to the capillary action. Surface tension tends to bend the fluidic interface in order to balance the normal pressure difference across it. The relationship between the surface tension and the normal pressure difference across a thin fluid interface is given by the Young–Laplace equation

$$\Delta p = \gamma \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \tag{16.5}$$

where Δp is the pressure difference, γ is the surface tension and R_1 and R_2 are the radii of curvature of the fluidic interface. The equation above (16.5) clearly demonstrates that an interface can sustain high pressure differences across it when the radius of curvature, which is determined by the sizes of the channels, is small such as in the case of micro–nanofluidic devices. Equation (16.5) can be simplified for a circular channel as

$$\Delta p = \frac{2\gamma \cos\theta}{a}, \tag{16.6}$$

where Δp is the pressure difference, γ is the surface tension, θ is the contact angle and a is the radius of curvature of the tubing. For example, a circular glass

capillary with a radius of 100 nm, filled with water ($\gamma = 71.97$ and $\theta = 24°$ [16.73]) has a pressure difference of 13 090 bar. Such high pressures are possible due to the small channel dimensions in microfluidic devices.

Alternatively, molecular interaction between a solid and a liquid can lead to a curved interface due to the preference for adhesive over cohesive forces. The curved interface, once established, leads to a pressure difference across it, which then tends to drive fluid flow in micro- and nanochannels. This property is very useful in the design of microfluidic diagnostic devices for transporting the liquid sample without external forces [16.74].

Another important factor arising because of adhesive and cohesive forces at the liquid–solid interface is called wettability. Wettability is the ability of a fluid to spread on a solid surface. It depends on the adhesive–cohesive force balance at the interface and contact angle measurement is used to measure this force. By changing the surface energy (e.g., by chemical treatment), we can manipulate the contact angle. Figure 16.3 shows the schematics on how to measure the contact angle. If the contact angle is less than 90°, the surface is considered hydrophilic and when greater than 90° is called hydrophobic. The wettability of a surface can be modulated by changing its chemistry, temperature or electrical charge.

A gradient in wettability along a surface can also drive fluid flow. For instance, a droplet on a surface will bead up if the cohesive interactions among its own molecules are preferred over adhesive forces with the surface. However, when presented with a gradient of adhesive interactions with the surface, the droplet will tend to move in the direction of the gradient with stronger adhesive forces in order to spread and minimize its overall energy and achieve the lowest possible energy state. A gradient in surface tension can be introduced using electrical, chemical and thermal means [16.37]. Digital microfluidics technology uses dynamic gradients in surface tension introduced on a surface electrically to control movement of droplets on it.

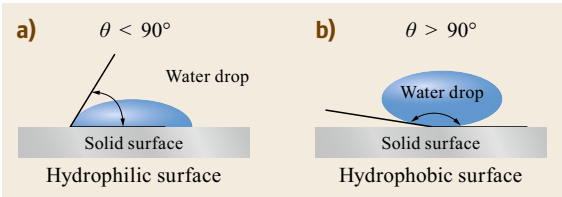


Fig. 16.3a,b Contact angle measurement of a drop. When the angle **(a)** $\theta < 90°$, the surface is hydrophilic and when **(b)** $\theta > 90°$, the surface is hydrophobic

In any microfluidic design, the influence of surface tension forces on the operation and functioning parameters of the device must be considered. For instance, hydrophilic surfaces are suitable for many microfluidic LoC devices to ensure that they fill smoothly and uniformly when samples are loaded. It is also desirable if the operating mechanism of the device exploits this dominant force in its working.

16.3.5 Electrokinetics

An interface such as between a solid and a polarizable fluid such as water develops a separation of charges known as a double layer [16.75]. This charged surface layer is insignificant in determining the properties of the flow in the macroscale. However, it may become a significant determinant of fluid flow in the microscale. Application of a pressure head across a microfluidic channel normally leads to a velocity profile that is a characteristic parabolic flow profile as shown in Fig. 16.4a. However, as discussed earlier, as the channel dimension is reduced, the flow resistance increases nonlinearly and in channels that are smaller than $\approx 1\,\mu$m, pressure-driven flow is not a suitable method. The primary reason for higher flow resistance is due to the walls of the channels that are close together and offer resistance (drag) to movement of the fluid layers adjacent to it.

In this scenario, the surface charging effects can be exploited to reduce the flow resistance. Application of an electric field along the length of the channel provides an electrostatic force on that thin layer of fluid close to the walls and causes *slip*. This phenomenon significantly reduces the drag associated with the walls and also drags the rest of the fluid in the channels as shown in Fig. 16.4b with a velocity that is uniform throughout the cross-section of the channel. This is known as the electrokinetic or electroosmotic flow. *Zeng* et al. [16.76] developed an electroosmotic pump that has no moving parts. They were able to achieve pressures over 20 bar and flow rates of 2.6 µl/min for a 2 kV applied potential. *Paul* et al. [16.77, 78] have generated high pressures using porous microstructures using the electrokinetic principle.

Similar to electroosmotic flow, the movement of particles suspended in a solution instead of the bulk fluid due to an applied electric field is called electrophoresis. Electrophoresis is a standard technique used to separate DNA based on size. By applying an electric field in a gel containing DNA strands (which have a negative charge) of varying size, smaller DNA moves faster and farther than the longer DNA strands and thus can be separated based on the size. Dielectrophoresis is another method used for separating par-

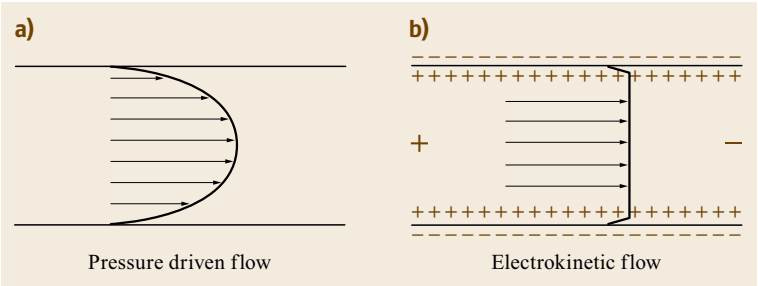


Fig. 16.4a,b Fluid flow profile for **(a)** pressure-driven and **(b)** electrokinetically driven flow

ticles. In dielectrophoresis, a particle with a different dielectric constant as compared to the fluid medium experiences a force when placed in a nonuniform electric field. The force depends on various factors such as medium, particle electrical properties, size, shape and strength of the electric field. Researchers have used this method to separate cells [16.79], particles [16.80] and droplets [16.81].

16.3.6 Geometric Control of Two Phase Flow

Since surface tension dominates in the micro- and the nanoscale, the interaction between two phases (oil and water or gas and a liquid) can be precisely controlled. This property is taken advantage of in droplet microfluidics. Aqueous droplets ($\approx$ 1 nanoliter) can be produced in an oil medium either actively (using valves [16.82] or electric fields [16.83]) or passively (geometry [16.38, 84]). Flow of two immiscible fluids such as oil and water can be brought together at precise angles such as shown in Fig. 16.5a, which leads to the production of uniform-volume droplets of one fluid in the other as the surface tension of the interface holds one fluid until the pressure buildup from the incoming flow overcomes it [16.85]. Similarly, the instabilities introduced in a coaxial flow of two immiscible phases either through a nozzle [16.86, 87] such as in Fig. 16.5b, or those that naturally occur [16.88] such as in Fig. 16.4c, lead to breakup of the coaxial flow into precise and uniform droplets and high rates ($\approx$ kHz).

Various groups have studied droplet manipulation, which includes mixing [16.90], fusion [16.91] and fission [16.39, 92]. Researchers have used these techniques on-chip for applications in sorting [16.93, 94], single-cell analysis [16.95, 96] and drug screening [16.97].

16.3.7 Nanoscale Phenomena

The S/V ratio becomes much larger in the nanoscale and in some cases that leads to breakdown of continuum mechanics in some cases and molecular interactions become dominant. For instance, nanoporous materials have been used to fabricate a Knudsen pump that can create a pressure difference across it due to a thermal gradient and with no mechanical moving parts [16.98]. The more frequent interaction of the air molecules with the walls of the nanopores than with other air molecules leads to this phenomenon.

Nanoscale channels and nanopillar arrays can be made to confine polymers and other long-chain molecules such as nucleic acids. Typically, in free solution these molecules fold themselves into a blob-like structure that is characterized by their persistence length. By forcing them into confined nanochannels, these molecules open up revealing their true linear structure. These nanofluidic channels have been used in a variety of applications including DNA sequencing and single-molecule detection [16.99].

Nanoscale channels also are so small that the double layers created due to the interaction of a polarizable fluid with the surface overlap or form a significant portion of the channel. This effect can lead to surprising phenomena such as exclusion of the co-ion of the type that is present on the walls of the channels [16.100, 101]. These devices can potentially be applied to fluidic electronics and other such applications [16.102, 103].

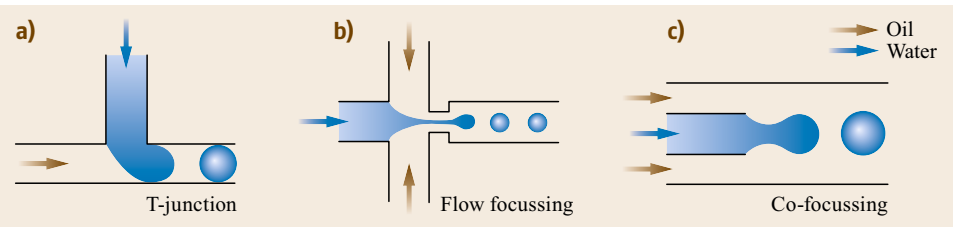


Fig. 16.5a–c Passive droplet generation geometries. **(a)** T-junction **(b)** flow focusing and **(c)** coflowing (after [16.89])

16.4 Fabrication of Micro–Nanofluidic Devices

Microfabrication technology was initially developed in the electronics industry in the 1950s and 1960s and has since been applied to construct micromechanical systems and micro–nanofluidic devices. The philosophy behind integrated circuit fabrication, which is to use common materials and fabrication methods to build multiple components so that they can be fabricated in parallel and tightly integrated, has been adopted to fabricate microchannels, microvalves, pumps, mixers, separation columns, amplification chambers and sensors, and to integrate them into lab-on-chip (LoC) and micro total analysis systems (μTAS). These systems can perform the entire sample processing and detection on a small form factor in a automated manner with high efficiency, increased throughput, lower cost, small reagent consumption, shorter analysis time and high sensitivity. Initially, silicon and glass were the main materials that were used for microfabricating these devices primarily because they had well-developed and mature microfabrication processes. However, unique requirements for biological assays such as disposability, need for surface functionalization and the need for inertness and biocompatibility provided impetus for the development of microfabrication methods in polymeric materials. Since the mid-1990s, fabrication methods have been developed for a range of elastomers, thermoplastics and thermosetting materials. More recently, the demand for devices that could be operated by minimally trained personnel for diagnosis purposes in resource-poor settings led to the development of fabrication methods to build microfluidic devices utilizing paper and more recently threads. Widespread availability of these materials, their low cost and simplicity of their fabrication methods and ability to transport fluids without any external power sources has led to rapid growth of this research field and a large number of functionalized paper- and thread-based microchips for detection and analysis of a variety of analytes have been reported during the past ten years. Microfabrication methods for hydrogels have been developed more recently due to interest in building cell culture scaffolds for tissue engineering applications The following sections describe the most extensively used materials in microfluidics along with their fabrication and integration methods as well as applications in biological investigations. Readers are encouraged to refer to the cited literature and subject-specific review papers for more detailed explanations of each process.

16.4.1 Silicon and Glass

Silicon was the first material adopted in microfluidics owing to the origin of microfluidics, which was the semiconductor industry that relied extensively on silicon as the main material abundantly available in the Earth's crust. Several techniques such as bulk micromachining, surface micromachining and buried channel methods have been developed for fabrication of silicon-based microfluidic devices [16.9, 104, 105].

In bulk micromachining, a photolithographic process is employed to pattern desired microfeatures on a substrate (typically silicon). Two different etching methods based on the etchant phase, i. e., wet (liquid) and dry (gas), have been used to remove silicon from the wafer and shape the microstructures. After etching, another substrate such as glass or silicon is bonded to the patterned piece to from enclosed channels, chambers and other features [16.106]. Alternatively, in surface micromachining, a thin film of a sacrificial material is deposited and patterned on the substrate to form the scaffold of the microfluidic structure. The structural layer is deposited next forming the side walls and the top surface of the microchannel or microfluidic chambers. Next, the reservoir areas are etched in the sacrificial layer to expose the underlying sacrificial material. Then the sacrificial layer is etched away to form the designed channel on the base substrate [16.107]. The sequential deposition of sacrificial and structural materials and their patterning can be continued to form multilayer microfluidic structures. SiO_2 and Si_3N_4 are the two most commonly used sacrificial materials while polysilicon (poly-Si) is mostly used to form the structural layer. Another process used in silicon micromachining that combines the bulk and surface micromachining is the buried channel technique. It is a multistep, bonding-free approach that involves patterning a thin narrow line on the substrates corresponding to the region where the microfluidic channel or structure is to be formed. This pattern is used to anisotropically etch a narrow trench. Subsequently, the thin nitride layer is deposited conformally to cover the side walls and the bottom of the trench and then etched anisotropically so that the nitride at the bottom of the trench is removed while that at the side walls form a protective coating. Next, the exposed bottom is etched isotropically producing a enlarged cavity buried into the substrate, which can form the channel structure. Finally, the nitride is stripped and sealing is achieved by deposition of another layer of a suitable material (poly-Si, SiO_2, Si_xN_y) conformally to a thickness that is more than half of the original trench width, which will enclose the channel. This method also enables the construction of channels at different depths by adjusting the depth of the trench [16.108–110].

Well-established fabrication techniques, the ability to obtain smooth sidewalls and anisotropic etches and the availability of several methods for surface modification were the main reason for the initial extensive use of silicon as the substrate material in microfluidics [16.111]. Silicon-based devices have been used in a wide range of applications from micropumps [16.112] and valves [16.113] to microreactors [16.114] and cell growth chambers [16.115]. Nevertheless, cost of the material and its opacity made it not ideal for biomedical and lab-on-chip applications [16.116, 117]. Moreover, high electrical and thermal conductivities of silicon necessitate employing electrical and thermal insulation layers in some applications, which increases the cost of fabrication [16.118]. These factors led to the development of microfabrication methods for more suitable substrates such as glass [16.119].

Glass is an amorphous silicon compound that can be used in microfabrication. Glass compared to silicon offers superior optical transparency and electrical insulation properties. Additionally, it is chemically inert, shows low fluorescence emissivity and is compatible with biological samples [16.111] as many of the biochemical assays and processes have been developed on Petri dishes and glass containers in the macroscale. Hence, it became the material of choice after the initial interest in silicon devices. The most widely used glass is 7740 Corning Pyrex due to its high thermal compatibility with silicon that results in smaller bending mismatches when high temperature is required for bonding [16.120, 121]. Soda lime glass and FOTURAN are two other types of glass that have proven to be useful in fabricating microfluidic devices.

Methods of fabricating glass microchannels are not as diverse and advanced as silicon [16.111]. Glass microchannels were mostly fabricated by mechanical, wet or dry etching approaches. The mechanical technique is similar to traditional drilling and often produces rougher surface finish [16.122]. Among the chemical methods, wet etching is the most common and can be used to fabricate isotropic structures [16.105]. It involves removing glass from the substrate by masking and submerging it in a selective liquid etchant like hydrofluoric acid, potassium hydroxide and tetramethylammonium hydroxide [16.123, 124]. The etchant used in this method and the glass type are two parameters that determine the etch rate. Adding some chemicals such as hydrochloric acid and HNO_3 can also be effective in tuning the etch rate [16.125]. The desired depth can be achieved by adjusting the etch rate and the duration of etching. Due to its isotropic nature, the aspect ratios obtained with this method is ≈ 1 or less. In dry etching, it is the momentum of ion species with the help of a masking process that removes the maskless glass. Alternatively, etchants can be introduced in the gas phase, ionized and provided momentum through an electric field in order to provide directionality to the etch and increase anisotropy. Dry etching has a significantly slower etch rate compared to wet etching [16.105]. After etching and definition of the microfluidic features using bulk micromachining, the substrate is bonded to another glass or silicon substrate using a variety of methods including fusion bonding, adhesive bonding and anodic bonding, which have been developed for bonding glass to glass and glass to silicon wafers [16.126, 127].

The intrinsic properties of glass make it a suitable choice for biomedical and biological fields. Its transparency also facilitates optical bioanalytical assays [16.128]. Therefore, glass microchannels have been extensively used for electrophoresis [16.129], polymerase chain reactions [16.72], and gas chromatography [16.130]. However, several shortcomings, such as relatively high cost of fabrication and lack of an easy technique for bonding glass to another substrate, limit its applications as the main material for microfabrication of microfluidic devices. Additionally, glass is not permeable to gases and cannot be used in more complex applications of microfluidics such as analysis of cells in which permeability to gases is vital [16.131]. Consequently, new materials needed to be explored to conduct more elaborate research experiments in microfluidics.

Although silicon and glass were the first materials that were used for microfluidic devices, they are now replaced by polymers to a great extent due to higher biocompatibility, easier fabrication process, lower fabrication cost, flexibility, permeability to gas, and good optical transparency of polymers [16.132]. However, silicon and glass are still important in applications where thermal, chemical and mechanical stability is required. For instance, rigid walls that are needed in nanoscale channels can be constructed by silicon or glass due to their high mechanical stability [16.131, 133].

16.4.2 Polymers

Drawbacks of silicon- and glass-based microfabrication such as relatively high cost of fabrication, complex and cumbersome fabrication procedure and silicon being optically opaque [16.22, 134], drove researchers towards seeking new materials. Polymers are among the most suitable for use in microfluidic devices, especially in biomedical applications. They offer desirable properties such as modifiable thermal, electrical and surface characteristic properties as well as low cost, chemical inertness, biocompatibility, and transparency that make them a proper choice for disposable bioassay devices [16.118]. Over the last decade,

polydimethylsiloxane (PDMS) has been the dominant polymer used in prototyping of microfluidics and LoC devices [16.135, 136]. Nevertheless other thermoplastic and thermosetting materials such as polycarbonate (PC), polystyrene (PS), cyclic olefin copolymer (COC), poly(methyl methacrylate) (PMMA) and polyimide (PI) [16.137, 138] have been used in commercial microfluidic devices due to their chemical inertness, minimal adsorption and ease of fabrication especially using standard processes like injection molding and hot embossing.

The advent of polymer casting methods (e.g., soft lithography) was an important step towards widespread prototyping and experimentation of LoCs [16.139, 140]. PDMS is the most extensively used polymer in soft lithography (Fig. 16.6) for fabrication of microfluidic devices [16.141, 142]. It is among the low-cost polymers that are transparent and permeable to gases, and its intrinsic mechanical properties render it the first choice for many microfluidic applications. Another advantage of this material is that it can replicate features in molds that are a few tens of nanometers in dimension [16.136]. This elastomeric polymer is commercially available in a kit that contains the base polymer and a curing agent, which is used to crosslink the polymer. The amount of the curing agent in PDMS defines the hardness of the final product. A larger ratio of the curing agent to base polymer results in a harder material. In the soft lithography method, photolithography or chemical etching of substrate is used to develop a negative replica of a desirable design called a master mold. PDMS base to curing agent mixture (typically 10 : 1 ratio) is then cast over the master followed by curing at 65 °C for 1–4 h to allow crosslinking of the prepolymer. Due to the low surface energy of PDMS, it can be peeled off from the master mold with ease. The cured PDMS needs to be bonded to another surface to form an enclosed channel. Various methods such as plasma treatment (most common) [16.144, 145], mechanical bonding [16.146] and chemical treatment [16.147] can be used depending on the application of the device and working pressures. The plasma bonding method includes exposure of surfaces to oxygen or other gases' plasma to generate free radicals on surfaces for bonding of PDMS with a variety of materials including PDMS, glass, parylene, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, and glassy carbon [16.139, 148].

Polymers in microfluidics are not limited to elastomers and thermoplastics are also commonly used. PMMA, as a low-cost polymer with superior mechanical strength, electrical insulation and transparency, can be a representative of this type of polymer. PMMA also shows high biocompatibility and is used in dental fillings [16.149], contact lenses [16.150] and bone cements [16.151]. Therefore, it is a suitable candidate for biomedical and disposable devices [16.152]. Microfluidic chips that exploit PMMA as their main manufacturing material are typically fabricated using a hot embossing technique [16.153] and based on their desired application can be used as the master molds for the subsequent soft lithographic process or as the microchannel itself [16.154, 155]. In the hot embossing method, a mold containing the features to be embedded into the plastic in negative is heated above the glass transition temperature of the plastic and pressed against a thin sheet with a significant pressure. The mold transfers heat to the plastic, which upon reaching temperatures above its glass transition temperature, flows to conform to the mold's topography. Subsequently, the mold and the plastic is cooled and the plastic removed to emboss the features permanently. Mass production of PMMA devices can be achieved via a microinjection molding method that facilitates shorter fabrication cycle times [16.156]. A variety of techniques can be found in the literature for bonding and integration of PMMA. Thermal bonding requires heating the PMMA channel layer up to glass transition temperature and pressing it over a cover plate [16.157]. In solvent bonding, which is conducted at room temperature, a solvent (plasma-enhanced ethanol [16.158], low azeotropic solvent [16.159], 2,4-pentadione [16.160]) is applied at the interfaces to produce a strong bond [16.161]. Although PMMA as a material presents several favorable proper-

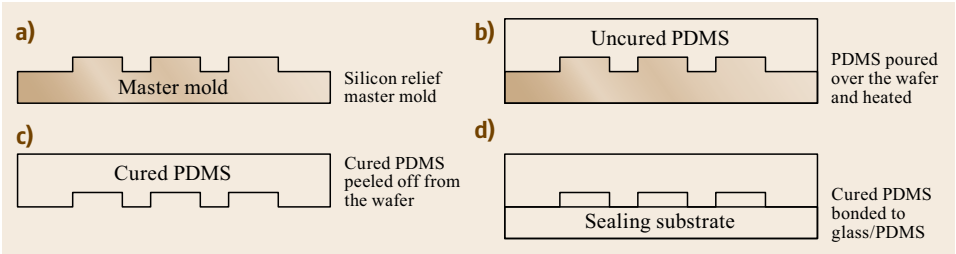


Fig. 16.6a–d Soft lithography method for replicating PDMS over a master mold. Uncured PDMS is poured over the mold and heated followed by bonding to another substrate to make closed microchannels (after [16.143])

ties that are ideal for biomicrofluidics, it is hydrophobic and presents a challenge in its use in devices that rely on capillary flow for sample transport. This shortcoming has been overcome by adding ionic charged surfactants or dynamic modification to form amine-terminated surfaces or by bulk copolymerization that results in modification of PMMA during the manufacturing process [16.162–164].

Polyimide (PI) is a thermosetting polymer that was initially introduced into microfluidics as a membrane for micropumps and microvalves [16.165, 166]. PI possesses high tensile strength, superior chemical stability and low water absorption. Its high resistance to chemicals along with high dielectric constant and excellent biocompatibility make it more suitable than other polymers for biomolecular applications [16.167, 168]. Fabrication methods used for making microfluidic structures include thermal injection molding [16.165] and, in the case of photosensitive PI, photolithography [16.167, 168]. Several methods can be employed for bonding PI to another substrate, which is commonly glass or PI. *Glasgow* et al. [16.169] reported solvent bonding as a technique for enclosing PI-based open channels. This method can exert a significant amount of stress on the device because of solvent evaporation. In another study, *Bayrashev* and *Ziaie* [16.170] used radio-frequency dielectric heating to achieve PI-PI bonding that led to bond strength of 1.5 MPa. *Min* and *Kim* [16.171] proposed thermomechanical bonding by pressing a three-dimensional PI-based device between two glass slides at 300 °C and 10 kPa for 1 h to achieve pressure tolerance of 70 bar. PIs are ideal in applications in which mechanical and electrical properties are important, such as electrophoresis and cell sorting and separation [16.172–174] where PI composites have been used to enhance its properties. The use of PI in microfluidics is still nascent but is likely to emerge as an important material with the advent of new additive manufacturing methods.

The broad spectrum of applications for polymers, the increasing demand for materials with specific properties that suit various research topics, and the compatibility of polymers with specific manufacturing processes have encouraged researchers to explore newer polymers that could be integrated in microfluidics devices. For instance, polystyrene (PS) is a synthetic thermoplastic polymer that is optically transparent and chemically inert. It has been used widely in conventional cell culture and the majority of cell culture and coculture protocols have been developed on this substrate. Recently, hot embossing-based microfabrication techniques have been developed for PS so that microfluidic devices can be used to perform cell-based assays and molecular biology studies [16.175, 176] using the same protocols as in macroscale Petri dishes. Despite its advantages, PS has low melting temperature and is attacked by organic solvents and aromatic hydrocarbons and hence not suitable for other chemical assays or microreactors. Polycarbonate (PC) is another polymer that has been used in microfluidic devices. It is preferred over PMMA in applications where high tolerance of operating temperatures is required and is commonly used in bioanalyses that include polymerase chain reactions [16.177, 178]. More recently, thermoplastic materials such as cyclic olefin copolymer (COC) have been introduced into microfluidics. COC has high moisture barrier properties and excellent optical properties that make it suitable for applications where fluorescence-based detection is required. Their glass transition temperatures also exceed 200 °C and therefore they can be used in assays that require heating and are compatible with autoclaving. These characteristics make it an appropriate choice for biodetector LoCs that utilize integrated circuits [16.179, 180]. Nevertheless, it cannot be used for pharmaceutical studies due to its hydrophobic surface that increases protein adsorption and cell adhesion [16.181].

16.4.3 Paper

Paper has been widely used as a material to develop diagnostic tools such as dipsticks and pregnancy tests [16.182, 183]. In 2007, Martinez et al. used a photolithographic approach to define hydrophobic barriers in paper to direct fluid flow and transport samples passively using capillary forces. They developed various bioassays using this format and called it microfluidic paper-based analytical devices (μPADs, Fig. 16.7a) [16.143]. The primary objective of μPADs is to offer easy-to-use, portable, and affordable devices that can be used for multi-analyte semiquantitative analyses [16.187, 188]. Ubiquity and low cost of paper along with its easy adaptability with chemical and medical applications make it suitable for portable and low-cost microfluidic systems [16.188]. Despite these advantages, μPADs suffer from risk of handling contamination [16.189] and long analysis times due to the slow capillary-based transportation of fluids [16.190], which can lead to evaporation of solvents and chemical instabilities.

Over time, a wide variety of methods have been used for fabrication of μPADs including wax printing [16.184, 191] (Fig. 16.7b), flexographic printing [16.185] (Fig. 16.7c), inkjet printing [16.192, 193], screen-printing [16.194], wax screen-printing [16.195], paper cutting [16.196, 197], photolithography [16.143, 198], laser treatment [16.199, 200], and plasma treatment [16.201]. Most of these techniques, except paper

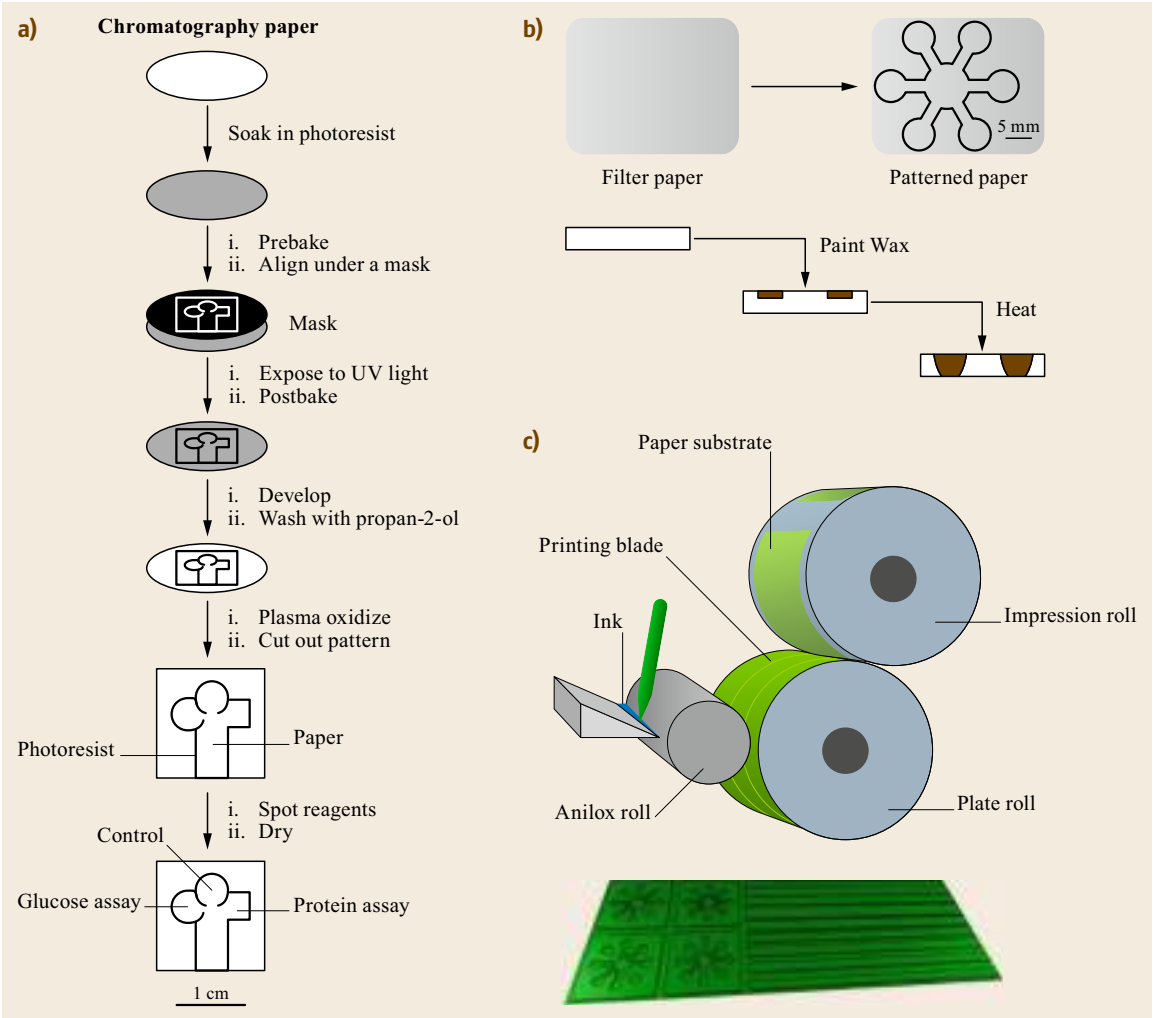


Fig. 16.7 **(a)** Paper-based microfluidics was first proposed by *Martinez* et al. [16.143]. They used a chromatography papers soaked in photoresist to be exposed under ultraviolet (UV) light to form the hydrophobic barrier required for fluid transport (after [16.184]). **(b)** In th wax printing method, wax needs to be printed on a hydrophilic paper followed by heating that helps the wax to penetrate into the paper and make the boundaries of the channel (after [16.185]). **(c)** A Flexographic unit that enables large-scale fabrication of μPADs. Rotation of impression role on which the paper substrate is fixed against the plate role that contains the pattern transfers the desired pattern onto the paper. Penetration depth is controlled by the pressure between impression and plate roles (after [16.186])

cutting, rely on patterning paper with hydrophobic materials to define the microchannel boundaries and confine the fluid flow. Early devices were made using photoresist as the hydrophobic barrier. However, higher cost of the photoresist ($\approx \$0.1$ for $100\,cm^2$) [16.202] compared to other materials made it unsuitable for μPADs that were meant to be low cost. Thus, new materials such as wax ($\approx \$0.01$ for $100\,cm^2$) [16.203] started to be used as the hydrophobic barrier. Lower-cost methods included inkjet printing of alkyl ketene dimer (AKD) ($\approx \$0.00001$ for $100\,cm^2$) [16.193] onto paper followed by polymerization by heating. Readers are encouraged to review the papers by *Xia* et al. [16.187] and *Yetisen* et al. [16.204] for further details about other paper microfluidic fabrication methods.

Paper-based microfluidic devices have been widely used in biomedical, environmental and food contamination applications. Detection in μPADs is based on the reaction between specific reagents functionalized on paper and target analytes, which produces a signal. The output signal can be detected by colorimet-

ric [16.205], electrochemical [16.206], chemiluminescent [16.207], electrochemiluminescent [16.208], fluorescent or photo-electrochemical [16.209] readouts. These methods differ in fabrication techniques, time and limit of detection, cost, and range of application. Colorimetric detection is more commonly compared to other methods, because it depends on visual or imaging-based readouts of the change in color intensity of a detection spot due to presence of target analytes and their concentration. Other detection techniques are more accurate and sensitive, but they require access to signal detection equipment such as electrochemical analyzers or fluorescent microscopes.

μPADs have also been used in environmental applications where precise detection of contaminants in surface water, soil or air is needed. For instance, *Mentele* et al. [16.210] developed a device that could simultaneously detect Fe, Ni and Cu in combustion ash. Many other strategies for detection of metals in water, soil and aerosol using μPADs have been proposed by other research groups [16.211–215].

The food industry has benefited from μPADs designed for detection of foodborne contamination. Detection methods in this area mostly rely on immunoassays or enzymatic reactions that occur in the detection zone. Using swab sampling followed by culturing in the appropriate media, *Jokerst* et al. [16.216] could detect *Listeria monocytogenes*, *E. coli O157:H7*, and *Salmonella typhimurium* in meat. Detection of Pb and Cd in beverages, dichlorvos (DDV) from vegetable skin and *E. coli* in ground beef and cucumber are other successful examples that show the capability of using paper-based microfluidics for affordable detection of contaminants in the food industry [16.217–219].

Although paper-based microfluidics is still in its early stages, its advantages such as being low cost and lightweight make it an appropriate choice for fabrication of diagnostic devices in less industrialized countries. Challenges that need to be immediately addressed are evaporation of reagents and precise control of flow along the porous media that can set the stage for development of the next generation of complex and efficient μPADs [16.220].

16.4.4 Threads

One of the important goals of microfluidics technology is to design low-cost, easy-to-fabricate, portable and disposable chips for bioassays and diagnosis applications. This has resulted in utilization of a few accessible materials such as paper and threads in microfluidic devices. Passive fluid transport along a thread is achieved by dominant capillary forces inside the pores of the fabric and along the interfiber cavities [16.221, 222]. Major advantages of thread over paper are the ability to develop 3-D devices, no need for micropatterning to produce fluidic paths, and availability of various type of threads with different wicking rates. However, lack of uniformity in the interfiber gaps and heterogeneities on the surface can cause limitations in the operation of the device and result in wicking rate inconsistencies [16.223–225].

It is essential to be able to predict the operation of the device and the fluid behavior in order to design efficient chips. One of the determinative elements in designing thread-based microfluidics is wicking rate. Modeling the wicking rate phenomenon in threads is a highly complex task that requires consideration of various parameters such as the size of the interfiber gaps, hydrophobicity and hydrophilicity, contact angle, viscosity, and tortuosity. However, it is believed that the Lucas–Washburn equation can be used to develop mathematical models to estimate the wicking rate [16.226]. Wicking rate has been comprehensively studied by *Wang* et al. [16.227] using a capillary rise method. They used the Hagen–Poiseuille equation and considered the effect of gravity to arrive at the following formula

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{R_\mathrm{D}^2}{8\eta h}\left(\frac{2\gamma\cos\theta}{R_\mathrm{S}} - \rho g h\right), \tag{16.7}$$

where γ, η, ρ and g denote surface tension, viscosity, density and gravity acceleration respectively while θ is the advancing contact angle, R_S shows mean static radius of the pores and R_D represents mean hydrodynamic radius. Similar equations can be found in the work of *Ballerini* et al. [16.228]. *Bhandari* et al. [16.229] showed that twist frequency and weaving coverage area of the yarn can be changed in order to adjust the desired wicking rate. *Safavieh* et al. [16.230] assumed that only 30% of the yarn was satiated with fluid and could propose a simple formula for the flow rate in hydrophilic yarns that obey Washburn's equation. They also showed that by increasing the wetting length, evaporation becomes important and proposed an equation for the volumetric flow rate of water, considering the saturation temperature, dew point and latent heat of water. Further details on models for prediction of wicking rates are reported by [16.231–234].

Natural threads are mostly coated with wax to increase their strength but this makes them unsuitable for microfluidic applications [16.230] since the presence of wax on the surface decreases the hydrophilicity of the material thus resulting in wicking rate reduction. Therefore, wet and dry methods are used to increase the wettability and hydrophilicity of the threads. Dry

methods are mostly based on plasma treatment while wet approaches hinge on mercerization.

Li et al. [16.225] performed a series of experiments to investigate the effect of plasma treatment as a dry method on wettability. Samples were analyzed with x-ray photoelectron spectroscopy (XPS) and it was shown that surfaces were highly oxidized and polarized, resulting in hydrophilization of the thread. *Rahel* et al. [16.235] used two alumina plates with discharge and induction electrodes on upper and lower plates and N_2, air, CO_2 and Ar as plasma gases in a surface dielectric barrier discharge system to hydrophilize polypropylene nonwoven fabric (PPNF). The authors concluded that N_2 was the most suitable gas to hydrophilize the fabric. Plasma treatment is reported by several other groups as a successful method for increasing the wicking rate [16.228, 236, 237].

Nilghaz et al. [16.238] utilized a wet method to increase wettability of unmercerized cotton cloth as their basic material by performing scouring treatment with NaOH and Na_2CO_3 (known as soda ash) to remove noncellulosic contamination. Briefly, the process included boiling the cotton for 5 min in the solutions followed by recovering the pH of the cotton by washing it with water and drying for 2 h. XPS, scanning electron microscopy (SEM), energy-dispersive x-ray spectroscopy (EDS) and light optical microscopy tests were carried out to compare the wicking rates of treated cottons and raw, untreated ones. Na_2CO_3 treatment was shown to give higher wicking rates. *Bhandari* et al. [16.229] investigated the effect of boiling tabby, crape and silk fabric in blocking solutions and demonstrated that the treated fabrics are capable of wicking colored dye solutions. Scouring treatment with a variety of solutions is also reported by other research teams to be a low-cost method for dewaxing and increasing the wicking rate [16.229, 232, 239].

Passive thread-based splitters and mixers were presented in the work of *Safavieh* et al. [16.230]. Mixing and splitting were attained by making handmade knots. Mixing ratios were analyzed quantitatively and it was demonstrated that different ratios of mixing can be achieved by using different knots.

Several scholars have tried to use various forms of textile-based microfluidics as chemiluminescence and electrochemiluminescence detectors [16.240, 241], electrochemical sensors [16.242], immunochromatographic assessment tools [16.243] and means of energy harvesting [16.231]. Thread-based microfluidic chips have also shown promising results in biomedical assays. *Agustini* et al. [16.244] showed the capability of a cotton thread-based microfluidic device for determination of acetaminophen and diclofenac in phosphate-buffered saline by a multiple pulse amperometry method. They used cylindrical graphite (used in mechanical pencils) with 0.5 mm diameter as working and reference electrodes and 0.7 mm-diameter graphite as a counter electrode. The cylindrical graphites were simply bonded to the device using a double-sided tape with a hydrophilic gauze deposited on top of the electrodes. *Nilghaz* et al. [16.245] introduced a device with a polyester substrate that could be used for blood typing by a chromatographic separation of blood. The authors immersed the threads in the antibody and dried them followed by addition of blood. Red blood cells then agglutinated and blood phases were separated resulting in creation of a band that could be easily visualized by the user. This method was reported to be a rapid method (< 1 min) that can be used in impoverished regions of the world. In this approach, shortcomings of paper-based methods such as higher blood volume consumption and substrate absorbency are resolved.

As a demonstration of the capability of performing electrochemical detection and separation by electrophoresis, *Wei* et al. [16.246], inspired by a dulcimer (a musical instrument), designed a chip to perform electrophoresis and electrodetection of $K_3Fe(CN)_6$ solution (Fig. 16.8). Their device consisted of polyester threads that acted as a route to transport the fluid while electric contacts were placed on protruding nuts as detector elements. The authors also showed that electrical current can be ten times higher in plasma-treated threads compared to nontreated threads.

All these methods display the capability of textile microfluidics as affordable platforms for diagnostics applications. However, we are still at the early stage of this technology and there remains further investigations to observe the capacity of using these devices in wearable clothing [16.223].

16.4.5 Hydrogels

The importance of using engineered materials that resemble characteristics of tissue environments for applications such as synthesizing extracellular matrices (ECM) brought up the idea of integrating hydrogels in microfluidics [16.247, 248]. Hydrogels are crosslinked polymers that are shaped from hydrophilic homopolymers, copolymers or macromers [16.249]. Heating, exposure to UV light and using chemical chelators are common methods to have the hydrogel chains crosslinked to form desired networks [16.247, 249]. Hydrogels have been used in many applications such as cell-to-cell interaction, drug delivery, artificial tissue constructs, and regenerative medicine [16.248–255]. Natural (collagen, agarose, alginate, chitosan, and gelatin) and synthetic (poly(acrylic acid), poly(ethylene oxide)) hydrogels, being optically transparent, can fa-

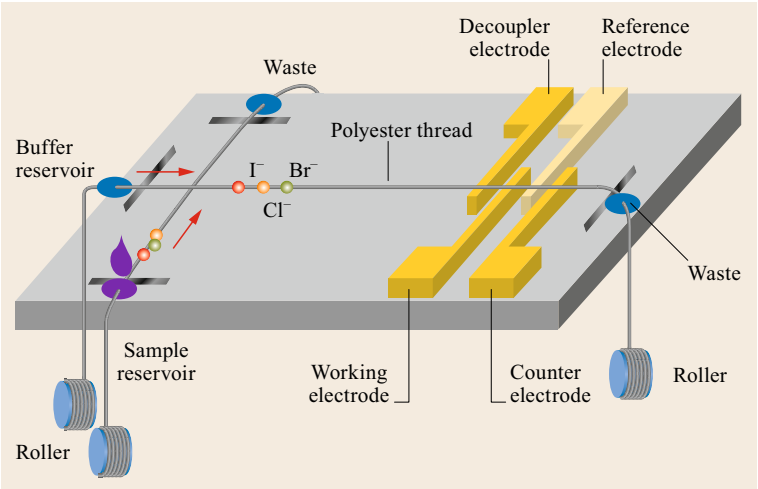


Fig. 16.8 Polyester thread-based device similar to the concept of a dulcimer designed for electrophoresis separation and electrochemical detection of ion samples and biological samples. Polyester threads act as conveyors of the solution to the detection zone where electrodes transmit signals associated with the detected sample (after [16.246])

cilitate free diffusion of molecules due to the presence of water-filled spaces between the polymeric chains [16.256–258]. Hydrogels are also nontoxic and can be accurately size calibrated [16.247]. In order to choose the appropriate hydrogel for a specific design, various properties of the gel such as swelling, porosity, diffusive characteristics, elasticity, and viscosity can be determinative [16.259–261].

Several methods are used for the integration of gels in microfluidics including a soft lithography-based approach with sacrificial or reusable templates, photopolymerization, and local integration. Briefly, soft lithography with reusable templates is very similar to conventional soft lithographic methods explained in the polymers section. Liquid hydrogels can be poured over a designed mold and solidified to from two-dimensional (2-D) or quasi-2-D structures in the gel [16.262, 263]. Although this method is straightforward and can be simply carried out for microfabrication, sealing and enclosing the channel is still a challenge in this approach. Therefore, sacrificial templates, that do not need sealing, can be used to form 3-D architectures in gels. A 3-D degradable template needs to be covered with the desired hydrogel to construct the replica of the template. The template can then be sacrificed to obtain the final device [16.186, 264–266]. This approach is typically followed when precise dimensions are not required such as studies of vascular systems (Fig. 16.9a). Another fabrication method is the local integration approach in which researchers take advantage of the laminar nature of flow in microfluidic systems. Various gels are co-introduced in a microchannel and due to the low Reynolds number, these flows form a laminated structure (Fig. 16.9b). Hence, the microchannel is subdivided into multiple layers separated by microslabs of hydrogels. The process is followed by solidification of the hydrogel using aforementioned methods [16.267, 268]. In the photopolymerization technique, for integrating microstructures into the microsystem, a UV light is employed to construct gel structures with high spatial resolution. Three types of photopolymerization are commonly used to form the structured gel components in a device: uniform exposure, printing with a mask and directed writing. High spatial resolution and high aspect ratio features are achievable using this method [16.248, 269–272].

Hydrogel technology is mainly applied to biological studies. Hydrogels have been very successful in tissue engineering as a means of mimicking native tissue architectures. *Miller* et al. [16.186] used a cytocompatible sacrificial template to construct a 3-D vascular network that could be lined with endothelial cells. *In-vitro* cell cultures that can be related to animal studies, in situ cardiac microtissues with real-time stress measurement, and bone-mimicking structures to study tumor cell migration are other examples that represent the capability of hydrogels in cell and tissue engineering studies [16.253, 273, 274].

Hydrogels are also being increasingly explored in organ-on-chips that are *in-vitro* devices in which microfabrication technologies are employed to develop microenvironments for coculturing of tissue- and organ-specific cells to mimic *in-vivo* physiology. Hydrogels could facilitate fabrication of 3-D cell cultures to perform more complex organ-on-chip studies in tissue microenvironments [16.57]. In one study, *Li* et al. [16.276] made a hepatic microtissue using PEG hydrogel that exhibited hepatocellular activities. These microtissues were then introduced in a microchannel for drug screening and could prove the capability of using hepatic microtissues for detection of hepatotoxicity and drug-drug interaction. Cell–cell interactions between human

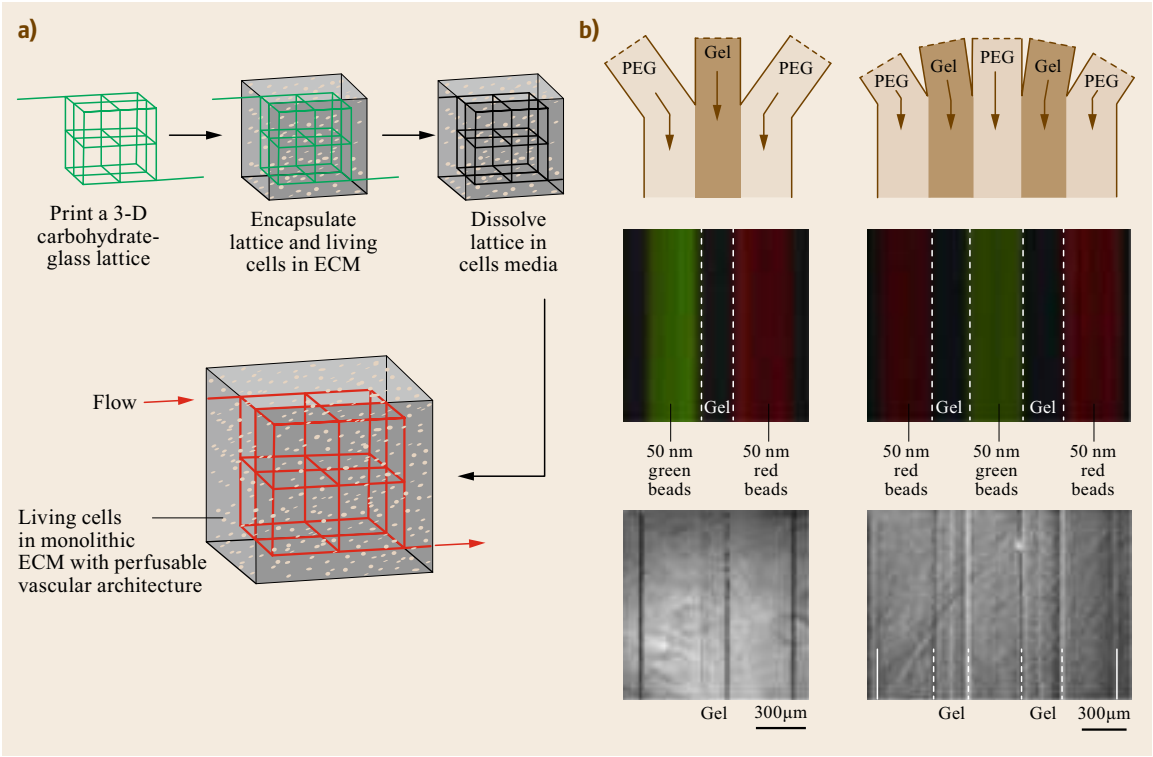


Fig. 16.9 **(a)** 3-D vascular network that was fabricated using a 3-D printed sacrificial template followed by encapsulation in ECM and dissolving in lattice cell media. (Reprinted from [16.267] with permission from Elsevier.) **(b)** Formation of microslabs of hydrogel on colaminar flows in a microchannel. (Reprinted from [16.275] with permission from Elsevier)

mesenchymal and endothelial cells utilizing fibrin hydrogel, analysis of cytotoxicity of anticancer drugs in 3-D structures in which cells were cultured in hydrogels to simulate liver, tumor and marrow, are other examples that show the potential of incorporating hydrogels in organ-on-chip studies [16.277, 278]. Further details on methods of integration of hydrogels in organ-on-chips can be found in more specific review papers [16.279].

16.5 Applications

The principles of micro–nanofluidics have been applied in a wide variety of areas including environmental sensing [16.280], medical diagnostics [16.281], drug discovery [16.282], drug delivery [16.283], microscale chemical production [16.284], combinatorial synthesis and assays [16.285], artificial organs [16.57] and micropropulsion, microscale energy systems [16.286]. The applications can be broadly classified into those that have been commercialized and others that are in the research phase.

16.5.1 Commercial Applications and Their Working Mechanism

The need to control fluid flow at the micro- and nanoscale is applicable in a wide variety of application areas. Some of these areas have become multimillion dollar revenue-generating markets in themselves. In this section, some of the commercial microfluidic devices are discussed.

Inkjets

Microfluidic nozzles for inkjet printers were one of the early commercial successes of microfluidic technology in the 1980s. Inkjets replaced the dot-matrix printing technology, which was slower and more expensive. In inkjet printing, tiny droplets of ink are sequentially placed on paper in a controlled way to form an image pattern. Today the inkjet printing market is worth $900 million and expected to reach $1 billion [16.11]. The inkjet technology has expanded to other areas such as bio-ink printing [16.287], printing

plastics and conductive materials [16.288] and carbon nanotubes [16.289]. The print head consists of hundreds of micronozzles that were machined using laser micromachining. The nozzles were integrated with microfluidic channels that connect all the nozzles to a central ink reservoir and microfabricated actuators that can be individually controlled to precisely dispense the ink only from the selected nozzles. There are two types of inkjet printers depending on their mechanism of actuation. One method was known as bubble jet where tiny resistors were used to heat and vaporize the ink to form bubbles locally underneath the nozzles ($\approx 30-100\,\mu m$). The expansion of the bubble pushes the confined ink out between the bubble and the nozzle forming a drop. Removal of heating collapses the bubble and creates a vacuum that pulls in more ink from the cartridge through the microfluidic channel. The bubble jet technology was pioneered by companies like Canon and Hewlett Packard and relied on the small thermal time constants associated with the microfluidic channels and devices to form and collapse the bubble within $80-200\,\mu s$ [16.290]. There can be up to 600 nozzles that can produce tiny droplets either in a sequence or simultaneously. Another technology patented by Epson uses a piezoelectric crystal, which is placed behind the nozzle as the actuation method to produce tiny droplets. The charging of the piezoelectric crystal makes it vibrate towards the nozzle and pushes a small amount of ink through a nozzle, and when discharged, it vibrates away from the nozzle and it sucks in a tiny amount of ink from the reservoir to repeat the process (Fig. 16.10). The inkjet printers can produce 6000 drops per seconds and are capable of printing six pages per second at the highest resolution [16.290].

Microvalves

Microvalves are used for controlling the fluid volume inside a channel. They are critical for manipulating fluid flow inside microfluidic devices. Based on the type of actuation microvalves can be classified as passive and active. *Au* et al. [16.291], *Oh* and *Ahn* [16.113], *Laser* and *Santiago* [16.112] and *Ashraf* et al. [16.292] have reviewed the different kinds of microvalve in detail. The global market value for microdispensing devices was estimated as \$5.5 billion and is estimated to increase to \$7 billion by 2020 [16.293]. The primary advantage of the microvalves over their macroscale counterparts is the significantly reduced dead volume, which is useful in many applications that require precise flow control at small flow rates. Examples include microdispensing of drugs, gas flow in vacuum systems and microreaction engineering.

Passive Valves. Passive valves were among the first microfluidic valves developed and use only their structural design to allow flow in one direction while offering a high fluidic resistance in the opposite direction. They do not have any external actuation mechanism and usually rely on the input pressure level to open a valve that is closed. Check valves and capillary valves are two well known types of passive valves. They are generally incorporated at the inlets and outlets of reciprocal displacement pumps. Passive valves only open when a forward pressure is applied and thus behave like a diode.

Check Valves. Check valves use a mechanical structure that is compliant to fluid flow in one direction while becoming rigid in the opposite direction. In the case of microvalves, these mechanical structures have often been cantilevers [16.275], membranes [16.294] or flexible structures [16.295]. For instance, *Tiren* et al. [16.275] fabricated a check valve using silicon microfabrication technology as shown in Fig. 16.11. The maximum pressure that the valve could sustain was calculated as 1000 bar and the maximum cantilever deflec-

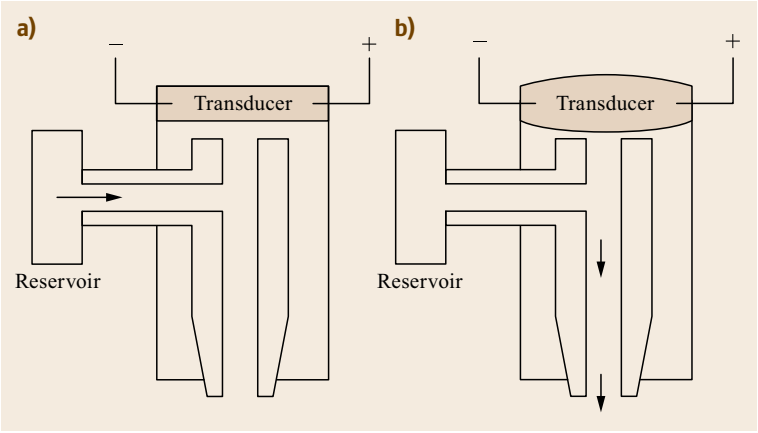


Fig. 16.10a,b Schematics of the working of a piezoelectric crystal for producing tiny droplets: **(a)** piezoelectric crystal vibrating away from the nozzle **(b)** piezoelectric crystal vibrating towards the nozzle

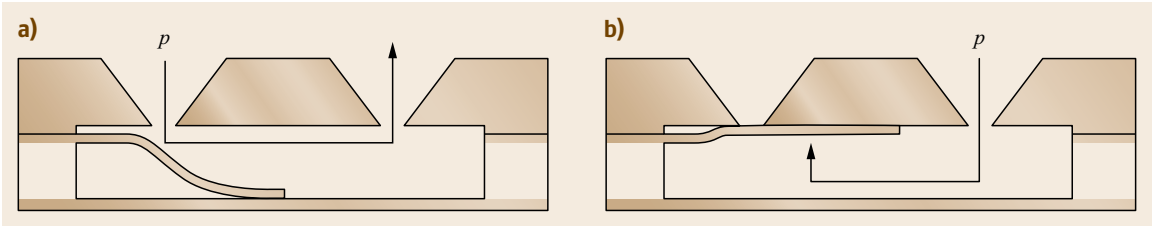


Fig. 16.11a,b Microfabricated check valve. **(a)** Forward pressure opens the valve **(b)** backward pressure closes the valve

tion was found to be 10 μm. *Hasselbrink* et al. [16.295] developed a passive valve by selectively polymerizing an ultraviolet-sensitive photopolymer in etched channels of glass substrate. The polymerized conformal plug could be moved in the valve chamber under the flow itself to provide a diode-like function, blocking the flow when its direction is from the outlet to the inlet while allowing it in the opposite direction. They showed that their passive valves could seal a pressure greater than 300 bar with actuation times less than 33 ms.

Capillary Valves. Capillary valves use the dominance of surface tension in microfluidics to perform valving action. They can be designed based on geometry [16.296] or surface property (hydrophilic or hydrophobic) [16.297]. For instance, *Yamada* and *Seki* [16.296] developed a microdispenser array made of PDMS in which nanoliter-sized droplets could be accurately dispensed. In their system, the hydrophobic and narrow channels acted as valves and were used to manipulate and transport fluid. Their geometry included a larger microchannel abruptly narrowing into a smaller channel. The pressure required to overcome the surface tension in the larger channel is much lower than that required to overcome it in the smaller channel as shown by the Young–Laplace equation. Therefore, this geometry creates a unidirection passive valve allowing liquid from the smaller to larger channel to flow freely while preventing the flow from the larger to smaller channel. Once the liquid is valved, pneumatic pressure can be further increased above the threshold required to force the fluid into the narrower channel. In this study the pressure threshold to perfuse the liquid in the larger (100 μm) and smaller (5 μm) channels was 0.001 to 0.197 bar respectively. Similarly, *Andersson* et al. [16.297] used octafluorocyclobutane (C_4F_8) patches as hydrophobic valves. They deposited the fluoropolymer (C_4F_8) lithographically in deep reactive ion etched silicon channels. The change in surface property changes the surface tension, which results in a high critical threshold of pressure required to perfuse the channel. The valve showed good sealing for water (up to 5 h) and a pressure of 0.0076 bar was required to break the seal.

Active Valves. Active valves require an external force of field for actuation. This force/field can be magnetic [16.298–300], electric [16.166, 301], piezoelectric [16.302], thermal [16.303–305], pneumatic [16.24, 272] or of other types [16.306].

Magnetically Actuated Valves. Magnetically actuated valves are usually diaphragm-based, where a large rigid structure attached to a membrane diaphragm is actuated by an external magnetic field to move it closer to the inlet or the outlet of the valve chamber and close it tightly. These valves usually have a ferromagnetic material such as nickel-iron alloy (NiFe) integrated into the boss or the diaphragm structure. Since the ferromagnetic material is strongly attracted to the magnetic field, it can be actuated by controlling a magnetic field external to the device. For instance, *Yanagisawa* et al. [16.298] developed an electromagnetically driven microvalve fabricated photolithographically. The valve consists of a magnetically suspended spring with a valve cap made of NiFe and an external coil for driving the valve cap. When a magnetic force is applied, the valve cap is pushed down closing the outlet forming a tight seal. They were able to control the valve's flow rate between 66 and 51401/min and an input pressure of 3×10^{-12} kPa. Similarly, *Byunghoon* et al. [16.299] designed a microvalve with a permanent magnet attached to a membrane actuated with an external solenoid for a glaucoma implant. The membrane was deflected using a current pulse and a pressure of 2.3 kPa was obtained for a 60 mA pulse. *Oh* et al. [16.300] fabricated a pinch-type valve for lab-on-chip application. The valve consisted of a solenoid actuator with a pinch-type plunger attached to a biomedical grade silicon tubing. Their pinch valve showed no leakage up to 200 kPa and zero dead volume.

Electrostatic Actuated Valves. Electrostatically actuated valves are also usually diaphragm based. Unlike the magnetic valves, the electrostatic ones have electrodes embedded into the membrane diaphragm and another rigid one placed close to it with the electrodes separated by a dielectric. Application of a high potential between the electrodes leads to electrostatic

attraction between the electrodes, which pulls the flexible diaphragm closer to the rigid electrode. As an example, *Goll* et al. [16.166] fabricated a microvalve consisting of three membrane layers. Two insulating layers sandwiched a gold layer of 3 μm thick and was controlled by electrostatic force by using electrodes in the upper and lower chamber separated by 25 μm. A voltage range between 60 and 150 V was required to close the valve compactly and at a pressure of 110 kPa nitrogen gas flow rate was measured as 12 ml/min. *Schaible* et al. [16.301] micromachined a silicon plate and a membrane to regulate pressure up to 1600 kPa and a response time of 1 ms. They used a miniaturized voltage amplifier with inputs between 3 and 12 V to actuate the valve.

Piezoelectrically Actuated Valves. Similar to the magnetic actuation, piezoelectric crystals that produce a mechanical stress when an electric field is applied can be embedded onto diaphragms of valves for valve actuation. These crystals are capable of producing large bending force with small displacements. Application of a potential to these crystals produces a mechanical displacement, which then moves the diaphragm and the structures connected to it against the inlet or outlet of the valve chamber to close it. As an example, *Rogge* et al. [16.302] designed a microvalve for gases and liquids that was piezoelectrically actuated. They achieved 25-fold hydraulic amplification with a stroke volume of 50 μm. Voltages up to 300 V were possible at a response time of 2 ms. The leakage flow rate was measured as 30 μl/min at a pressure of 200 kPa for nitrogen and 0.013 μl/min for water at a pressure of 100 kPa.

Thermally and Thermopneumatically Actuated Valves. Thermally actuated valves can be classified into bimetallic [16.303], thermopnuematic [16.304] and shape memory alloy [16.305] actuation. The bimetallic actuation utilizes the differential thermal expansions of different materials in a bilayer to introduce a bending moment that can actuate a boss attached to it to close the inlet or outlet in the valve chamber. Microheaters are typically embedded with the bimetallic elements to introduce local control of actuation. As an example, *Barth* [16.303] deposited 30 μm Ni film on a silicon membrane to fabricate a bimetallic valve. This microvalve functioned at 1035 kPa and had a flow rate of 1 l/min of air with an actuation time of 0.2 s. Another thermal actuation method is the use of phase change to generate a pressure in a confined chamber that can then deflect a membrane, which can seal the inlet or outlet of the valve. *Rich* and *Wise* [16.304] demonstrated such a microvalve having a cavity sealed with volatile fluid below a membrane. The cavities were partially filled with pentane to improve thermal efficiency and polysilicon heaters were placed 9 μm above the cavities. The microvalve used 500 mW to generate a pressure of 16 Pa. The leakage flow rates were measured as 1 μl/min. An alternative to bimetallic actuators is the use of shape memory alloys. *Kohl* et al. [16.305] developed microvalves based on shape memory alloy (SMA) thin films. The microvalve is in a normally open position. NiTiPd thin film coating was used as SMA (Fig. 16.12) and sustained pressure up to 250 kPa for a gas flow rate of 360 ml/min. A 110 mW power was required for NiTiPd microvalves. Figure 16.12 shows the open (a) and closed (b) positions of the SMA microvalve.

Phase change valves comprise of hydrogel valves [16.272] and paraffin microvalves [16.307]. Smart hydrogels can change their state reversibly with some external stimuli. *Beebe* et al. [16.272] developed a smart hydrogel-based microvalve that can be actuated pH. The hydrogels were photopolymerized in liquid phase directly inside the channels. A flexible PDMS membrane was coupled with the pH-sensitive

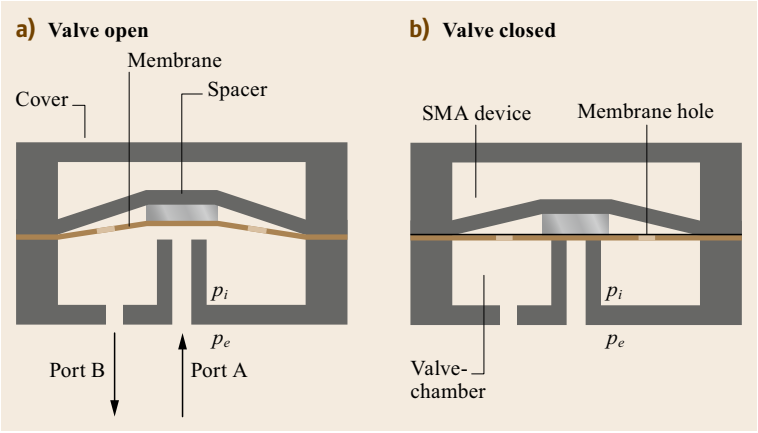


Fig. 16.12a,b Smart material alloy (NiTiPd) microvalves: **(a)** normally open position. **(b)** closed position after actuation (after [16.305])

hydrogel to expand as the pH of the solution changed to 11, sealing the channel completely. The hydrogel expansion generated an internal pressure of 184 kPa with a maximum displacement of up to 185 μm. The membrane completely shut off the channel in 19 s. *Liu* et al. [16.307] heated paraffin wax to change the phase to open and close channels. A paraffin plug was heated so that it changed phase from solid to liquid and by using external pneumatic pressure the plug was moved inside a microchannel. Upon cooling, the paraffin plug changed to solid phase, sealing the entire channel. No leakage was detected over a period of 15 min for a pressure of 1725 kPa. The time response for paraffin-based microvalves was around 2–10 s.

Pneumatic Actuated Valves. In pneumatic actuated valves a positive pressure is applied to expand a flexible membrane to compress a channel thereby sealing it. Pneumatic valves include in-plane [16.129] and inline [16.24] actuation. As an example, *Lagally* et al. [16.129] used a latex membrane for in-plane actuating pneumatic valves to develop a fully integrated PCR-capillary electrophoresis (PCR-CE) system on a glass chip. Two aluminum manifolds were used as hydrophobic vents and a vacuum was used to clamp the latex membrane and inlet ports in place. The manifolds were connected to external solenoid valves for pneumatic actuation. A vacuum was used to load the sample and transport it to the PCR chamber. The latex membrane was pressurized to 89 kPa to seal the PCR sample. *Unger* et al. [16.24] developed multilayer soft lithographically fabricated microvalves. The inline microvalve consists of two PDMS blocks with channels. The two PDMS blocks are then bonded to each other such that they form a cross structure (Fig. 16.13). The upper PDMS channel is pneumatically pressurized and it selectively closes off the bottom channel at the intersections. The channels were 100 μm wide and 10 μm in height. A pressure of 60 kPa was used to close the valve within few milliseconds.

Other Actuation Valves. Several other methods have also been used to actuate microvalves including electrokinetic and electrochemical techniques. *Neagu*

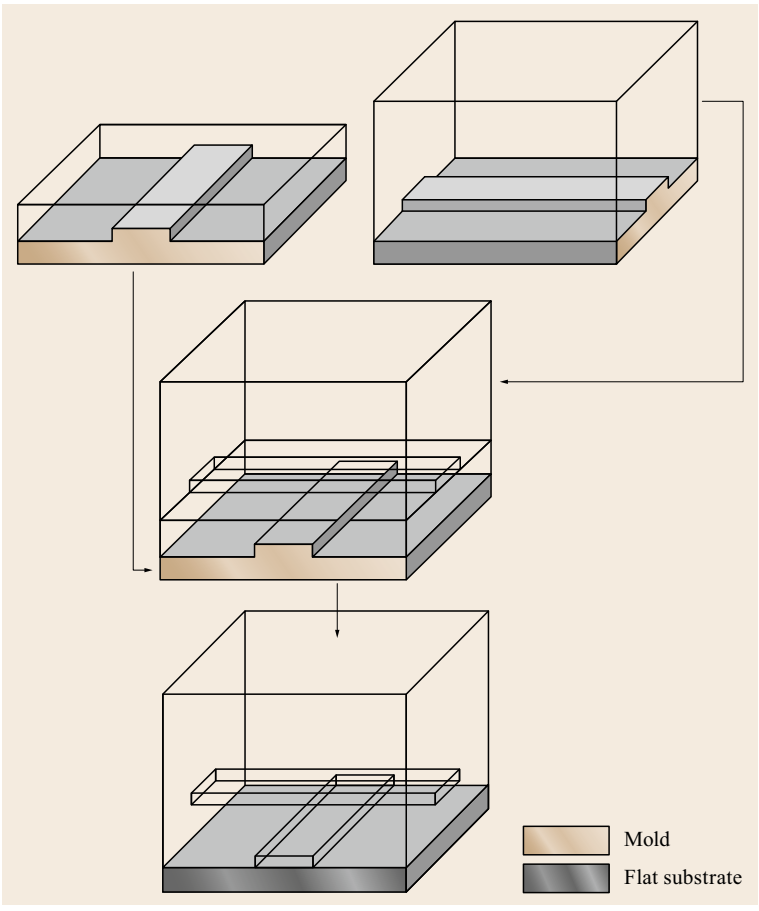


Fig. 16.13 Schematics of the Quake valve setup showing overlapping of the control channels and the fluid channel

et al. [16.306] developed an electrochemical-based microvalve. They utilize the pressure generated due to production of oxygen by electrolysis to deflect a membrane. By supplying 1.6 V and 50 μA, the membrane was deflected by 30–70 μm and generated a pressure of 200 kPa in a few seconds.

Valves were also actuated electrokinetically as demonstrated by *Kirby* et al. [16.308]. Here a potential applied to a fluid generates electrokinetic force that produces a substantial pneumatic pressure in a microchannel. This pressure was used to move a porous polymer plug to seal off a secondary valve chamber and prevent flow between its inlet and outlet. They generated plugs that were 60 μm long by photopatterning of polymer monoliths inside microchannels, which were electrokinetically operated by creating pressure gradients. The valves had a working pressure of 350 bar. They showed a voltage addressable electrokinetic valve useful for liquid chromatography.

Micropumps

Transport of fluid is one of the important aspects of microfluidic devices. Micropumps have been developed using several principles and actuation mechanisms. They can be broadly classified as mechanical [16.309], nonmechanical (electrokinetic and magnetokinetic [16.310], phase change [16.311]) and other novel technologies [16.312]. *Au* et al. [16.291], *Laser* and *Santiago* [16.112] and *Ashraf* et al. [16.292] have reviewed the micropumps in detail. Micropumps can be broadly classified into mechanical and nonmechanical pumps.

Mechanical Micropumps. Mechanical micropumps are characterized by the presence of moving parts in their actuation mechanism. They can be further classified by different actuating mechanisms into rotary displacement pumps [16.309] and vibrating diaphragm pumps [16.310–320].

Rotary Pumps. Rotary pumps typically have a rotor that is positioned such that it sweeps a defined volume of fluid from the inlet to the outlet of the pump. The mechanical action and the direction of motion determine pressure generated. As an example, *Döpper* et al. [16.309] fabricated gear pumps using microfabricated gear wheels. Gear wheels with diameters of 592 and 1192 μm and a height of 500 μm were realized using a combination of lithography, electroplating and molding (LIGA) and precision engineering. They used oil as the pumping medium to achieve a back pressure of more than 1200 kPa and flow rates up to 1 ml/min. They were able to control the amount of oil displaced precisely using this pump and also pump viscous fluids easily because of the good back pressure obtained in their gear pump.

Vibrating Diaphragm Pumps. A vibrating diaphragm pump also known as a membrane pump combines the reciprocating action of the diaphragm or membrane and valves on either side of the flexible diaphragm/membrane to pump fluid. The diaphragm sits atop the pump chamber and vibrates back and forth upon actuation. The motion of the diaphragm displaces the fluid in the chamber in all directions. However, two check valves placed at the inlet and the outlet rectify the flow and cause a net displacement of the fluid over a cycle. These positive displacement pumps can be used to pump any fluid or gas. Vibrating diaphragm pumps can be further classified according to actuation mechanism as electrostatic, electromagnetic, piezoelectric and thermopneumatic pumps. As an example, *Zengerle* et al. [16.310] have developed a flexible diaphragm pump that can be actuated electrostatically with two passive check valves. The unique function of their pump is that it can be run bidirectionally depending on the frequency and pressure. They designed a flap valve that works in the forward direction at low frequency (0.1–800 Hz) and reverses the direction at high frequency (2–6 kHz). They achieved a maximum flow rate of 850 μl/min in the forward direction and 200 μl/min in the reverse direction. The pump had a maximum back pressure of 31 kPa (water) in the forward and 0.7 kPa in the reverse direction.

Dario et al. [16.311] designed and fabricated a microanalyzer for detecting heavy metals in water. Their device consisted of a micropump with a built-in passive valve and an active valve fabricated by a thermoplastic molding process. They actuated the diaphragm using electromagnetic force. Their pump produced a head pressure of 5.5 kPa with a maximum flow rate of 13 ml/s. *Andersson* et al. [16.312] have developed a valveless diffuser micropump for microfluidic analytical systems using peizoelectric actuation. The pump was fabricated on a silicon glass stack (Fig. 16.14) with three subsequent deep ion reactive etching steps. The diaphragm was moved using a piezoactuator placed adjacent to it. They achieved a maximum pump rate of 0.3 μl/min for glycerol and 2.3 μl/min for water.

Piezoelectric actuated pumps are the most researched and have found commercial applications. Several groups including *van Lintel* et al. [16.313], *Stemme* and *Stemme* [16.314], *Forster* et al. [16.315], *Gerlach* and *Wurmus* [16.316], *Koch* [16.317], *Olsson* et al. [16.318] and many more have developed pumps based on piezoelectric actuation. Piezoelectric actuated micropumps are available from companies like ThinXXS2000 [16.319] and Debiotech S.A. [16.320].

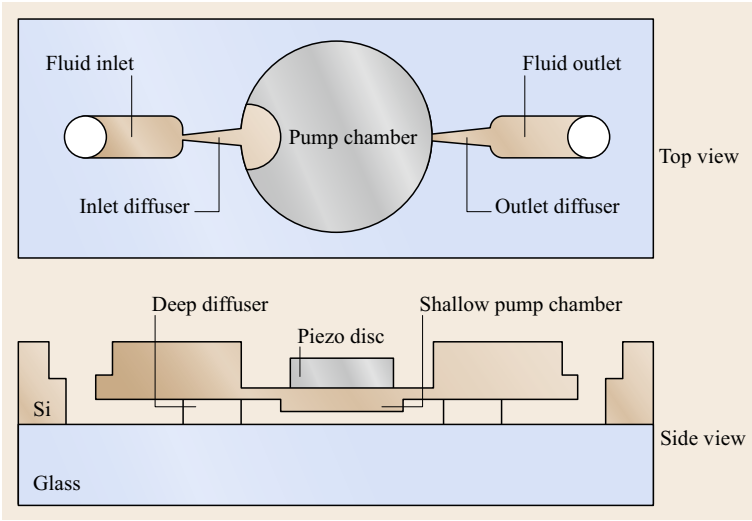


Fig. 16.14 Piezoelectrically actuated micropump fabricated using deep ion reactive etching process (after [16.297])

Phase Change Micropumps. Phase change pumps are a subset of reciprocating pumps where a change in phase (liquid to gas), usually induced by a microheater underneath, generates a pressure that drives the fluid forward. Bubble pumps and electrochemical pumps are types of phase change micropumps.

Bubble Pumps. Bubble pumps expand and collapse a bubble periodically to generate a pressure difference to move the fluid forward. Liquid is heated to expand and its phase change pushes the rest of the fluid in a microchannel. Heating is controlled by applying a voltage. *Tsai* and *Lin* [16.321] reported a bubble pump that was fabricated from two material layers. They pumped isopropyl alcohol at a flow rate of 45 µl/min at a maximum pressure of 400 Pa operating at 20 V.

Electrochemical Pumps. Electrochemical pumps use bubbles generated from electrochemical reactions as the phase change mechanism to generate the pressure to move the fluid forward. *Böhm* et al. [16.322] reported an electrochemical pump to dispense volumes of five nanoliter range. They used electrolysis of water to generate bubbles to drive the fluid forward.

Nonmechanical Pumps. Nonmechanical pumps do not use moving parts as a component of their mechanism but directly convert nonmechanical energy (electrical, magnetic, chemical) into mechanical kinetic energy to pump the fluid in microchannels. They can be categorized as electrohydrodynamic [16.323], magnetohydrodynamic [16.324] and electroosmotic [16.76, 325] micropumps.

Electrohydrodynamic Pumps. Electrohydrodynamic (EHD) pumps move dielectric liquids using electrostatic forces. Application of an electric field between two electrodes placed in a dielectric field leads to some induced charges. The electric field acts on those induced charges producing an electrostatic force that then drags the surrounding fluid along with it. As an example, *Fuhr* et al. [16.323] fabricated an EHD pump where traveling waves were applied to arrays of electrodes. They used dielectric force to transport the fluid by applying an electric field to a fluid having a permittivity gradient. Their pump operated at a maximum voltage of 40 V and generated a maximum flow rate of 2 µl/min.

Magnetohydrodynamic Pumps. Magnetohydrodynamic (MHD) pumps use Lorentz force to induce a body force on the fluid by subjecting the current-carrying ions in an aqueous liquid to a magnetic field. Usually the electrodes are placed in the fluid in such a way that the direction of the current and the magnetic field lines are orthogonal. This produces a body force in the mutually perpendicular direction. As an example, *Jang* and *Lee* [16.324] reported a MHD where they were able to pump fluid through a 40 nm-long microchannel with a hydraulic diameter of 1 mm. The permanent magnets used in their pumps had a magnetic flux density of 0.44 T and generated a maximum flow rate of 63 µl/min with a maximum pressure of 170 Pa.

Electroosmotic Pumps. Electroosmotic pumps use the surface charge created by the electric double layer when a liquid comes into contact with the walls of the microchannel [16.326]. Electric field applied along

the axis of the microchannel tends to impose a body force on the charged layer of fluid adjacent to the walls, which subsequently drags the bulk of the fluid along with it. As an example, *Zeng* et al. [16.76] developed an electroosmotic pump using capillaries packed with silica particles. Their capillaries had large diameters (500–700 μm) and the pump generated a maximum flow rate of 1 μl/min and a pressure of 1 kPa. *Debesset* et al. [16.325] report an electroosmotic pump with flow rates as high as 50 μl/s using an alternating current (AC) electric field to pairs of unequal width electrodes (Fig. 16.15). They developed this pump for chromatography. The electrodes were placed in a circular array and the fluid was pumped in a closed-loop channel.

Other/Novel Micropumps. New technologies like flexural plate waves [16.327], electrowetting [16.328] and acoustic streaming [16.329] have been used to pump fluids.

Flexural Plate Wave Devices. A micropump based on flexural plate waves was developed by *Luginbuhl* et al. [16.327]. In flexural plate wave devices, acoustic waves propagate through a membrane fabricated using thin-film technology to transport fluid. Their device consisted of dual interdigitated transducers fabricated on a thin film membrane made of silicon nitride, platinum and sol-gel-derived piezoelectric ceramic. The device operated at a frequency between 2–3 MHz and actuated a 2×8 mm thin-film membrane. They reported a flow rate of 255 nl/min.

Electrowetting Micropumps. A change in wettability due to applied electric potential moves the fluid. Surface tension is the main reason for transport of fluid in electrowetting micropumps. When a voltage is applied to a dielectric layer, the interfacial energy decreases for the solid–liquid interface causing the fluid to move. *Yun* et al. [16.328] report a surface tension-driven pump by continuous electrowetting. They used the surface tension-induced motion of a mercury drop to transport fluid inside the microchannel. The micropump has a three-layer stack of silicon wafers bonded together. The mercury drop is inserted into a SU-8 channel fabricated on the bottom silicon wafer. When the mercury drop moves it drags an electrolyte solution that deflects the membrane on the second silicon wafer. The third layer consists of passive check valve forming the inlet and outlet of the pump. They achieved a flow rate of 70 μl/min at a voltage of 2.3 V with a maximum pump pressure of 800 Pa.

Acoustic Streaming. Surface waves were used by *Wang* et al. [16.327] to transport fluid inside a microchannel. Interdigital transducers were fabricated to generate surface waves, which propagated towards the microchannel and interacted with the liquid. The compressional component of the wave would be diffracted at the Rayleigh angle into the liquid. The surface waves transfer heat to the substrate and liquid, which increases the saturation pressure of water vapor and will increase the evaporation rate, pushing the liquid forward. Piezoelectric actuators operated at 2–3 MHz to generate a flow rate of 255 nl/min.

Micromixers

Micromixers are important components of lab-on-chip devices for applications in drug delivery, sequencing, amplification and biochemical reactions. *Hessel* et al. [16.69] and *Nguyen* and *Wu* [16.70] have contributed detailed reviews on micromixers. Based on the actuation method, micromixers can be broadly classified as passive and active.

Passive Mixers. Passive mixers do not require external energy and depend solely on diffusion and chaotic advection processes for mixing. The microfabricated structure forces the two combining fluids into shapes

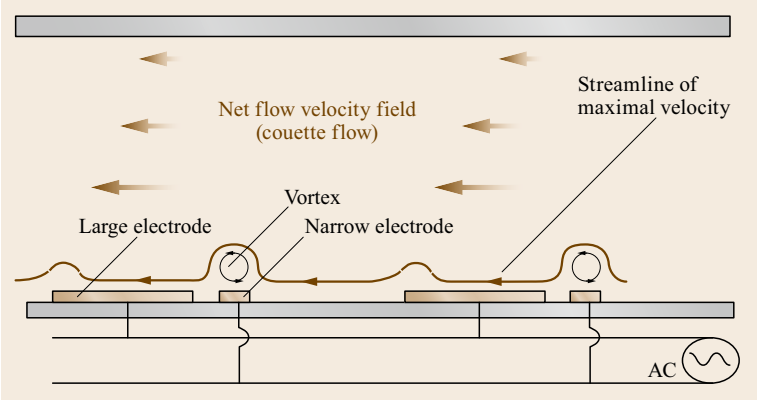


Fig. 16.15 Electroosmotic flow generated using AC voltage applied to pairs of asymmetric electrodes (after [16.325])

and forms that increases the effective surface area between the two fluids and enhance diffusional transport between them. Different types of passive mixers include lamination, injection, chaotic advection and droplet-based mixers.

Lamination. In lamination micromixers, the inlet stream is split into substreams and then the streams are joined again to form a laminate. Diffusional mixing in microfluidic channels can be enhanced by increasing the surface area of contact and reducing the diffusion length. The lamination micromixers generally consist of two inlets connected to a long microchannel. They are also called T-mixers or Y-mixers [16.330, 331] because of the design (Fig. 16.16). The long microchannel provides good contact area for diffusion and the width of the microchannel is small enough for fast diffusion process.

Wong et al. [16.332] reported a T-mixer manufactured on a silicon substrate and bonded it to glass. They applied different pressures to study the effects on mixing. They observed best mixing performance between Reynolds numbers of 400 and 500. They report for a T-mixer 67 μm in hydraulic channel diameter, an applied pressure of 5.5 bar is sufficient to generate complete mixing in less than a millisecond after the two fluids are in contact. *Yi* and *Bau* [16.333] studied the effect of rectangular cross-sections aligned at 90° on steady laminar flows. The bend in their design (Fig. 16.17b) helped in inducing vortices to stir the fluid and this improved mixing. They observed that the vortices disappear quickly at low Reynolds numbers and as a result they had to introduce a number of bends to achieve the desired effect. Using this design they were able to achieve a Reynolds number of 30 where mixing happened right after the bend. *Bessoth* et al. [16.334] used parallel flow lamination design to develop a micromixer. They fabricated a device with structures from a glass/silicon/glass sandwich (Fig. 16.17c), which had a total volume of ≈ 600 nl. They report flow rates between 1–200 ml/min to achieve 95% mixing in 15 ms. *Knight* et al. [16.335] report a continuous flow focusing mixer for fast mixing in microchannels. They showed mix times of less than 10 μs for a sample flow rate of nl/s. They adjusted the flow rates between the sample and sheath flow to achieve this fast mixing. They also reported a Reynolds number of 0.5 for their mixer.

Injection. In an injection mixer, the solute flow is split into many streams and then injected back into the solvent flow at multiple locations. An array of nozzles create microplumes on top of the stream creating a shorter path and increasing the contact area to accelerate diffusion. *Miyake* et al. [16.336] have developed an injection mixer to increase the mixing efficiency in microchannels. They report a homogeneous mixing of 0.5 μl volume in 1.2 s at a flow rate of 0.75 μl/s and a Reynolds number of 0.018.

Chaotic Advection. Chaotic advection is another process by which mixing can occur passively at low Reynolds numbers besides diffusion. Chaotic advection can be generated by special geometries incorporated in the microchannels or by external force. The shape of the channel is altered for splitting, folding, stretching or breaking the flow of fluid. The Reynolds number in such devices can vary anywhere between 10 and 100.

Lin et al. [16.337] developed a mixer that operated at Reynolds numbers ranging between 200 to 2000 and was able to mix 450 pl volume in 50 μs. They introduced vertical pillars (10 μm diameter) in a staggered fashion perpendicular to the direction of flow to achieve turbulent mixing.

Hong et al. [16.338] developed a 2-D micromixer with a modified Tesla valve-type structure. It generated a Coanda effect that caused chaotic advection

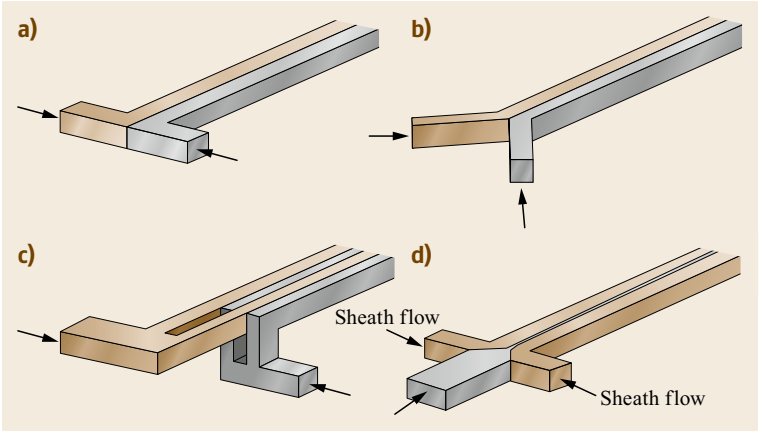


Fig. 16.16a–d Micromixing in microchannels. **(a)** T-mixer **(b)** Y-mixer **(c)** parallel lamination mixer **(d)** continuous flow focusing mixer

and improved mixing. The mixer was fabricated using cyclic olefin copolymer by both thermal bonding and hot embossing techniques. The mixer operated well at Reynolds numbers of 5 and above.

Stroock et al. [16.339] investigated the staggered herringbone structures for mixing. Their mixer operated in low Reynolds numbers ranging between 1 and 100. They fabricated channels 200 μm wide and 70 μm high in PDMS.

Droplet Mixing. Diffusion path can also be reduced by mixing inside droplets. The mixing happens inside the droplet due to its movement, which causes an internal flow field. This technique was used by *Hosokawa* et al. [16.340]. They fabricated a droplet micromixer in PDMS and sealed the channels using a PMMA sheet. A hydrophobic capillary vent was used to fuse two droplets. They showed mixing of two 5 nl droplets pneumatically inside a microchannel.

Active Mixers. In active mixers, an external perturbation is required for actuation. The various perturbations that have been used include pressure disturbance, and electrohydrodynamic, dielectrophretic, electrokinetic, magnetohydrodynamic, acoustic and thermal pulses.

Pressure Disturbance. *Deshmukh* et al. [16.341] used micropumps to achieve pulsatile flow in a continuous flow microfluidic mixer. They fabricated a T-mixer (Fig. 16.17b) within which the micropumps attached to the two inlet arms alternatively drive and stop the flow inside the channel to create a segmented flow. Their micromixer was integrated into a microfluidic device and was fabricated using deep reactive-ion etching (DRIE). Their mixer operated at a low Reynolds number of 0.01.

Electrohydrodynamic Mixing. In electrohydrodynamic mixing, the two fluids that merge into a central flow in a T-channel mixer subjected to anytime varying electric field generate an electrohydrodynamic force (Fig. 16.17d). The superimposition of the time-varying EHD body force over the constant pressure-driven flow causes the advection process and mixing. *Moctar* et al. [16.342] fabricated electrodes such that they were perpendicular to the interface between the two fluids, creating a transversal flow. They used a flow rate of 0.26 μl/s, which corresponded to a Reynolds number of 0.0174. By optimizing the voltage and frequency they observed time for good mixing took 0.1 s over a very short distance (≈ 250 μm).

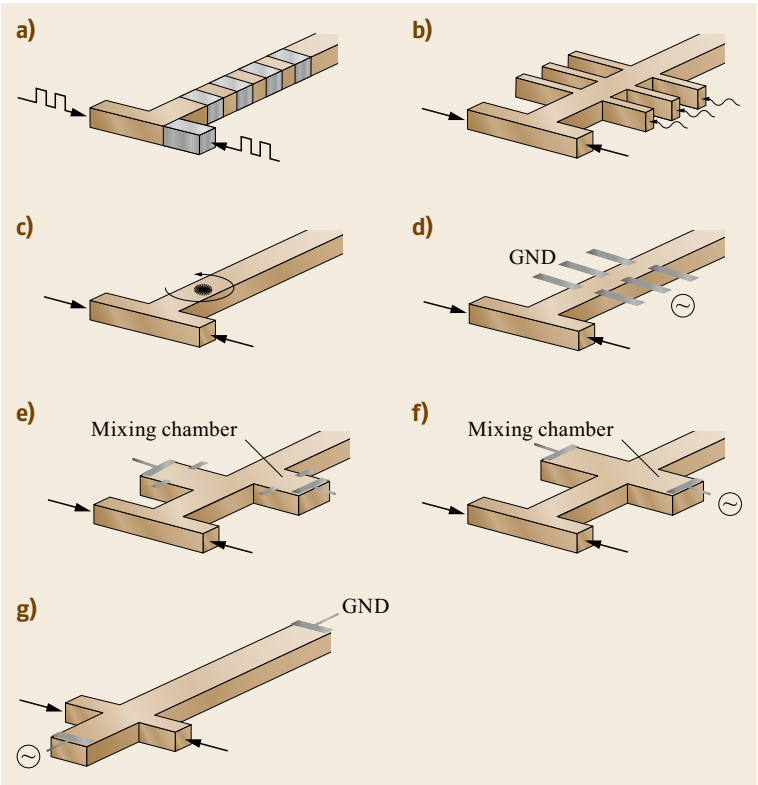


Fig. 16.17a–g Schematics of different types of active mixers

Dielectrophoresis. Dielectrophoresis is the the force acting on a particle when it is subjected to nonuniform electric field. This effect can be used to move a particle towards or away from an electrode (Fig. 16.17e). *Deval* et al. [16.343] reported a dielectrophoretic mixer that used electrical fields as well as geometry to produce chaotic advection. The micromixer operated at 0.02 Reynolds number. The micromixer was integrated into microchannels fabricated using a photolithography technique on silicon and glass.

Electrokinetic Mixing. Similar to pressure disturbance in microchannels for mixing, electrokinetic perturbation can be used to generate segmented flow (Fig. 16.17f). *Tang* et al. [16.344] used periodic application of an electric field to induce electrokinetic flow superimposed on the pressure drive flow to generate alternating segments of the two fluids for mixing. They fabricated the micromixer chamber in PDMS material. Their micromixer operated at 0.04 Reynolds number and they found that the mixing in channels is highly dependent on applied frequency.

Magnetohydrodynamic Mixing. In the presence of a magnetic field orthogonal to an applied direct current (DC, voltage on the electrodes would generate Lorentz force, which can be used for mixing. *Bau* et al. [16.345] developed a micromixer using the magnetohydrodynamic force. The mixer was made from ceramic tapes and a gold paste was used for printing the electrode.

Acoustic Mixer. Acoustic wave energy can also be used to mix fluids inside microfluidic chambers. *Vivek* and *Kim* [16.346] developed an acoustic wave mixer in fabricated silicon. Acoustic waves were produced using a zinc oxide membrane located under the mixing chamber. The vibration was controlled by changing the frequency and voltage of the signal.

Thermal Mixing. Thermal energy was used for mixing fluids. Since diffusion coefficient is dependent on temperature, mixing can be accelerated by increasing the temperature of the chamber. *Tsai* and *Lin* [16.347] developed a thermal mixer based on advection created due to thermal bubble generation. They used a thermal bubble actuated micropump to oscillate the flow creating disturbance inside the mixing chamber. They found that 200 Hz was the optimized frequency for fast mixing in a channel 200 μm wide and 50 μm deep.

Microreactor

Micro–nanofluidic devices can be fabricated in which channels and chambers have extremely small volumes (pl–μl). If these chambers are used as reactors, then reactants can be combined in a precise way and allowed to react for a controlled duration of time. The ability to precisely control the residence time of the reaction allows microfluidic reactors to achieve reaction chemistries that are not possible in large-batch-type reactors. For instance, reactions that produce the desired compound as an intermediate can be performed in the microreactor where the reaction stream is quenched before the intermediate is converted to the final product. The large surface-to-volume ratio of the microreactor also allows highly exothermic reactions to be performed, which may not be possible in a large-batch reactor. Furthermore, microreactors can be parallelized for batch production with higher yields. Reviews regarding chemical synthesis [16.348], novel chemical compounds [16.349], and photocatalytic reactions [16.350] using microfluidic microreactors and production process [16.351] are available.

The microreactor technology has seen much interest from industry. For instance, Lonza has developed a Flowplate lab microfluidic reactor designed for laboratory research for process development. This platform technology was developed to enable easy transfer of technology for commercialization. A production-scale Flowplate A6 microreactor has also been developed that can synthesize products in the kilogram range [16.352]. Other companies such as Little Things Factory GmbH have also developed microreactors on silicon and glass substrates [16.353]. They also supply starter kits for training purposes through SigmaAldrich Inc. and are also capable of developing customized solutions for specific reactions. Researchers at Hitachi Plant Technologies Ltd. [16.351] carried out three reactions (bromination, nitration and ester reduction) in microreactors and compared the yield to conventional batch processes. They reported a 40% increase in yield for the bromination reaction by using microreactor technology. Microreactor technology has been used for studying emulsification processes and nanoparticle generation [16.351]. *Lee* et al. [16.354] have used the microfluidic microreactor technology to synthesize radiolabeled chemical for injection into mice. A micro-computer tomography (micro-CT) image showed the accumulation of the radiolabeled chemical *in-vivo* in mice. Microreactors optimize reaction parameters. In such a study [16.355] the Paal–Knorr reaction product was continuously monitored using infrared analysis for reaction optimization using several multitrajectory strategies.

Lab-on-Chip Systems

Many processes in biochemical analysis, especially those associated with molecular assays, have a number of unit operations that have to be sequentially

performed in order to identify a particular molecular species from the sample. Starting in the 1990s, microfluidic device were designed to miniaturize, integrate and automate many of the unit operations so that these assays could be performed in a wide variety of settings from the bedside and in the field without any particular user expertise. These devices came to be known as lab-on-chip as they performed all the operations that happen in a laboratory effectively on a microfabricated chip.

Microamplification Devices. DNA amplification is one of the most important processes in molecular biology and was one of the earliest process implemented in microfluidic devices. Polymerase chain reaction (PCR) was the first amplification method that was developed in the early 1980s to copy a defined stretch of DNA a million times to enable its detection. It uses a temperature cycle from 55 to 95 °C to perform the biochemical operations of denaturing the DNA, attaching markers known as primers and initiating the polymerase enzyme to copy the strand between the primers. Repetition of this temperature cycle, a process that takes several hours, leads to an exponential increase in the number of copies producing a million copies in 30 cycles. Based on the cycling temperatures, PCR can be broadly classified as temperature-cycling PCR and loop-mediated isothermal amplification (LAMP). LAMP is an isothermal process where the amplification takes place at a constant temperature (typically 65 °C). The cycling-temperature PCR process was miniaturized first in 1995 by *Northrup* et al. [16.19], which led to faster amplification. Since then a large number of amplification devices have been developed that can be categorized based on types of heating and materials [16.356, 357] and method of amplification (chamber [16.358], continuous [16.357] and droplets [16.359]).

Heating Methods for Amplification. Various heating methods (contact and noncontact) have been developed for carrying out PCR reactions in microfluidic chips. Contact methods include thin-film heating elements [16.360], metal or heating blocks [16.361] and other methods [16.362]. *Daniel* et al. [16.360] fabricated a PCR device in silicon by micromachining using an anisotropic wet etching process. They used thin-film platinum metal for heating the chamber. *Khandurina* et al. [16.361] developed an integrated system for rapid PCR analysis on a microfluidic chip. They utilized dual Peltier thermoelectric elements for heating the samples. Other technologies like commercial thin-film resistance heaters [16.362] have been used for PCR. The authors carried out PCR in a plastic microfluidic chip for bacterial detection. Graphite ink electrodes were used as heaters for heating 29 nl volume samples.

Noncontact methods of heating for PCR include hot air cycling [16.363], infrared light irradiation [16.364], laser-mediated heating [16.365] and other methods [16.366]. *Wittwer* et al. [16.363] report a thermal cycler for PCR based on heat transfer with air in a sealed capillary tube. They control the air temperature using a thermocouple and heating coil. *Oda* et al. [16.364] developed a noncontact heating method for PCR in glass chips. They used a tungsten lamp as a heating source and cooled it by natural air convection. *Tanaka* et al. [16.365] used a diode laser to carry out photothermal temperature control of enzyme-catalyzed reaction in a microchip. Other noncontact heating methods include microwave-assisted heating demonstrated by *Fermér* et al. [16.366].

Fluid Handling Methods for Amplification. PCR has been performed in a single chamber, as a continuous phase and in an emulsion (droplets). In single-chamber PCR, the sample is injected into the chamber and the whole chip is heated and cooled through required PCR temperatures. *Belgrader* et al. [16.367] report a PCR device microfabricated in silicon microchambers with integrated optics. They were able to achieve fast cycle times (17 s) because of efficient heating and cooling of the reactions inside the chambers. Instead of heating the fluid in chambers a new method known as continuous flow PCR has been developed where the fluid is cycled between three regions that are maintained at constant but different temperatures [16.368]. Continuous flow-based PCR devices can further be classified into closed-loop devices [16.369], oscillatory device [16.369] and fixed-loop devices [16.72]. In oscillatory PCR devices the sample is moved back and forth between chambers in an oscillating motion to perform PCR. The thermosiphonic effect was used to transport the sample in a fixed circuit in both closed-loop and fixed-loop devices. The only difference being the number of cycles in a fixed loop is predetermined before the fabrication of the device. *Liu* et al. [16.368] have shown closed-loop PCR on microfluidic chips using a 12 nl volume sample. *Bu* et al. [16.369] fabricated a microfluidic chip with an integrated microperistaltic pump and chambers for carrying out PCR. They had three different chambers maintained at three different temperatures for PCR reactions. The sample is pumped back and forth (oscillatory motion) in the chamber to carry out the reactions. *Kopp* et al. [16.72] developed continuous flow micro-PCR chips that use fixed-loop amplification to amplify 176 base pair DNA. The amplification was carried out in less than 20 min, which compared to conventional PCR was much faster, show-

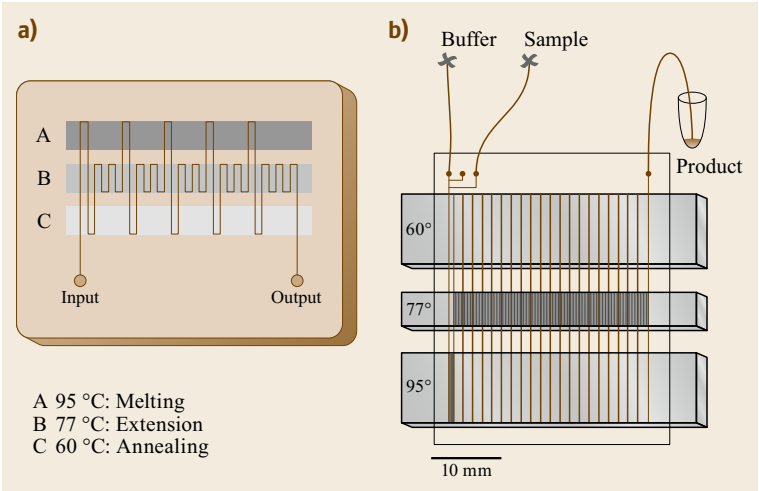


Fig. 16.18a,b Schematics of a continuous flow micro-PCR device. **(a)** Channel design and different heating zones. **(b)** Device layout consists of three inlets and one outlet. The sample and buffer are injected in the two inlets

ing the capabilities of microfluidics as a versatile and useful technology for amplification (Fig. 16.18). PCR in droplets has been reported by various groups [16.370–372] and is primarily used to segment the sample into a number of parallel reactors so that multiple amplifications can be performed. Such segmentation is useful in amplifying DNA that occurs at low concentration in a background of DNA at high concentration. Droplets of PCR mixture are cycled through specific PCR cycle temperatures to perform the PCR reaction within the droplets. *Leman* et al. [16.373] have developed a PCR in droplets technique using droplet microfluidic technology. They report PCR reactions in 20 femtoliter droplets produced at high frequency (3 kHz). The current PCR market size is around $3.28 billion and is estimated to rise to $4.94 billion by 2021 [16.374]. Companies that have commercialized microfluidic PCR platforms include Thermo Fisher Scientific Inc., Bio-Rad Laboratories Inc., F. Hoffmann-La Roche Ltd, Qiagen N.V., Agilent Technologies Inc., Fluidigm Corporation, Takara Bio Inc. and Affymetrix Inc.

Sequencing and Separation. Every living cell carries a genetic code that is required for major functions like protein synthesis, proliferation and apoptosis. Instructions to cells are embedded as codes in the DNA for every function the cell performs starting from a stem cell and differentiating into the specialized cell form. Understanding how this code works is crucial for understanding the genetic, developmental and biological behavior of all living cells. It also helps in deciphering complex processes like cell signaling and mutation. Finally, the cure for cancer and other genetic diseases lies in understanding how the genetic code interacts with the outside environment. A number of microfluidic devices have been developed that aim at low-cost, fast and portable identification of the DNA sequence present in medical, biological or environmental samples. Reviews on microfluidic DNA sequencing technology [16.375–377], emerging technologies [16.378], next generation [16.379–384] and DNA microarrays [16.385].

Sanger Sequencing. One of the earliest methods for sequencing DNA was developed by *Sanger* and is considered as the gold standard for comparison with other methods [16.386]. In Sanger sequencing, a primer, DNA polymerase along with a deoxynucleotide triphosphate (e.g., deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxycytidine triphosphate (dCTP) and deoxythymidine triphosphate (dTTP)) are added to four tubes containing the target sequence. Next, a chain terminating dideoxynucleotide triphosphate (ddNTP, e.g., ddGTP, ddATP, ddTTP and ddCTP) are added one in each tube at a low concentration. Amplification of the DNA leads to random insertions of the chain-terminating nucleotide into the growing chain, producing extended chains of varying lengths corresponding to where that nucleotide was present in the original sequence. After the polymerization reaction, all the extended sequences in each tube are loaded on a gel in different lanes and separated to arrive at the sequence of DNA in the sample. This sequence of steps is ideal for automation using microfluidic technology.

Microfluidic Sanger Sequencing Devices. Many researchers have developed microfluidic devices to carry out DNA sequencing. *Effenhauser* et al. [16.387] developed a micromachined device to sequence DNA using the capillary electrophoresis principle to separate out short DNA sequences 10–25 base pairs in length. They were able to separate the DNA strands in 45 s. *Woolley*

and *Mathies* [16.388] sequenced DNA using microfabricated capillary electrophoresis. The DNA fragments were fluorescently labeled and were separated to 433 bases in about 10 min. *Burns* et al. [16.21] reported an integrated device that uses microfabricated microfluidic channels, heaters and fluorescence detectors to analyze nanoliter-size DNA samples. They were able to show separation for a 50 base pair DNA sample in 30 s. *Blazei* et al. [16.389] developed a microfluidic platform to perform the Sanger sequencing. They incorporated DNA purification and PCR steps into their microfluidic design. Caliper Life Sciences Inc. [16.390] commercialized the technology for DNA sequencing using electrophoresis.

Next-Generation Sequencing. The Sanger sequencing was time consuming and expensive. In order to sequence DNA strands in parallel so that it is quick and cost effective, next-generation sequencing technologies have been developed over the past 15 yr. Many of these next-generation technologies rely on micro–nanofabricated structures and microfluidics for their operation. Some of the well known next-generation sequencing platforms include Illumina sequencing [16.391], Ion Torrent semiconductor sequencing [16.392, 393], single-molecule real-time sequencing [16.394] and nanopore DNA sequencing [16.395].

Illumina Sequencing Technology. Illumina sequencing [16.391] is one of the most popular types of next-generation sequencing because of the cost effectiveness of the process. In this method, the DNA to be sequenced is first sheared by the shotgun method and then amplified. Next, adapters are prepared and added to the target DNA sequence. A flow cell containing oligonucleotides (oligos) with complementary adapters is used for binding the DNA strands. Once the DNA has attached to the oligos they are amplified by the bridge amplification technique. The DNA strands are then denatured and amplified again to form linear arrays of clusters in the flow cell. For sequencing, nucleotides tagged with fluorophores are added, which compete for polymerization and are simultaneously read using a laser in a high-throughput way. The patterned flow cell device used for Illumina sequencing increases the throughput dramatically. Millions of nanowells ($\approx$ 500 nm) are patterned into glass substrate where the DNA primers are deposited [16.392].

Ion Torrent Semiconductor Sequencing. Ion Torrent Inc. developed a sequencing technology based on sequencing chemistry using a semiconductor technology detection system. In this technology, first the DNA is broken down into millions of smaller pieces and each piece is attached to a bead coated with a specific sequence. They are then amplified to cover the entire bead and made to flow over a flow cell with millions of microwells such that each microwell is occupied by only one bead. Then the chip is flooded with one of the four nucleotides. Whenever a nucleotide is incorporated into a single strand of DNA a hydrogen ion is released. The hydrogen ion changes the pH in the solution, which is measured by an ion-sensitive sensor fabricated beneath the well. The change in pH is converted to voltage and a change in voltage is measured to indicate a nucleotide was incorporated in the sequence. This process is repeated every 15 s with a new nucleotide washing over the chip to continue sequencing. Also if two nucleotides are incorporated consecutively, the voltage value doubles indicating a double nucleotide polymerization [16.393]. This process is carried out simultaneously in millions of wells to obtain a high-throughput sequencing technology.

Single-Molecule Real-Time Sequencing. Single-molecule real-time sequencing (SMRT) is based on sequencing by synthesis. It uses a SMRT cell chip consisting of zero-mode waveguides (ZMWs). ZMW is a nanophotonic visualization chamber ($\approx$ 70 nm in diameter) that can analyze volumes as low as 20 zeptoliters (10^{-21} l). It enables detection of individual molecules against the required background of nucleotides. The ZMWs are illuminated from the bottom and the wavelength of light is too large for it to pass through the waveguide because only the lower 20–30 nm of each ZMW is penetrated by excited light. For sequencing, a DNA template polymerase is immobilized at the bottom of the ZMW and phospholinked nucleotides (each labeled with a different color) are added to the ZMWs chamber. As the nucleotide is incorporated into the sequence, the DNA polymerase cleaves the fluorescent label, which is detected by a sensor. The signal-to-noise ratio is greatly improved due to the low volume of the sample [16.394]. This process occurs in thousands of ZMWs, increasing the throughput of the technology.

Nanopore Sequencing. Nanopore sequencing [16.395] sequences DNA using a method called strand sequencing. The strand sequencing is carried out by a protein nanopore ($\approx$ 2–5 nm diameter), which is inserted into a membrane made of synthetic polymer. A current is applied across the membrane such that it flows only through the aperture of the nanopore due to high resistance of the membrane. When a single molecule flows through the nanopore it causes characteristic disruption in the current, which is measured to identify the molecule. In strand sequencing, a DNA

enzyme complex attaches itself to the nanopore and pushes one nucleotide at a time into the nanopore. The enzyme unzips the double-stranded DNA and feeds the single-strand DNA into the nanopores one base at a time and the characteristic disruption in current is created by a particular combination of bases in a specific part of the nanopores. Because these disruptions in currents are specific to the combination of nucleotides the order of the bases can be determined. This process can be multiplexed to create an array of nanopores for high-throughput sequencing.

Emerging Technologies. Recently, *Macosko* et al. were able to use microfluidic droplet technology to sequence thousands of cells individually and in a high-throughput manner [16.396]. They use a technology called Drop-Seq, where thousands of individual cells are encapsulated into nanoliter droplets with a unique barcode to identify the cells and sequence them together. Figure 16.19 shows the schematics of the experiment. In this study, they analyzed 44 808 mouse retinal cells and were able to identify 39 different subpopulations in retina cells.

Hybridization. Hybridization, which is the process of binding two complementary single strands of DNA to make a double strand, is one of the oldest techniques for DNA recognition. With the advent of the technology to synthesize DNA artificially in the early 1950s, single-stranded DNA could be synthesized with various unique sequences and assembled into an array (known as DNA microarray) that could be used to probe the presence of thousands of nucleic acid sequences in a sample all at the same time. The method to assemble unique sequences of DNA into microarrays started in the early 1990s and three main methods were developed, namely: Oxford Gene Technology [16.397], Affymetrix GeneChip [16.398] and DNA microarray technology [16.399]. Oxford Gene Technology's microarrays rely on immobilization of the molecular probes on a solid surface and detect the complementary sequence by hybridization [16.397]. Affymetrix GeneChips work on a similar principle. Sequence information is used to directly design arrays of synthetic oligonucleotides. These thousands of arrays of nucleotides are spatially patterned on a glass chip using a photolithography technique [16.398]. These chips can be used for obtaining information on gene expression levels. When the target sequence hybridizes with the synthesized oligonucleotide, a fluorescent signal is generated, which is read optically. Another high-throughput method to monitor gene expression levels was reported by *Schena* et al. [16.399]. Microarrays of complementary DNA strands were fabricated using a high-speed robotic printing technology on glass substrate. Differential expressions were measured using a two-color fluorescence hybridization technique. Reviews on application [16.400] and recent progress in DNA microarrays can be found in the literature [16.401, 402]. There has been considerable interest in using mi-

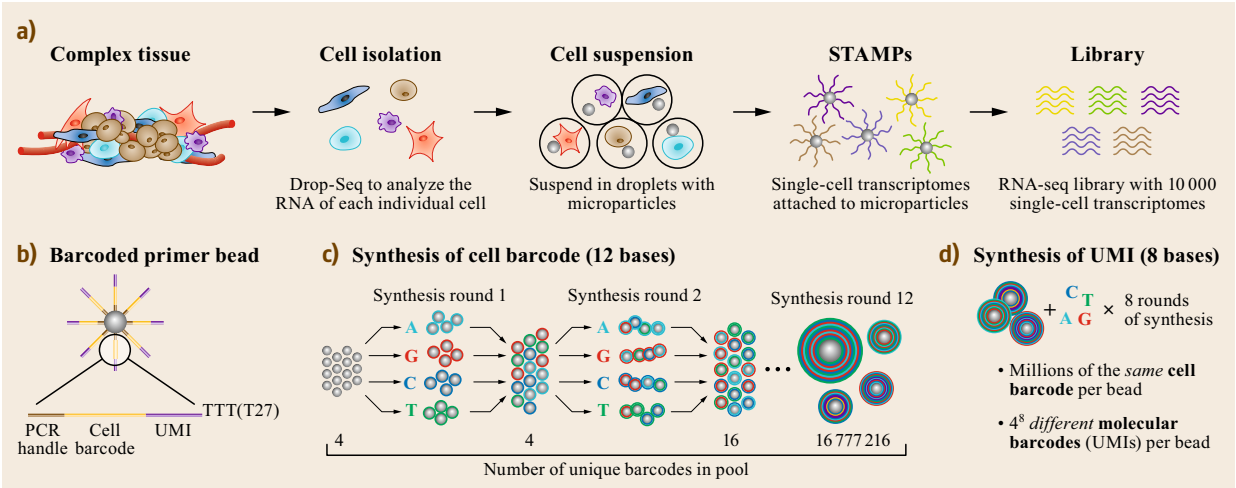


Fig. 16.19a–d Drop-Seq technology. **(a)** Encapsulation of a cell in a nanoliter droplet along with barcoded primer beads suspended in a buffer solution. After lysing the cell and hybridization of messenger RNA (mRNA), the droplets are broken and microparticles collected and washed. The mRNA strands are reverse transcribed to find the complementary DNA (cDNA) sequence. **(b)** The microfluidic device to encapsulate cell and a bead coated with barcoded primer. **(c)** The molecular element of a Drop-Seq library. **(d)** Reconstruction of thousands of single-cell transcriptomes. (Reprinted from [16.396] with permission from Elsevier)

crofluidic technology for DNA microarrays [16.385, 403]. The high surface area-to-volume ratio in microfluidic channels increases the sensitivity of the DNA hybridization making microfluidics a suitable platform for the DNA microarray technology. *Liu* et al. [16.404] developed a complete microfluidic device to carry out DNA hybridization. Their device consisted of 12 000 microarrays with integrated pumps, valves, mixers for fluid handling and storage to carry out the DNA hybridization in an automated fashion. They were able to detect sample RNA concentrations as low as 0.375 pM. The global market for DNA microarrays was $3.93 billion (2016) and is expected to rise to $5.3 billion by 2020 [16.405].

16.5.2 Research Applications

In addition to applications described previously, which have been widely adopted and have found commercial success, there have been a large number of recent application areas that are still in the research and development phase. This section of the chapter describes some of the more promising application areas that are likely to find widespread usage.

Organ-on-Chip

Microfabrication and microfluidic technology has been applied to create continuously perfused channels containing living cells to simulate tissue- and organ-like functionality. By mimicking the physiology, biophysical and biochemical environments including gas exchange, these devices replicate *in-vivo* conditions closely, which is not possible in a conventional cell culture. High-resolution live imaging and drug testing for pharmacological studies are possible using such microfluidic devices. Understanding specific organ disease models, drug and toxicity testing and biomarker identifications are some major application areas. Reviews on perspective of this field [16.57] and applications [16.58] are available for further reading. Based on the application, organ-on-chips can be classified as microphysiological organ models [16.406], target identification and validation [16.407], drug screening [16.408], pharmacokinetics and body-on-chip systems [16.409] and other applications [16.410].

Organ Models. *Hu* et al. [16.406] developed a lung-on-chip, which mimics the functionality of an *in-vivo* lung. They cultured epithelial, alveolar epithelial and immune cells on an elastic membrane. Their device consisted of a main cell culture channel with two side channels to allow gas exchange, which replicates the inhaling and exhaling of the lungs. The central channel has a lower channel, which has a culture medium continuously flowing to simulate the blood flow. Their nanotoxicology studies using this device revealed that cyclic mechanical strain elevates toxic and inflammatory responses of a lung to silica nanoparticles.

Target Identification and Validation. *Song* et al. [16.407] developed a device where they could microengineer a model for chemokine-mediated interactions between breast cancer cells and microvascular endothelium. They fabricated a microfluidic device comprising of two PDMS layers sandwiching a thin, porous polyester membrane (400 nm pore size). The channel (800 μm wide and 60 μm high) in the top layer intersects the regionally distinct and perpendicularly oriented bottom PDMS layer. Human dermal microvascular endothelial cells were cultured in the top channel. The device reproduced serial interactions of tumor cells with endothelia by activating endothelia over specific regions where both the channels intersect. By producing defined flow rates within the range of physiological effects and site-specific interactions of tumor cells with endothelia they were able to mimic the natural setting. Their system revealed that endothelial recruitment of breast cancer cells induced a chemokine involved in cancer metastasis.

Drug Screening. *Aref* et al. [16.408] used an organ-on-chip system to test drug efficacy. They microengineered a 3-D assay of epithelial and mesenchymal transition (EMT) during cancer progression. The microfluidic tissue culture device used in this study consisted of a main central channel containing extracellular gel matrix and two media channels running parallel on either side of it. All the channels were fabricated using PDMS soft lithography. They were able to recreate the EMT-induced tumor dispersion and phenotypic changes in cancer cells by culturing lung cancer cells next to an endothelial cell cultured microchannel. Twelve drugs were tested of which some were approved by the US Food and Drug Administration (FDA). The drugs were injected in the vascular channel and their ability to inhibit the EMT was studied. They report a three orders of magnitude difference in concentrations between their method and conventional 2-D assays.

Pharmacokinetics and Body-on-Chip Systems. *Kim* et al. [16.409] developed a gut-on-chip using a 3-D microfluidic system. Soft lithography was used to fabricate the gut-on-chip device using PDMS as the material. A PDMS membrane (30 μm thick) with 10 μm holes patterned in it was sandwiched between upper and lower microchannels (1000 μm wide and 150 μm high). A silicon wafer with circular pillars deposited photolithographically 10 μm was used as a mold for

making the membrane with pores. The membrane was lined with Caco-2 cells, which separated two cell culture chambers. They mimicked the cyclic deformation of the Caco-2 monolayer by applying cyclic suction to a tubing connected to vacuum chamber and monitored using a feedback loop from the computer. Their mass spectroscopy results generated results consistent with the previous finding.

Other Applications. *Kim* et al. [16.410] used the organ-on-chip platform to test nanomedicine. They used a blood-vessel-on-chip to study the translocation of nanoparticles in an endothelium-coated microfluidic channel. The device consisted of two channels (upper and lower). The upper channel (200 μm wide and 200 μm tall) and lower channel (400 μm wide and 200 μm tall) were fabricated using soft lithography in PDMS. Electrodes were installed in the upper channel before bonding with a microporous polyester membrane. They were able to validate their data by comparing it with animal studies conducted on a rabbit model of atherosclerosis.

High-Throughput Screening (HTS)

High-throughput screening is widely used in drug discovery in order to test the effect of drug candidates on specific proteins, ligands, biomolecules or cells associated with a disease process in order to determine the few that affect the biochemical pathways that determine the outcome of a disease. Currently, HTS assays are performed on microwell plates using robotic instrumentation for fluid handling, which make them expensive. Micro- and nanofluidic devices have been increasingly investigated in order to increase the throughput while lowering the cost and yet providing the same level of automation as conventional plate-based assays. The devices for HTS can be classified as cell-based assays [16.411], binding assays [16.412] and whole-organism assays [16.413].

Cell-Based Assay. Cell-based assays use cells of the organ type involved in the disease to perform high-content screening of the effect of drug candidates on the target protein of interest. Conventionally, they are performed on multiwall plates where cells seeded in the wells can be exposed to different drug candidates and its effects observed on the whole-cell basis. Researchers have used microfluidics in cell-based assays. *Barbulovic-Nad* and *Wheeler* [16.414] have reviewed microfluidic cell-based assays. *Chin* et al. [16.415] report a microfluidic cell-based assay consisting of 10 000 microwells. The micrsowells were microfabricated on a glass cover slip. The dimension of the wells were tunable according to the application ranging from 20–500 μm in diameter and 10–500 μm in height. Rat neural cells were cultured in the microwells and were enclosed in a microfluidic system such that all the cells received the same media including any chemicals secreted by the cells. Proliferation of over 3000 single cells was analyzed over a period of a few days. They also propose that this platform can be extended to monitor cells in a high-throughput way.

Binding Assay. *Delamarche* et al. [16.416] developed a binding assays for HTS application. They patterned the surface of a microchannel with antigens by using the capillary forces generated because of the small dimensions of the channel. Bovine serum albumin was used as a blocking agent and they were able to pattern antigens in narrow stripes on the surface. By using each line to detect a specific analyte from the sample solution this process can be highly parallelized. They propose that such an assay can reduce the analysis time ten-fold and reduce the sample and reagent volume by 100-fold compared to conventional well-plate techniques.

Whole-Organism Assays. Microfluidics has been used to perform high-throughput behavioral and imaging assays. Recently, *Ghaemi* and *Selvaganapathy* [16.413] developed a microfluidic high-throughput assay to study the effects of drugs on *Drosophila*. They also developed a high-throughput method to load the *Drosophila* in microfluidic devices in an automated way [16.417]. *Rezai* et al. [16.418] studied the biology of human diseases by testing different drugs on *C. elegans* using a whole-organism assay. The electrotaxis property was used to sort *C. elegans* in a whole organism assay in a fast and automated method [16.419].

Bioprinting

Additive manufacturing methods such as 3-D printing have become popular in the construction of 3-D constructs in thermoplastic polymers and metals. Recently, the same technology is being increasingly applied to build tissue-like constructs for biomedical applications such as drug discovery, toxicology and potentially organ replacement. One particular challenge in artificial tissue construction is the need to integrate microscale vascular networks inside the 3-D structures. Still in its early stage, bioprinting has been applied in regenerative medicine to generate transplantable tissues including liver, bone, skin and cartilage. The current bioprinting technology still needs to solve problems like high-resolution cell depositing, controlled cell patterning and integration within complex 3-D tissues. *Mandrycky* et al. [16.420] have a review of this topic. A combination of advances in stem cell biology and bioprinting technology could enable a platform for high-throughput

printing of organs and drug screening methods in the future.

Micro Fuel Cells and Energy Devices

One of the main requirement for microsystems, portable devices and POC applications is a reliable means of power generation. Microfluidic fuel cells can offer portable and fully integrated energy sources. *Safdar* et al. [16.421] have recently reviewed microfluidic technology for energy generation. Some examples of microfluidic fuel cells include microfluidic biofuel cells [16.422], microfluidic photocatalytic fuel cells [16.423] and microfluidic electrochemical fuel cells [16.424]. The microfluidic electrochemical fuel cells use carbon electrodes, metals or metal alloys as catalysts to carry out a redox reaction for generating energy. *Esquivel* et al. [16.424] have developed a paper-based microfluidic electrochemical fuel cell. The fuel cell is capable of producing current densities between $1-5\,\mathrm{mW/cm^2}$. A conjugated pad was used to store the electrolyte solution and a fuel on the device with integrated electrodes. The fuel cell functions by soaking the pad in water. The diluted fuel was oxidized underwater at the platinum-ruthenium electrode to generate current. In the future, a need for microfluidic fuel cells to generate higher power densities with minimal operation and fabrication cost will drive this field forward.

16.6 Outlook and Future Directions

Microfluidics and nanofluidics are enabling technologies that utilize the unique features of the fluid flow at these scale to perform operations with some advantage over its macroscale equivalents. Over the past 20 years the microfabrication technology required to make micro- and nanofluidic devices has developed significantly. Currently, devices in a wide range of materials from metals, inorganic materials such as silicon and glass, polymeric materials such as PMMA, PDMS, PC, PS, PI, COC and COP using a variety of manufacturing processes to achieve resolutions and feature sizes in the micro- and nanoscale have been well developed. Similarly, devices have been designed for a wide variety of applications such as precise fluid dispensing, medical and environmental diagnosis, analytical process such as chromatography, sorting and electrophoresis, microscale chemical production and synthesis, combinatorial screening and others. Some of these applications areas have found significant commercial markets and have generated hundreds of millions of dollars in revenue per year. The future evolution of this field is likely to focus on two aspects, namely new materials and fabrication processes and new applications. Some of the potential emerging innovations in this field are listed below.

16.6.1 Additive Manufacturing

The revolution in additive manufacturing of materials will have a significant impact in the way that micro- and nanofluidic devices are made. Already 3-D printers using fused deposition modeling (FDM) have been developed and commercialized for common thermoplastic polymers that are used in microfluidic devices. In the near future, additive manufacturing methods will be developed for other polymer systems such as thermosets or elastomers. Methods for the combination of elastomers with thermoplastic polymers will also be developed to integrate multiple materials in the same device. Simultaneously, 3-D printing methods for gels and other soft materials will also be extensively developed, catering to the tissue engineering applications. Finally, additive manufacturing methods other than FDM, such as layer object manufacturing, will be investigated for micromanufacturing.

16.6.2 Artificial Organs

The development of microfabrication methods for shaping gels, extracellular matrices and cells is in its early stages currently. With the development of increasingly sophisticated techniques to assemble these materials, they will be used to fabricate tissue-like constructs and subsequently artificial organs that combine multiple cell types, ECMs, and inorganic materials in the 3-D complexity present in human organs. These artificially constructed organs can be used in drug discovery applications, as *in-vivo* bioreactors or as true fully functional organs with multifunctional capacity.

16.6.3 Biopharmaceutical Production

The ability of the microfluidic devices to precisely control the residence time in a microreactor and its modular nature are of importance in cost-effective production of specialty chemicals. Increasingly these same characteristics will be of importance in production of biopharmaceuticals such as proteins and cells. For instance, PEGylation of proteins is an important process in biopharmaceutical production. Microfluidic reactors

can optimize the residence time such that the desired mono-PEGylated product is produced over the undesired di- and tri-PEGlayted ones. Similarly, the ability to fabricate parallel culture chambers will be important in the production of stem cells and immunotherapy cells for treatment of diseases such as cancer. Microscale culture chambers integrated with microfluidic channels to perfuse and maintain nutrients, pH and other culture parameters as well as sensors to measure them will be important in growing cell-based biopharmaceuticals under controlled and documented conditions required for medical products.

16.6.4 Droplets in Food and Cosmetics

Methods for monodisperse droplet formation and the ability to form single, double and triple emulsions have also been demonstrated. Recently, some efforts have been made to identify scalable methods of droplet production. In the near future, more research will be conducted on the development of scalable high-throughput and high-volume production of droplets. Simultaneously, the application areas will spread from medicine and biology, where the sample sizes are small, to more mainstream applications like in food and cosmetics where the batch sizes are extremely large. For instance, the development of scalable high-throughput and cheap techniques to produce microdroplets or gels loaded with specific flavor, odorant and other sensory chemicals can be used for programmed release that will make the eating experience more pleasurable.

16.6.5 Continuous Health Monitoring

Many of the microfluidic and nanofluidic devices that are made currently are designed for single use and are disposable in nature. These are suitable for applications such as disease diagnosis and biochemical assays in the field or hospital settings. Lately, there has been considerable interest in wearable sensors and in the use of physical sensors for continuous monitoring of parameters related to health. In the near future, methods to extend such monitoring using chemical and biochemical sensors will be actively investigated. Such sensors will require microfluidic technology that can clean, compensate and calibrate even with multiple use. These will require incorporation of new materials that are self cleaning, with active cleaning technologies, microfluidic components that are not prone to contamination and retention, as well as components that are reliable and have the same performance characteristics on extended usage. The development of these technologies will enable on-body or in-home continuous biochemical health monitoring of individuals that could serve as a sentinel to identify and detect onset of disease conditions early.

In summary, the outlook for micro- and nanofluidics is bright and the field is extending into a number of significant and interesting avenues that will make it the enabling technology of the future. Developments in the fundamental physics of fluid flow and its control, microfabrication methods, microfluidic components, and applications in new and emerging areas are all anticipated.

References

16.1 D.J. Beebe, G.A. Mensing, G.M. Walker: Physics and applications of microfluidics in biology, Annu. Rev. Biomed. Eng. **4**, 261–286 (2002)

16.2 A. Manz, N. Graber, H.M. Widmer: Miniaturized total chemical analysis systems: A novel concept for chemical sensing, Sens. Actuat. B: Chem. **1**, 244–248 (1990)

16.3 E.H. Yoo, S.Y. Lee: Glucose Biosensors: An overview of use in clinical practice, Sensors (Basel) **10**, 4558–4576 (2010)

16.4 Y. Feng, Y. Zhang, C. Ying, D. Wang, C. Du: Nanopore-based fourth-generation DNA sequencing technology, Genom. Proteom. Bioinform. **13**, 4–16 (2015)

16.5 MarketsandMarkets: Microfluidics market by application, material, diagnostic end user-2020, http://www.marketsandmarkets.com/Market-Reports/microfluidics-market-1305.html (2016)

16.6 Cepheid Inc.: Cepheid geneXpert IV, http://www.cepheid.com/us/cepheid-solutions/systems/genexpert-systems/genexpert-iv (2016)

16.7 D.D. Dalma-Weiszhausz, J. Warrington, E.Y. Tanimoto, C.G. Miyada: The affymetrix genechip platform: An overview, Methods Enzymol **410**, 3–28 (2006)

16.8 F.F. Reuss: Sur un nouvel effet de l'electricite galvanique, Mem. Soc. Imp. Nat. Mosc. **2**, 327–337 (1809)

16.9 S.C. Terry, J.H. Jerman, J.B. Angell: A gas chromatographic air analyzer fabricated on a silicon wafer, IEEE Trans. Electron. Devices **26**, 1880–1886 (1979)

16.10 J.K. Fink: Inkjet inks. In: *The Chemistry of Printing Inks and Their Electronics and Medical Applications*, ed. by M. Scrivener (Wiley, New York 2014) pp. 1–10

16.11 Yole Développement: Inkjet printhead market technology trends (2016)

16.12 S.P. Fodor, R.P. Rava, X.C. Huang, A.C. Pease, C.P. Holmes, C.L. Adams: Multiplexed biochemical assays with biological chips, Nature **364**, 555–556 (1993)

16.13 Kalorama Information (Market Research Group), LLC: Market research report, http://www.kaloramainformation.com/Microarrays-9618385 (2016)
16.14 D.R. Reyes, D. Iossifidis, P.-A. Auroux, A. Manz: Micro total analysis systems. 1. Introduction, theory and technology, Anal. Chem. **74**, 2623–2636 (2002)
16.15 C.T. Culbertson, T.G. Mickleburgh, S.A. Stewart-James, K.A. Sellens, M. Pressnall: Micro total analysis systems: Fundamental advances and biological applications, Anal. Chem. **86**, 95–118 (2014)
16.16 M.L. Kovarik, D.M. Ornoff, A.T. Melvin, N.C. Dobes, Y. Wang, A.J. Dickinson, P.C. Gach, P.K. Shah, N.L. Allbritton: Micro total analysis systems: Fundamental advances and applications in the laboratory, clinic, and field, Anal. Chem. **85**, 451–472 (2013)
16.17 D.E.W. Patabadige, S. Jia, J. Sibbitts, J. Sadeghi, K. Sellens, C.T. Culbertson: Micro total analysis systems: Fundamental advances and applications, Anal. Chem. **88**, 320–338 (2016)
16.18 A.J. Tudos, G.J. Besselink, R.B. Schasfoort: Trends in miniaturized total analysis systems for point-of-care testing in clinical chemistry, Lab Chip **1**, 83–95 (2001)
16.19 M. A. Northrup, M.T. Ching, R.M. Whlte, R.T. Watson: Proc. Transducers '93, Yokohama, Japan, (1993), p. 924
16.20 A.T. Woolley, D. Hadley, P. Landre, A.J. deMello, R.A. Mathies, M.A. Northrup: Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device, Anal. Chem. **68**, 4081–4086 (1996)
16.21 M.A. Burns, B.N. Johnson, S.N. Brahmasandra, K. Handique, J.R. Webster, M. Krishnan, T.S. Sammarco, P.M. Man, D. Jones, D. Heldsinger, C.H. Mastrangelo, D.T. Burke: An integrated nanoliter DNA analysis device, Science **282**, 484–487 (1998)
16.22 H. Becker, L.E. Locascio: Polymer microfluidic devices, Talanta **56**, 267–287 (2002)
16.23 Y. Xia, G.M. Whitesides: Soft lithography, Angew. Chem. Int. Edn. **37**, 550–575 (1998)
16.24 M.A. Unger, H.-P. Chou, T. Thorsen, A. Scherer, S.R. Quake: Monolithic microfabricated valves and pumps by multilayer soft lithography, Science **288**, 113–116 (2000)
16.25 S. Quake: A nanoliter rotary device for polymerase chain reaction, Electrophoresis **23**, 1531–1536 (2002)
16.26 Y.-C. Wang, M.H. Choi, J. Han: Two-dimensional protein separation with advanced sample and buffer isolation using microfluidic valves, Anal. Chem. **76**, 4426–4431 (2004)
16.27 A.Y. Fu, H.-P. Chou, C. Spence, F.H. Arnold, S.R. Quake: An integrated microfabricated cell sorter, Anal. Chem. **74**, 2451–2457 (2002)
16.28 V. Studer, R. Jameson, E. Pellereau, A. Pepin, Y. Chen: A microfluidic mammalian cell sorter based on fluorescence detection, Microelectron. Eng. **73**, 852–857 (2004)
16.29 D. Lipomi, R. Martinez, L. Cademartiri, G. Whitesides: Soft lithographic approaches to nanofabrication. In: *Polym. Sci. Compr. Ref.*, Vol. 10 (2012) pp. 211–231
16.30 P. Kim, K.W. Kwon, M.C. Park, S.H. Lee, S.M. Kim, K.Y. Suh: Soft lithography for microfluidics: A review, Biochip J **2**, 1–11 (2008)
16.31 D. Qin, Y. Xia, G.M. Whitesides: Soft lithography for micro- and nanoscale patterning, Nat. Protoc. **5**, 491–502 (2010)
16.32 G.M. Whitesides, E. Ostuni, S. Takayama, X. Jiang, D.E. Ingber: Soft lithography in biology and biochemistry, Annu. Rev. Biomed. Eng. **3**, 335–373 (2001)
16.33 H.A. Stone, A.D. Stroock, A. Ajdari: Engineering flows in small devices: Microfluidics toward a lab-on-a-chip, Annu. Rev. Fluid Mech. **36**, 381–411 (2004)
16.34 G.F. Christopher, S.L. Anna: Microfluidic methods for generating continuous droplet streams, J. Phys. D: Appl. Phys. **40**, R319 (2007)
16.35 P. Garstecki, M.J. Fuerstman, H.A. Stone, G.M. Whitesides: Formation of droplets and bubbles in a microfluidic T-junction – scaling and mechanism of break-up, Lab Chip **6**, 437–446 (2006)
16.36 P. Tabeling: *Introduction to Microfluidics* (Oxford Univ. Press, Oxford 2005)
16.37 J. Atencia, D.J. Beebe: Controlled microfluidic interfaces, Nature **437**, 648–655 (2005)
16.38 C.N. Baroud, F. Gallaire, R. Dangla: Dynamics of microfluidic droplets, Lab Chip **10**, 2032–2045 (2010)
16.39 H.N. Joensson, H. Andersson Svahn: Droplet Microfluidics – A tool for single-cell analysis, Angew. Chem. Int. Edn. **51**, 12176–12192 (2012)
16.40 S.Y. Teh, R. Lin, L.H. Hung, A.P. Lee: Droplet microfluidics, Lab Chip **8**, 198–220 (2008)
16.41 Z. Guan, Y. Zou, M. Zhang, J. Lv, H. Shen, P. Yang, H. Zhang, Z. Zhu, C.J. Yang: A highly parallel microfluidic droplet method enabling single-molecule counting for digital enzyme detection, Biomicrofluidics **8**, 014110 (2014)
16.42 C. Benz, H. Retzbach, S. Nagl, D. Belder: Protein-protein interaction analysis in single microfluidic droplets using FRET and fluorescence lifetime detection, Lab Chip **13**, 2808–2814 (2013)
16.43 J.W. Park, S.C. Na, T.Q. Nguyen, S.M. Paik, M. Kang, D. Hong, I.S. Choi, J.H. Lee, N.L. Jeon: Live cell imaging compatible immobilization of Chlamydomonas reinhardtii in microfluidic platform for biodiesel research, Biotechnol. Bioeng. **112**, 494–501 (2015)
16.44 W. Li, H. Dong, G. Tang, T. Ma, X. Cao: Controllable microfluidic fabrication of Janus and microcapsule particles for drug delivery applications, RSC Adv **5**, 23181–23188 (2015)
16.45 C.H. Yang, C.Y. Wang, A.M. Grumezescu, A.H. Wang, C.J. Hsiao, Z.Y. Chen, K.S. Huang: Core-shell structure microcapsules with dual pH-responsive drug release function, Electrophoresis

35, 2673–2680 (2014)
16.46 N. Bardiya, J.-W. Choi, S.-I. Chang: Analysis of single nucleotide polymorphism in human angiogenin using droplet-based microfluidics, BioChip J **8**, 15–21 (2014)
16.47 J. Shuga, Y. Zeng, R. Novak, Q. Lan, X. Tang, N. Rothman, R. Vermeulen, L. Li, A. Hubbard, L. Zhang, R.A. Mathies, M.T. Smith: Single molecule quantitation and sequencing of rare translocations using microfluidic nested digital PCR, Nucl. Acids Res. **41**, e159 (2013)
16.48 N. Chronis, M. Zimmer, C.I. Bargmann: Microfluidics for in vivo imaging of neuronal and behavioral activity in caenorhabditis elegans, Nat. Methods **4**, 727–731 (2007)
16.49 K. Chung, M.M. Crane, H. Lu: Automated on-chip rapid microscopy, phenotyping and sorting of C. Elegans, Nat. Methods **5**, 637–643 (2008)
16.50 P. Rezai, A. Siddiqui, P.R. Selvaganapathy, B.P. Gupta: Electrotaxis of Caenorhabditis elegans in a microfluidic environment, Lab Chip **10**, 220–226 (2010)
16.51 K. Chung, Y. Kim, J.S. Kanodia, E. Gong, S.Y. Shvartsman, H. Lu: A microfluidic array for large-scale ordering and orientation of embryos, Nat. Methods **8**, 171–176 (2011)
16.52 R. Ghaemi, P. Rezai, B.G. Iyengar, P.R. Selvaganapathy: Microfluidic devices for imaging neurological response of drosophila melanogaster larva to auditory stimulus, Lab Chip **15**, 1116–1122 (2015)
16.53 E.M. Lucchetta, J.H. Lee, L.A. Fu, N.H. Patel, R.F. Ismagilov: Dynamics of drosophila embryonic patterning network perturbed in space and time using microfluidics, Nature **434**, 1134–1138 (2005)
16.54 Y.C. Shen, D. Li, A. Al-Shoaibi, T. Bersano-Begey, H. Chen, S. Ali, B. Flak, C. Perrin, M. Winslow, H. Shah, P. Ramamurthy, R.H. Schmedlen, S. Takayama, K.F. Barald: A student team in a University of Michigan biomedical engineering design course constructs a microfluidic bioreactor for studies of zebrafish development, Zebrafish **6**, 201–213 (2009)
16.55 L.L. Bischel, B.R. Mader, J.M. Green, A. Hutenlocher, D.J. Beebe: Zebrafish entrapment by restriction array (ZEBRA) device: A low-cost, agarose-free zebrafish mounting technique for automated imaging, Lab Chip **13**, 1732–1736 (2013)
16.56 F. Yang, C. Gao, P. Wang, G.-J. Zhang, Z. Chen: Fish-on-a-chip: Microfluidics for zebrafish research, Lab Chip **16**, 1106–1125 (2016)
16.57 S.N. Bhatia, D.E. Ingber: Microfluidic organs-on-chips, Nature **201**, 4 (2014)
16.58 E.W. Esch, A. Bahinski, D. Huh: Organs-on-chips at the frontiers of drug discovery, Nat. Rev. Drug Discov. **14**, 248–260 (2015)
16.59 I.M. Derrington, T.Z. Butler, M.D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, J.H. Gundlach: Nanopore DNA sequencing with MspA, Proc. Nat. Acad. Sci. **107**, 16060–16065 (2010)
16.60 A.M. Streets, Y. Huang: Microfluidics for biological measurements with single-molecule resolution, Curr. Opin. Biotechnol. **25**, 69–77 (2014)
16.61 C. Liu, Y. Qu, Y. Luo, N. Fang: Recent advances in single-molecule detection on micro- and nano-fluidic devices, Electrophoresis **32**, 3308–3318 (2011)
16.62 J. Fu, R.B. Schoch, A.L. Stevens, S.R. Tannenbaum, J. Han: A patterned anisotropic nanofluidic sieving structure for continuous-flow separation of DNA and proteins, Nat. Nanotechnol. **2**, 121–128 (2007)
16.63 S. Prakash, J. Yeom: Energy and environmental applications. In: *Nanofluidics and Microfluidics*, ed. by W. Andrew (Elsevier, Oxford 2014) pp. 241–269
16.64 F.M. White, I. Corfield: *Viscous Fluid Flow* (McGraw-Hill, New York 2006)
16.65 R.A. Freitas: *Nanomedicine, Basic Capabilities*, Vol. I (Landes Bioscience Georgetown, Oxford 1999)
16.66 D.L. Englert, M.D. Manson, A. Jayaraman: Flow-based microfluidic device for quantifying bacterial chemotaxis in stable, competing gradients, Appl. Environ. Microbiol. **75**, 4557–4564 (2009)
16.67 J. Oakey, J. Allely, D.W. Marr: Laminar-flow-based separations at the microscale, Biotechnol. Prog. **18**, 1439–1442 (2002)
16.68 A. Hatch, A.E. Kamholz, K.R. Hawkins, M.S. Munson, E.A. Schilling, B.H. Weigl, P. Yager: A rapid diffusion immunoassay in a T-sensor, Nat. Biotechnol. **19**, 461–465 (2001)
16.69 V. Hessel, H. Löwe, F. Schönfeld: Micromixers – A review on passive and active mixing principles, Chem. Eng. Sci. **60**, 2479–2501 (2005)
16.70 N.-T. Nguyen, Z. Wu: Micromixers – A review, J. Micromech. Microeng. **15**(2), R1 (2005)
16.71 X. Xuan, D. Li: Analytical study of joule heating effects on electrokinetic transportation in capillary electrophoresis, J. Chromatogr. A **1064**, 227–237 (2005)
16.72 M.U. Kopp, A.J. De Mello, A. Manz: Chemical amplification: Continuous-flow PCR on a chip, Science **280**, 1046–1048 (1998)
16.73 W.L. Tong, M.K. Tan, J.K. Chin, K.S. Ong, Y.M. Hung: Coupled effects of hydrophobic layer and vibration on thermal efficiency of two-phase closed thermosyphons, RSC Adv **5**, 10332–10340 (2015)
16.74 A.A. Darhuber, S.M. Troian: Principles of microfluidic actuation by modulation of surface stresses, Annu. Rev. Fluid Mech. **37**, 425–455 (2005)
16.75 Z. Stojek: The electrical double layer and its structure. In: *Electroanalytical Methods*, ed. by F. Scholz (Springer, Berlin, Heidelberg 2010) pp. 3–9
16.76 S. Zeng, C.-H. Chen, J.C. Mikkelsen, J.G. Santiago: Fabrication and characterization of electroosmotic micropumps, Sens. Actuat. B: Chem. **79**, 107–114 (2001)
16.77 P.H. Paul, D.W. Arnold, D.J. Rakestraw: Electrokinetic generation of high pressures using porous microstructures. In: *Micro Total Analysis Systems*, ed. by D.J. Harrison, A. van den Berg (Springer, Dordrecht 1998) pp. 49–52

16.78 P.H. Paul, D.J. Rakestraw, D.W. Arnold, K.R. Hencken, J.S. Schoeniger, D.W. Neyes: Electrokinetic high pressure hydraulic system, US Patent (Application) 6277257 B1 (2001)
16.79 S. Fiedler, S.G. Shirley, T. Schnelle, G. Fuhr: Dielectrophoretic sorting of particles and cells in a microsystem, Anal. Chem. **70**, 1909–1915 (1998)
16.80 J.G. Kralj, M.T. Lis, M.A. Schmidt, K.F. Jensen: Continuous dielectrophoretic size-based particle sorting, Anal. Chem. **78**, 5019–5025 (2006)
16.81 K. Ahn, C. Kerbage, T.P. Hunt, R. Westervelt, D.R. Link, D. Weitz: Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices, Appl. Phys. Lett. **88**, 24104–24104 (2006)
16.82 S. Zeng, B. Li, X. Su, J. Qin, B. Lin: Microvalve-actuated precise control of individual droplets in microfluidic devices, Lab Chip **9**, 1340–1343 (2009)
16.83 Z.Z. Chong, S.H. Tan, A.M. Ganan-Calvo, S.B. Tor, N.H. Loh, N.T. Nguyen: Active droplet generation in microfluidics, Lab on a Chip **16**, 35–58 (2016)
16.84 H. Gu, M.H. Duits, F. Mugele: Droplets formation and merging in two-phase flow microfluidics, Int. J. Mol. Sci. **12**, 2572–2597 (2011)
16.85 T. Thorsen, R.W. Roberts, F.H. Arnold, S.R. Quake: Dynamic pattern formation in a vesicle-generating microfluidic device, Phys. Rev. Lett. **86**, 4163–4166 (2001)
16.86 S.L. Anna, N. Bontoux, H.A. Stone: Formation of dispersions using flow focusing in microchannels, Appl. Phys. lett. **82**, 364–366 (2003)
16.87 L. Yobas, S. Martens, W.-L. Ong, N. Ranganathan: High-performance flow-focusing geometry for spontaneous generation of monodispersed droplets, Lab on a Chip **6**, 1073–1079 (2006)
16.88 C. Cramer, P. Fischer, E.J. Windhab: Drop formation in a co-flowing ambient fluid, Chem. Eng. Sci. **59**, 3045–3058 (2004)
16.89 D.J. Collins, A. Neild, A. de Mello, A.-Q. Liu, Y. Ai: The Poisson distribution and beyond: Methods for microfluidic droplet production and single all encapsulation, Lab Chip **15**, 3439–3459 (2015)
16.90 H. Song, M.R. Bringer, J.D. Tice, C.J. Gerdts, R.F. Ismagilov: Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels, Appl. Phys. Lett. **83**, 4664–4666 (2003)
16.91 D.R. Link, E. Grasland-Mongrain, A. Duri, F. Sarrazin, Z. Cheng, G. Cristobal, M. Marquez, D.A. Weitz: Electric control of droplets in microfluidic devices, Angew. Chem. Int. Edn. **45**, 2556–2560 (2006)
16.92 D. Link, S.L. Anna, D. Weitz, H. Stone: Geometrically mediated breakup of drops in microfluidic devices, Phys. Rev. Lett. **92**, 054503 (2004)
16.93 A.T.-H. Hsieh, N. Hori, R. Massoudi, P.J.-H. Pan, H. Sasaki, Y.A. Lin, A.P. Lee: Nonviral gene vector formation in monodispersed picolitre incubator for consistent gene delivery, Lab on a Chip **9**, 2638–2643 (2009)
16.94 Y.-C. Tan, Y.L. Ho, A.P. Lee: Microfluidic sorting of droplets by size, Microfluid. Nanofluidics **4**, 343–348 (2008)
16.95 M. Srisa-Art, I.C. Bonzani, A. Williams, M.M. Stevens: A.J. deMello, J.B. Edel: Identification of rare progenitor cells from human periosteal tissue using droplet microfluidics, Analyst **134**, 2239–2245 (2009)
16.96 J.C. Baret, Y. Beck, I. Billas-Massobrio, D. Moras, A.D. Griffiths: Quantitative cell-based reporter gene assays using droplet-based microfluidics, Chem. Biol. **17**, 528–536 (2010)
16.97 E. Brouzes, M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J.B. Hutchison, J.M. Rothberg, D.R. Link, N. Perrimon, M.L. Samuels: Droplet microfluidic technology for single-cell high-throughput screening, Proc. Nat. Acad. Sci. **106**, 14195–14200 (2009)
16.98 S. McNamara, Y.B. Gianchandani: On-chip vacuum generated by a micromachined Knudsen pump, J. Microelectromech. Syst. **14**, 741–746 (2005)
16.99 W. Reisner, J.N. Pedersen, R.H. Austin: DNA confinement in nanochannels: Physics and biological applications, Rep. Prog. Phys. **75**, 106601 (2012)
16.100 T.-C. Kuo, D.M. Cannon, Y. Chen, J.J. Tulock, M.A. Shannon, J.V. Sweedler, P.W. Bohn: Gateable nanofluidic interconnects for multilayered microfluidic separation systems, Anal. Chem. **75**, 1861–1867 (2003)
16.101 S.A. Gajar, M.W. Geis: An ionic liquid-channel field-effect transistor, J. Electrochem. Soc. **139**, 2833–2840 (1992)
16.102 P. Abgrall, N.T. Nguyen: Nanofluidic devices and their applications, Anal. Chem. **80**, 2326–2341 (2008)
16.103 M.A. Gijs: Device physics: Will fluidic electronics take off?, Nat. Nanotechnol. **2**, 268–270 (2007)
16.104 A. Manz, D.J. Harrison, E.M. Verpoorte, J.C. Fettinger, A. Paulus, H. Lüdi, H.M. Widmer: Planar chips technology for miniaturization and integration of separation techniques into monitoring systems: Capillary electrophoresis on a chip, J. Chromatogr. A **593**, 253–258 (1992)
16.105 C. Iliescu, H. Taylor, M. Avram, J. Miao, S. Franssila: A practical guide for the fabrication of microfluidic devices using glass and silicon, Biomicrofluidics **6**, 016505 (2012)
16.106 G.T. Kovacs, N.I. Maluf, K.E. Petersen: Bulk micromachining of silicon, Proc. IEEE **86**, 1536–1551 (1998)
16.107 C.D. James, M. Okandan, S.S. Mani, P.C. Galambos, R. Shul: Monolithic surface micromachined fluidic devices for dielectrophoretic preconcentration and routing of particles, J. Micromech. Microeng. **16**, 1909 (2006)
16.108 M.J. de Boer, R.W. Tjerkstra, J. Berenschot, H.V. Jansen, G. Burger, J. Gardeniers, M. Elwenspoek, A. van den Berg: Micromachining of buried micro channels in silicon, J. Microelectromech. Syst. **9**, 94–103 (2000)
16.109 S.-J. Paik, S. Byun, J.-M. Lim, Y. Park, A. Lee, S. Chung, J. Chang, K. Chun: In-plane single-crystal-silicon microneedles for minimally invasive microfluid systems, Sens. Actuat. A: Phys. **114**,

276–284 (2004)
16.110 N.-T. Nguyen, S.T. Wereley: *Fundamentals and Applications of Microfluidics* (Artech House, Norwood 2002)
16.111 W.I. Wu, P. Rezai, H.H. Hsu, P.R. Selvaganapathy: Materials and methods for the microfabrication of microfluidic biomedical devices. In: *Microfluidic Devices for Biomedical Applications*, ed. by X.-J.J. Li, Y. Zhou (2013) pp. 3–62
16.112 D.J. Laser, J.G. Santiago: A review of micropumps, J. Micromech. Microeng. **14**, R35 (2004)
16.113 K.W. Oh, C.H. Ahn: A review of microvalves, J. Micromech. Microeng. **16**, R13 (2006)
16.114 R. Tiggelaar, P. Van Male, J. Berenschot, J. Gardeniers, R. Oosterbroek, M. De Croon, J. Schouten, A.M. Van Den Berg: Elwenspoek: Fabrication of a high-temperature microreactor with integrated heater and sensor patterns on an ultrathin silicon membrane, Sens. Actuat. A: Phys. **119**, 196–205 (2005)
16.115 Y. Nakashima, T. Yasuda: Cell differentiation guidance using chemical stimulation controlled by a microfluidic device, Sens. Actuat. A: Phys. **139**, 252–258 (2007)
16.116 E.G. Lavoie, T. Wangdi, B.I. Kazmierczak: Innate immune responses to Pseudomonas aeruginosa infection, Microbes Infect **13**, 1133–1145 (2011)
16.117 M.A. Hopcroft, W.D. Nix, T.W. Kenny: What is the young's modulus of silicon?, J. Microelectromech. Syst. **19**, 229–238 (2010)
16.118 K. Yamada, T.G. Henares, K. Suzuki, D. Citterio: Paper-based inkjet-printed microfluidic analytical devices, Angew. Chem. Int. Edn. **54**, 5294–5310 (2015)
16.119 R.R. Anderson, W. Hu, J.W. Noh, W.C. Dahlquist, S.J. Ness, T.M. Gustafson, D.C. Richards, S. Kim, B.A. Mazzeo, A.T. Woolley: Transient deflection response in microcantilever array integrated with polydimethylsiloxane (PDMS) microfluidics, Lab on a Chip **11**, 2088–2096 (2011)
16.120 M.-A. Grétillat, F. Paoletti, P. Thiébaud, S. Roth, M. Koudelka-Hep, N. De Rooij: A new fabrication method for borosilicate glass capillary tubes with lateral inlets and outlets, Sens. Actuat. A **60**, 219–222 (1997)
16.121 F. Gretillat, M.-A. Gretillat, N.F. de Rooij: Improved design of a silicon micromachined gyroscope with piezoresistive detection and electromagnetic excitation, J. Microelectromech. Syst. **8**, 243–250 (1999)
16.122 C. Iliescu, B. Chen: Thick and low-stress PECVD amorphous silicon for MEMS applications, J. Micromech. Microeng. **18**, 015024 (2007)
16.123 J. Brugger, G. Beljakovic, M. Despont, N. De Rooij, P. Vettiger: Silicon micro-nanomechanical device fabrication based on focused ion beam surface modification and KOH etching, Microeleetronic Eng **35**, 401–404 (1997)
16.124 J.S. You, D. Kim, J.Y. Huh, H.J. Park, J.J. Pak, C.S. Kang: Experiments on anisotropic etching of Si in TMAH, Sol. Energy Mater. Sol. Cells **66**, 37–44 (2001)
16.125 G. Spierings: Wet chemical etching of silicate glasses in hydrofluoric acid based solutions, J. Mater. Sci. **28**, 6261–6273 (1993)
16.126 U. Gösele, Q.-Y. Tong: Semiconductor wafer bonding, Annu. Rev. Mater. Sci. **28**, 215–241 (1998)
16.127 M.A. Schmidt: Wafer-to-wafer bonding for microstructure formation, Proc. IEEE **86**, 1575–1585 (1998)
16.128 V.I. Chin, P. Taupin, S. Sanga, J. Scheel, F.H. Gage, S.N. Bhatia: Microfabricated platform for studying stem cell fates, Biotechnol. Bioeng. **88**, 399–415 (2004)
16.129 E.T. Lagally, P.C. Simpson, R.A. Mathies: Monolithic integrated microfluidic DNA amplification and capillary electrophoresis analysis system, Sens. Actuat. B **63**, 138–146 (2000)
16.130 A.C. Lewis, J.F. Hamilton, C.N. Halliday, J. Rhodes, K.D. Bartle, P. Homewood, R.J. Grenfell, B. Goody, A.M. Harling, P. Brewer: Microfabricated planar glass gas chromatography with photoionization detection, J. Chromatogra. A **1217**, 768–774 (2010)
16.131 G.M. Whitesides: The origins and the future of microfluidics, Nature **442**, 368–373 (2006)
16.132 A. Plecis, Y. Chen: Fabrication of microfluidic devices based on glass–PDMS–glass technology, Microelectron. Eng. **84**, 1265–1269 (2007)
16.133 D. Mijatovic, J. Eijkel, A. Van Den Berg: Technologies for nanofluidic systems: Top-down versus bottom up – a review, Lab on a Chip **5**, 492–500 (2005)
16.134 F.H. Labeed, H.O. Fatoyinbo: *Microfluidics in Detection Science: Lab-on-a-Chip Technologies* (Royal Society of Chemistry, London 2014)
16.135 J. Friend, L. Yeo: Fabrication of microfluidic devices using polydimethylsiloxane, Biomicrofluidics **4**, 026502 (2010)
16.136 J.C. McDonald, G.M. Whitesides: Poly (dimethylsiloxane) as a material for fabricating microfluidic devices, Accounts Chem. Res. **35**, 491–499 (2002)
16.137 N. Lobontiu, E. Garcia: *Mechanics of Microelectromechanical Systems* (Springer, New York 2004)
16.138 H. Becker, C. Gärtner: Polymer microfabrication technologies for microfluidic systems, Anal. Bioanal. Chem. **390**, 89–111 (2008)
16.139 D.C. Duffy, J.C. McDonald, O.J. Schueller, G.M. Whitesides: Rapid prototyping of microfluidic systems in poly(dimethylsiloxane), Anal. Chem. **70**, 4974–4984 (1998)
16.140 J.C. McDonald, S.J. Metallo, G.M. Whitesides: Fabrication of a configurable, single-use microfluidic device, Anal. Chem. **73**, 5645–5650 (2001)
16.141 T. Yang, S. Jung, H. Mao, P.S. Cremer: Fabrication of phospholipid bilayer-coated microchannels for on-chip immunoassays, Anal. Chem. **73**, 165–169 (2001)
16.142 R.S. Martin, A.J. Gawron, S.M. Lunte: Dual-electrode electrochemical detection for poly(dimethylsiloxane)-fabricated capillary electrophoresis microchips, Anal. Chem. **72**, 3196–3202 (2000)
16.143 A.W. Martinez, S.T. Phillips, M.J. Butte, G.M. Whitesides: Patterned paper as a plat-

form for inexpensive, low-volume, portable bioassays, Angew. Chem. Int. Edn. **46**, 1318–1320 (2007)
16.144 S. Bhattacharya, A. Datta, J.M. Berg, S. Gangopadhyay: Studies on surface wettability of poly (dimethyl) siloxane (PDMS) and glass under oxygen-plasma treatment and correlation with bond strength, J. Microelectromech. Syst. **14**, 590–597 (2005)
16.145 P. Rezai, P.R. Selvaganapathy, G.R. Wohl: Plasma enhanced bonding of polydimethylsiloxane with parylene and its optimization, J. Micromech. Microeng. **21**, 065024 (2011)
16.146 J. Soo Ko, H.C. Yoon, H. Yang, H.B. Pyo, K. Hyo Chung, S.K. Jin, Y.K. Tae: A polymer-based microfluidic device for immunosensing biochips, Lab Chip **3**, 106–113 (2003)
16.147 M.A. Eddings, M.A. Johnson, B.K. Gale: Determining the optimal PDMS–PDMS bonding technique for microfluidic devices, J. Micromech. Microeng. **18**, 067001 (2008)
16.148 P. Rezai, P.R. Selvaganapathy, G.R. Wohl: Plasma enhanced bonding of polydimethylsiloxane (PDMS) with parylene. In: *Proc. 16th Int. Solid-State Sens., Actuat. Microsyst. Conf.* (2011) pp. 1340–1343
16.149 M. Atai, L. Solhi, A. Nodehi, S.M. Mirabedini, S. Kasraei, K. Akbari, S. Babanzadeh: PMMA-grafted nanoclay as novel filler for dental adhesives, Dent. Mater. **25**, 339–347 (2009)
16.150 P.C. Nicolson, J. Vogt: Soft contact lens polymers: An evolution, Biomaterials **22**, 3273–3283 (2001)
16.151 S.B. Kim, Y.J. Kim, T.L. Yoon, S.A. Park, I.H. Cho, E.J. Kim, I.A. Kim, J.W. Shin: The characteristics of a hydroxyapatite-chitosan-PMMA bone cement, Biomaterials **25**, 5715–5723 (2004)
16.152 M.B. Wabuyele, S.M. Ford, W. Stryjewski, J. Barrow, S.A. Soper: Single molecule detection of double-stranded DNA in poly(methylmethacrylate) and polycarbonate microfluidic devices, Electrophoresis **22**, 3939–3948 (2001)
16.153 G. Jenkins, C.D. Mansfield: *Microfluidic Diagnostics: Methods and Protocols* (Humana, Totowa 2013)
16.154 M. Heckele, W. Bacher, D.K. Müller: Hot embossing – The molding technique for plastic microstructures, Microsyst. Technol. **4**, 122–124 (1998)
16.155 J. Liu, L. Wang, W. Ouyang, W. Wang, J. Qin, Z. Xu, S. Xu, D. Ge, L. Wang, C. Liu, L. Wang: Fabrication of PMMA nanofluidic electrochemical chips with integrated microelectrodes, Biosens. Bioelectron. **72**, 288–293 (2015)
16.156 G. Julien, C. Thierry, M. Patrice: Microinjection molding of thermoplastic polymers: A review, J. Micromech. Microeng. **17**, R96 (2007)
16.157 S.A. Soper, S.M. Ford, S. Qi, R.L. McCarley, K. Kelly, M.C. Murphy: Peer reviewed: Polymeric microelectromechanical systems, Anal. Chem. **72**, 642A–651A (2000)
16.158 H. Klank, J.P. Kutter, O. Geschke: CO_2-laser micromachining and back-end processing for rapid production of PMMA-based microfluidic systems, Lab on a Chip **2**, 242–246 (2002)
16.159 C.-H. Lin, C.-H. Chao, C.-W. Lan: Low azeotropic solvent for bonding of PMMA microfluidic devices, Sens. Actuat. B **121**, 698–705 (2007)
16.160 Z. Chen, Y. Gao, J. Lin, R. Su, Y. Xie: Vacuum-assisted thermal bonding of plastic capillary electrophoresis microchip imprinted with stainless steel template, J. Chromatogra. A **1038**, 39–245 (2004)
16.161 C.-W. Tsao, D.L. DeVoe: Bonding of thermoplastic polymer microfluidics, Microfluid. Nanofluidics **6**, 1–16 (2009)
16.162 S.A. Soper, A.C. Henry, B. Vaidya, M. Galloway, M. Wabuyele, R.L. McCarley: Surface modification of polymer-based microfluidic devices, Anal. Chim. Acta **470**, 87–99 (2002)
16.163 J. Wang, A. Muck Jr., M.P. Chatrathi, G. Chen, N. Mittal, S.D. Spillman, S. Obeidat: Bulk modification of polymeric microfluidic devices, Lab Chip **5**, 226–230 (2005)
16.164 F. Dang, L. Zhang, H. Hagiwara, Y. Mishina, Y. Baba: Ultrafast analysis of oligosaccharides on microchip with light-emitting diode confocal fluorescence detection, Electrophoresis **24**, 714–721 (2003)
16.165 C. Goll, W. Bacher, B. Buestgens, D. Maas, W. Menz, W. Schomburg: Microvalves with bistable buckled polymer diaphragms, J. Micromech. Microeng. **6**, 77 (1996)
16.166 C. Goll, W. Bacher, B. Büstgens, D. Maas, R. Ruprecht, W. Schomburg: An electrostatically actuated polymer microvalve equipped with a movable membrane electrode, J. Micromech. Microeng. **7**, 224 (1997)
16.167 Z. Azeem, P. Andrea, S.E. Winnie, D. Maria: Fabrication of polyimide based microfluidic channels for biosensor devices, J. Micromech. Microeng. **25**, 035022 (2015)
16.168 R.R. Richardson Jr., J.A. Miller, W.M. Reichert: Polyimides as biomaterials: Preliminary biocompatibility testing, Biomaterials **14**, 627–635 (1993)
16.169 I.K. Glasgow, D.J. Beebe, V.E. White: Design rules for polyimide solvent bonding, Sens. Mater. **11**, 269–278 (1999)
16.170 A. Bayrashev, B. Ziaie: Silicon wafer bonding with an insulator interlayer using RF dielectric heating. In: *IEEE Int. Conf* (2002) pp. 419–422
16.171 K. Min, D. Kim: Pressure tolerant multilayered polymer film microfluidics by one-step bonding process for high throughput emulsion generation. In: *16th Int. Conf. Miniaturized Syst. Chem. Life Sci.* (2012) pp. 1798–1800
16.172 B.-K. Zhu, S.-H. Xie, Z.-K. Xu, Y.-Y. Xu: Preparation and properties of the polyimide/multi-walled carbon nanotubes (MWNTs) nanocomposites, Compos. Sci. Technol. **66**, 548–554 (2006)
16.173 X. Jiang, Y. Bin, M. Matsuo: Electrical and mechanical properties of polyimide–carbon nanotubes composites fabricated by in situ polymerization, Polymer **46**, 7418–7424 (2005)

16.174 Y. Zhao, J.T. Yeow, L.A. Zheng: Simple and quick fabrication method of microfluidic cell sorter using Dielectrophoresis. In: *The 7th IEEE Int. Conf. Nano–Molecular Medicine and Engineering* (2013) pp. 32–35
16.175 E.W. Young, E. Berthier, D.J. Guckenberger, E. Sackmann, C. Lamers, I. Meyvantsson, A. Hutenlocher, D.J. Beebe: Rapid prototyping of arrayed microfluidic systems in polystyrene for cell-based assays, Anal. Chem. **83**, 1408–1417 (2011)
16.176 P.M. van Midwoud, A. Janse, M.T. Merema, G.M. Groothuis, E. Verpoorte: Comparison of biocompatibility and adsorption properties of different plastics for advanced microfluidic cell and tissue culture models, Anal. Chem. **84**, 3938–3944 (2012)
16.177 M.-S. Choi, J.-C. Yoo: Automated centrifugal-microfluidic platform for DNA purification using laser burst valve and voriolis effect, Appl. Biochem. Biotechnol. **175**, 3778–3787 (2015)
16.178 D. Ogończyk, J. Węgrzyn, P. Jankowski, B. Dąbrowski, P. Garstecki: Bonding of microfluidic devices fabricated in polycarbonate, Lab Chip **10**, 1324–1327 (2010)
16.179 R.K. Jena, C. Yue: Cyclic olefin copolymer based microfluidic devices for biochip applications: Ultraviolet surface grafting using 2-methacryloyloxyethyl phosphorylcholine, Biomicrofluidics **6**, 012822 (2012)
16.180 J.S. Jeon, S. Chung, R.D. Kamm, J.L. Charest: Hot embossing for fabrication of a microfluidic 3-D cell culture platform, Biomed. Microdevices **13**, 325–333 (2011)
16.181 J. Steigert, S. Haeberle, T. Brenner, C. Müller, C. Steinert, P. Koltay, N. Gottschlich, H. Reinecke, J. Rühe, R. Zengerle: Rapid prototyping of microfluidic chips in COC, J. Micromech. Microeng. **17**, 333 (2007)
16.182 V. Gubala, L.F. Harris, A.J. Ricco, M.X. Tan, D.E. Williams: Point of care diagnostics: Status and future, Anal. Chem. **84**, 487–515 (2011)
16.183 C. Renault, X. Li, S.E. Fosdick, R.M. Crooks: Hollow-channel paper analytical devices, Anal. Chem. **85**, 7976–7979 (2013)
16.184 L. Cai, Y. Wu, C. Xu, Z. Chen: A simple paper-based microfluidic device for the determination of the total amino acid content in a tea leaf extract, J. Chem. Educ. **90**, 232–234 (2012)
16.185 J. Olkkonen, K. Lehtinen, T. Erho: Flexographically printed fluidic structures in paper, Anal. Chem. **82**, 10246–10250 (2010)
16.186 J.S. Miller, K.R. Stevens, M.T. Yang, B.M. Baker, D.-H.T. Nguyen, D.M. Cohen, E. Toro, A.A. Chen, P.A. Galie, X. Yu: Rapid casting of patterned vascular networks for perfusable engineered three-dimensional tissues, Nat. Mater. **11**, 768–774 (2012)
16.187 Y. Xia, J. Si, Z. Li: Fabrication techniques for microfluidic paper-based analytical devices and their applications for biological testing: A review, Biosens. Bioelectron. **77**, 774–789 (2016)
16.188 X. Li, D.R. Ballerini, W. Shen: A perspective on paper-based microfluidics: Current status and future trends, Biomicrofluidics **6**, 011301 (2012)
16.189 K.M. Schilling, A.L. Lepore, J.A. Kurian, A.W. Martinez: Fully enclosed microfluidic paper-based analytical devices, Anal. Chem. **84**, 1579–1585 (2012)
16.190 E.W. Washburn: The dynamics of capillary flow, Phys. Rev. **17**, 273 (1921)
16.191 Y. Lu, W. Shi, L. Jiang, J. Qin, B. Lin: Rapid prototyping of paper-based microfluidics with wax for low-cost, portable bioassay, Electrophoresis **30**, 1497–1500 (2009)
16.192 K. Abe, K. Suzuki, D. Citterio: Inkjet-printed microfluidic multianalyte chemical sensing paper, Anal. Chem. **80**, 6928–6934 (2008)
16.193 X. Li, J. Tian, W. Shen: Progress in patterned paper sizing for fabrication of paper-based microfluidic sensors, Cellulose **17**, 649–659 (2010)
16.194 C. Renault, K. Scida, K.N. Knust, S.E. Fosdick, R.M. Crooks: Paper-based bipolar electrochemistry, J. Electrochem. Sci. Technol. **4**, 146–152 (2013)
16.195 W. Dungchai, O. Chailapakul, C.S. Henry: A low-cost, simple, and rapid fabrication method for paper-based microfluidics using wax screen-printing, Analyst **136**, 77–82 (2011)
16.196 E.M. Fenton, M.R. Mascarenas, G.P. López, S.S. Sibbett: Multiplex lateral-flow test strips fabricated by two-dimensional shaping, ACS Appl. Mater. Interfaces **1**, 124–129 (2008)
16.197 S. Jahanshahi-Anbuhi, P. Chavan, C. Sicard, V. Leung, S.Z. Hossain, R. Pelton, J.D. Brennan, C.D. Filipe: Creating fast flow channels in paper fluidic devices to control timing of sequential reactions, Lab on a Chip **12**, 5079–5085 (2012)
16.198 E. Carrilho, S.T. Phillips, S.J. Vella, A.W. Martinez, G.M. Whitesides: Paper microzone plates, Anal. Chem. **81**, 5990–5998 (2009)
16.199 G. Chitnis, Z. Ding, C.-L. Chang, C.A. Savran, B. Ziaie: Laser-treated hydrophobic paper: An inexpensive microfluidic platform, Lab on a Chip **11**, 1161–1165 (2011)
16.200 C.L. Sones, I.N. Katis, P.J. He, B. Mills, M.F. Namiq, P. Shardlow, M. Ibsen, R.W. Eason: Laser-induced photo-polymerisation for creation of paper-based fluidic devices, Lab Chip **14**, 4567–4574 (2014)
16.201 X. Li, J. Tian, T. Nguyen, W. Shen: Paper-based microfluidic devices by plasma treatment, Anal. Chem. **80**, 9131–9134 (2008)
16.202 A.W. Martinez, S.T. Phillips, B.J. Wiley, M. Gupta, G.M. Whitesides: FLASH: A rapid method for prototyping paper-based microfluidic devices, Lab on a Chip **8**, 2146–2150 (2008)
16.203 E. Carrilho, A.W. Martinez, G.M. Whitesides: Understanding wax printing: A simple micropatterning process for paper-based microfluidics, Anal. Chem. **81**, 7091–7095 (2009)
16.204 A.K. Yetisen, M.S. Akram, C.R. Lowe: Paper-based microfluidic point-of-care diagnostic devices, Lab on a Chip **13**, 2210–2251 (2013)
16.205 W. Wang, W.-Y. Wu, J.-J. Zhu: Tree-shaped paper strip for semiquantitative colorimetric detection of protein with self-calibration, J. Chromatogra A

1217, 3896–3899 (2010)
16.206 W. Dungchai, O. Chailapakul, C.S. Henry: Electrochemical detection for paper-based microfluidics, Anal. Chem. **81**, 5821–5826 (2009)
16.207 L. Ge, S. Wang, S. Ge, J. Yu, M. Yan, N. Li, J. Huang: Electrophoretic separation in a microfluidic paper-based analytical device with an on-column wireless electrogenerated chemiluminescence detector, Chem. Commun. (Camb.) **50**, 5699–5702 (2014)
16.208 L. Wu, C. Ma, L. Ge, Q. Kong, M. Yan, S. Ge, J. Yu: Paper-based electrochemiluminescence origami cyto-device for multiple cancer cells detection using porous AuPd alloy as catalytically promoted nanolabels, Biosens. Bioelectron. **63**, 450–457 (2015)
16.209 L. Ge, J. Yan, X. Song, M. Yan, S. Ge, J. Yu: Three-dimensional paper-based electrochemiluminescence immunodevice for multiplexed measurement of biomarkers and point-of-care testing, Biomaterials **33**, 1024–1031 (2012)
16.210 M.M. Mentele, J. Cunningham, K. Koehler, J. Volckens, C.S. Henry: Microfluidic paper-based analytical device for particulate metals, Anal. Chem. **84**, 4474–4480 (2012)
16.211 P. Rattanarat, W. Dungchai, D. Cate, J. Volckens, O. Chailapakul, C.S. Henry: Multilayer paper-based device for colorimetric and electrochemical quantification of metals, Anal. Chem. **86**, 3555–3562 (2014)
16.212 P. Rattanarat, W. Dungchai, D.M. Cate, W. Siangproh, J. Volckens, O. Chailapakul, C.S. Henry: A microfluidic paper-based analytical device for rapid quantification of particulate chromium, Anal. Chim. Acta **800**, 50–55 (2013)
16.213 D.M. Cate, P. Nanthasurasak, P. Riwkulkajorn, C. L'Orange, C.S. Henry, J. Volckens: Rapid detection of transition metals in welding fumes using paper-based analytical devices, Ann. Occup. **58**, 413–423 (2014)
16.214 L. Feng, X. Li, H. Li, W. Yang, L. Chen, Y. Guan: Enhancement of sensitivity of paper-based sensor array for the identification of heavy-metal ions, Anal. Chim. Acta **780**, 74–80 (2013)
16.215 B.M. Jayawardane, R.W. Cattrall, S.D. Kolev: The use of a polymer inclusion membrane in a paper-based sensor for the selective determination of Cu (II), Anal. Chim. Acta **803**, 106–112 (2013)
16.216 J. Jokerst, J.A. Adkins, B. Bisha, M. Mentele, L. Goodridge, C. Henry: A Paper-based analytical device for the colorimetric detection of foodborne pathogenic bacteria. In: *Proc. 15th Int. Conf. Miniaturized Syst. Chem. Life Sci.* (2011) pp. 2116–2118
16.217 W. Liu, J. Kou, H. Xing, B. Li: Paper-based chromatographic chemiluminescence chip for the detection of dichlorvos in vegetables, Biosens. Bioelectron. **52**, 76–81 (2014)
16.218 Y. Wang, J. Ping, Z. Ye, J. Wu, Y. Ying: Impedimetric immunosensor based on gold nanoparticles modified graphene paper for label-free detection of escherichia coli 0157:H7, Biosens. Bioelectron. **49**, 492–498 (2013)
16.219 J. Shi, F. Tang, H. Xing, H. Zheng, B. Lianhua, W. Wei: Electrochemical detection of Pb and Cd in paper-based microfluidic devices, J. Braz. Chem. Soc. **23**, 1124–1130 (2012)
16.220 D.M. Cate, J.A. Adkins, J. Mettakoonpitak, C.S. Henry: Recent developments in paper-based microfluidic devices, Anal. Chem. **87**, 19–41 (2014)
16.221 W.K. Tomazelli Coltro, C.M. Cheng, E. Carrilho, D.P. Jesus: Recent advances in low-cost microfluidic platforms for diagnostic applications, Electrophoresis **35**, 2309–2324 (2014)
16.222 D.R. Ballerini, X. Li, W. Shen: Patterned paper and alternative materials as substrates for low-cost microfluidic diagnostics, Microfluid. Nanofluidics **13**, 769–787 (2012)
16.223 A. Nilghaz, D. Ballerini, W. Shen: Exploration of microfluidic devices based on multi-filament threads and textiles: A review, Biomicrofluidics **7**, 051501 (2013)
16.224 A. Nilghaz, D.R. Ballerini, X.-Y. Fang, W. Shen: Semiquantitative analysis on microfluidic thread-based analytical devices by ruler, Sens. Actuat. B: Chem. **191**, 586–594 (2014)
16.225 X. Li, J. Tian, W. Shen: Thread as a versatile material for low-cost microfluidic diagnostics, ACS Appl. Mater. Interfaces **2**, 1–6 (2009)
16.226 K.T. Hodgson, J.C. Berg: The effect of surfactants on wicking flow in fiber networks, J. Colloid Interface Sci. **121**, 22–31 (1988)
16.227 N. Wang, A. Zha, J. Wang: Study on the wicking property of polyester filament yarns, Fibers Polym **9**, 97–100 (2008)
16.228 D.R. Ballerini, X. Li, W. Shen: Flow control concepts for thread-based microfluidic devices, Biomicrofluidics **5**, 014105 (2011)
16.229 P. Bhandari, T. Narahari, D. Dendukuri: Fab-Chips: A versatile, fabric-based platform for low-cost, rapid and multiplexed diagnostics, Lab on a Chip **11**, 2493–2499 (2011)
16.230 R. Safavieh, G.Z. Zhou, D. Juncker: Microfluidics made of yarns and knots: From fundamental properties to simple networks and operations, Lab on a Chip **11**, 2618–2624 (2011)
16.231 G.O.F. Parikesit, F. Prasetia, G.A. Pribadi, D.C. Simbolon, G.Y. Pradhana, A.R. Prastowo, A. Gunawan, K. Suryopratomo, I. Kusumaningtyas: Textile-based microfluidics: Modulated wetting, mixing, sorting, and energy harvesting, J. Text. Inst. **103**, 1077–1087 (2012)
16.232 S. Xing, J. Jiang, T. Pan: Interfacial microfluidic transport on micropatterned superhydrophobic textile, Lab on a Chip **13**, 1937–1947 (2013)
16.233 T. Liu, K.-F. Choi, Y. Li: Wicking in twisted yarns, J. Colloid Interface Sci. **318**, 134–139 (2008)
16.234 J. Wiener, P. Dejlová: Wicking and wetting in textiles, Autex Res. J. **3**, 64–71 (2003)
16.235 J. Ráhel', M. Šimor, M. Černák, M. Štefečka, Y. Imahori, M. Kando: Hydrophilization of polypropylene nonwoven fabric using surface barrier discharge, Surf. Coat. Technol. **169/170**, 604–608 (2003)

16.236 M. Reches, K.A. Mirica, R. Dasgupta, M.D. Dickey, M.J. Butte, G.M. Whitesides: Thread as a matrix for biomedical assays, ACS Appl. Mater. Interfaces **2**, 1722–1728 (2010)
16.237 D.R. Ballerini, X. Li, W. Shen: An inexpensive thread-based system for simple and rapid blood grouping, Anal. Bioanal. Chem. **399**, 1869–1875 (2011)
16.238 A. Nilghaz, D.H. Wicaksono, D. Gustiono, F.A.A. Majid, E. Supriyanto, M.R.A. Kadir: Flexible microfluidic cloth-based analytical devices using a low-cost wax patterning technique, Lab Chip **12**, 209–218 (2012)
16.239 R.S. Malon, K.Y. Chua, D.H. Wicaksono, E.P. Corcoles: Cotton fabric-based electrochemical device for lactate measurement in saliva, Analyst **139**, 3009–3016 (2014)
16.240 W. Guan, C. Zhang, F. Liu, M. Liu: Chemiluminescence detection for microfluidic cloth-based analytical devices (muCADs), Biosens. Bioelectron. **72**, 114–120 (2015)
16.241 W. Guan, M. Liu, C. Zhang: Electrochemiluminescence detection in microfluidic cloth-based analytical devices, Biosens. Bioelectron. **75**, 247–253 (2016)
16.242 T. Choudhary, G.P. Rajamanickam, D. Dendukuri: Woven electrochemical fabric-based test sensors (WEFTS): A new class of multiplexed electrochemical sensors, Lab Chip **15**, 2064–2072 (2015)
16.243 G. Zhou, X. Mao, D. Juncker: Immunochromatographic assay on thread, Anal. Chem. **84**, 7736–7743 (2012)
16.244 D. Agustini, M.F. Bergamini, L.H. Marcolino-Junior: Low cost microfluidic device based on cotton threads for electroanalytical application, Lab Chip **16**, 345–352 (2016)
16.245 A. Nilghaz, D.R. Ballerini, L. Guan, L. Li, W. Shen: Red blood cell transport mechanisms in polyester thread-based blood typing devices, Anal. Bioanal. Chem. **408**, 1365–1371 (2016)
16.246 Y.-C. Wei, L.-M. Fu, C.-H. Lin: Electrophoresis separation and electrochemical detection on a novel thread-based microfluidic device, Microfluid. Nanofluid. **14**, 583–590 (2013)
16.247 B.G. Chung, K.H. Lee, A. Khademhosseini, S.H. Lee: Microfluidic fabrication of microengineered hydrogels and their application in tissue engineering, Lab Chip **12**, 45–59 (2012)
16.248 X. Zhang, L. Li, C. Luo: Gel integration for microfluidic applications, Lab Chip **16**, 1757–1776 (2016)
16.249 B.V. Slaughter, S.S. Khurshid, O.Z. Fisher, A. Khademhosseini, N.A. Peppas: Hydrogels in regenerative medicine, Adv. Mater. **21**, 3307–3329 (2009)
16.250 N.A. Peppas, B. Kim: Stimuli-sensitive protein delivery systems, J. Drug Deliv. Sci. Technol. **16**, 11–18 (2006)
16.251 S.Y. Cheng, S. Heilman, M. Wasserman, S. Archer, M.L. Shuler, M. Wu: A hydrogel-based microfluidic device for the studies of directed cell migration, Lab Chip **7**, 763–769 (2007)
16.252 B.R. Lee, J.W. Hwang, Y.Y. Choi, S.F. Wong, Y.H. Hwang, D.Y. Lee, S.H. Lee: In situ formation and collagen-alginate composite encapsulation of pancreatic islet spheroids, Biomaterials **33**, 837–845 (2012)
16.253 F.R. Walter, S. Valkai, A. Kincses, A. Petneházi, T. Czeller, S. Veszelka, P. Ormos, M.A. Deli, A. Dér: A versatile lab-on-a-chip tool for modeling biological barriers, Sens. Actuat. B: Chem. **222**, 1209–1219 (2016)
16.254 J. Wan: Microfluidic-based synthesis of hydrogel particles for cell microencapsulation and cell-based drug delivery, Polymers **4**, 1084–1108 (2012)
16.255 G.D. Nicodemus, S.J. Bryant: Cell encapsulation in biodegradable hydrogels for tissue engineering applications, Tissue Eng. Part B: Rev. **14**, 149–165 (2008)
16.256 P. Worthington, D.J. Pochan, S.A. Langhans: Peptide hydrogels-versatile matrices for 3-D cell culture in cancer medicine, Front. Oncol. **5**, 92–92 (2014)
16.257 E.M. Ahmed: Hydrogel: Preparation, characterization, and applications: A review, J. Adv. Res. **6**, 105–121 (2015)
16.258 B. Amsden: Solute diffusion within hydrogels, mechanisms and models, Macromolecules **31**, 8382–8395 (1998)
16.259 J.-I. Horinaka, R. Yasuda, T. Takigawa: Rheological properties of concentrated solutions of galactomannans in an ionic liquid, Carbohydr. Polym. **89**, 1018–1021 (2012)
16.260 H. Zhang, W. Davison: Diffusional characteristics of hydrogels used in DGT and DET techniques, Anal. Chim. Acta **398**, 329–340 (1999)
16.261 V. Normand, D.L. Lootens, E. Amici, K.P. Plucknett, P. Aymard: New insight into agarose gel mechanical properties, Biomacromolecules **1**, 730–738 (2000)
16.262 D. Kim, D.J. Beebe: Hydrogel-based reconfigurable components for microfluidic devices, Lab on a Chip **7**, 193–198 (2007)
16.263 C. Luo, X. Ni, L. Liu, S.I.M. Nomura, Y. Chen: Degassing-assisted patterning of cell culture surfaces, Biotechnol. Bioeng. **105**, 854–859 (2010)
16.264 L.M. Bellan, M. Pearsall, D.M. Cropek, R. Langer: A 3-D interconnected microchannel network formed in gelatin by sacrificial shellac microfibers, Adv. Mater. **24**, 5187–5191 (2012)
16.265 X.-Y. Wang, Z.-H. Jin, B.-W. Gan, S.-W. Lv, M. Xie, W.-H. Huang: Engineering interconnected 3-D vascular networks in hydrogels using molded sodium alginate lattice as the sacrificial template, Lab on a Chip **14**, 2709–2716 (2014)
16.266 L.M. Bellan, S.P. Singh, P.W. Henderson, T.J. Porri, H.G. Craighead, J.A. Spector: Fabrication of an artificial 3-dimensional vascular network using sacrificial sugar structures, Soft Matter **5**, 1354–1357 (2009)
16.267 A.P. Wong, R. Perez-Castillejos, J.C. Love, G.M. Whitesides: Partitioning microfluidic channels with hydrogel to construct tunable 3-D cellular microenvironments, Biomaterials **29**,

1853–1861 (2008)
16.268 S. Chung, R. Sudo, P.J. Mack, C.-R. Wan, V. Vickerman, R.D. Kamm: Cell migration into scaffolds under co-culture conditions in a microfluidic platform, Lab on a Chip **9**, 269–275 (2009)
16.269 H. Xu, M.M. Ferreira, S.C. Heilshorn: Small-molecule axon-polarization studies enabled by a shear-free microfluidic gradient generator, Lab on a Chip **14**, 2047–2056 (2014)
16.270 S. Park, D. Kim, S.Y. Ko, J.-O. Park, S. Akella, B. Xu, Y. Zhang, S. Fraden: Controlling uniformity of photopolymerized microscopic hydrogels, Lab on a Chip **14**, 1551–1563 (2014)
16.271 N. Zaari, P. Rajagopalan, S.K. Kim, A.J. Engler, J.Y. Wong: Photopolymerization in microfluidic gradient generators: microscale control of substrate compliance to manipulate cell response, Adv. Mater. **16**, 2133–2137 (2004)
16.272 D.J. Beebe, J.S. Moore, J.M. Bauer, Q. Yu, R.H. Liu, C. Devadoss, B.-H. Jo: Functional hydrogel structures for autonomous flow control inside microfluidic channels, Nature **404**, 588–590 (2000)
16.273 A. Aung, I.S. Bhullar, J. Theprungsirikul, S.K. Davey, H.L. Lim, Y.-J. Chiu, X. Ma, S. Dewan, Y.-H. Lo, A. McCulloch: 3D cardiac μtissues within a microfluidic device with real-time contractile stress readout, Lab on a Chip **16**, 153–162 (2016)
16.274 S. Bersini, J.S. Jeon, G. Dubini, C. Arrigoni, S. Chung, J.L. Charest, M. Moretti, R.D. Kamm: A microfluidic 3D in vitro model for specificity of breast cancer metastasis to bone, Biomaterials **35**, 2454–2461 (2014)
16.275 J. Tiren, L. Tenerz, B. Hök: A batch-fabricated non-reverse valve with cantilever beam manufactured by micromachining of silicon, Sens. Actuat. **18**, 389–396 (1989)
16.276 C.Y. Li, K.R. Stevens, R.E. Schwartz, B.S. Alejandro, J.H. Huang, S.N. Bhatia: Micropatterned cell–cell interactions enable functional encapsulation of primary hepatocytes in hydrogel microtissues, Tissue Eng. Part A **20**, 2200–2212 (2014)
16.277 S. Trkov, G. Eng, R. Di Liddo, P.P. Parnigotto, G. Vunjak-Novakovic: Micropatterned 3-dimensional hydrogel system to study human endothelial mesenchymal stem cell interactions, J. Tissue Eng. Regen. Medi. **4**, 205 (2010)
16.278 J.H. Sung, M.L. Shuler: A micro cell culture analog (μCCA) with 3-D hydrogel culture of multiple cell lines to assess metabolism-dependent cytotoxicity of anti-cancer drugs, Lab on a Chip **9**, 1385–1394 (2009)
16.279 M. Verhulsel, M. Vignes, S. Descroix, L. Malaquin, D.M. Vignjevic, J.-L. Viovy: A review of microfabrication and hydrogel engineering for micro-organs on chips, Biomaterials **35**, 1816–1832 (2014)
16.280 H.-F. Li, J.-M. Lin: Applications of microfluidic systems in environmental analysis, Anal. Bioanal. Chem. **393**, 555–567 (2009)
16.281 C.D. Chin, T. Laksanasopin, Y.K. Cheung, D. Steinmiller, V. Linder, H. Parsa, J. Wang, H. Moore, R. Rouse, G. Umviligihozo, E. Karita, L. Mwambarangwe, S.L. Braunstein, J. van de Wijgert, R. Sahabo, J.E. Justman, W. El-Sadr, S.K. Sia: Microfluidics-based diagnostics of infectious diseases in the developing world, Nat. Med. **17**, 1015–1019 (2011)
16.282 P.S. Dittrich, A. Manz: Lab-on-a-chip: Microfluidics in drug discovery, Nat. Rev. Drug. Discov. **5**, 210–218 (2006)
16.283 I.U. Khan, C.A. Serra, N. Anton, T.F. Vandamme: Production of nanoparticle drug delivery systems with microfluidics tools, Expert Opin. Drug. Deliv. **12**, 547–562 (2015)
16.284 A.J. Demello: Control and detection of chemical reactions in microfluidic systems, Nature **442**, 394–402 (2006)
16.285 P.M. Valencia, E.M. Pridgen, M. Rhee, R. Langer, O.C. Farokhzad, R. Karnik: Microfluidic platform for combinatorial synthesis and optimization of targeted nanoparticles for cancer therapy, ACS Nano **7**, 10671–10680 (2013)
16.286 P.S. Glockner, G.F. Naterer: Recent advances in nano-electromechanical and microfluidic power generation, Int. J. Energy Res. **31**, 603–618 (2007)
16.287 A.B. Dababneh, I.T. Ozbolat: Bioprinting technology: A current state-of-the-art review, J. Manuf. Sci. Eng. **136**, 061016 (2014)
16.288 G. Cummins, M.P.Y. Desmulliez: Inkjet printing of conductive materials: A review, Circuit World **38**, 193–213 (2012)
16.289 R.P. Tortorich, J.-W. Choi: Inkjet printing of carbon nanotubes, Nanomaterials **3**, 453–468 (2013)
16.290 H.P. Le: Progress and trends in ink-jet printing technology, J. Imaging Sci. Technol. **42**, 49–62 (1998)
16.291 A.K. Au, H. Lai, B.R. Utela, A. Folch: Microvalves and micropumps for biomems, Micromachines **2**, 179–220 (2011)
16.292 M.W. Ashraf, S. Tayyaba, N. Afzulpurkar: Micro electromechanical systems (MEMS) based microfluidic devices for biomedical applications, Int. J. Mol. Sci. **12**, 3648–3704 (2011)
16.293 IndustryARC: Dispensing systems market by technology by end user application 2020, http://industryarc.com/Report/250/Dispensing-systems-Market-analysis-report.html (2016)
16.294 D.C.S. Bien, S.J.N. Mitchell, H.S. Gamble: Fabrication and characterization of a micromachined passive valve, J. Micromech. Microeng. **13**, 557 (2003)
16.295 E.F. Hasselbrink, T.J. Shepodd, J.E. Rehm: High-pressure microfluidic control in lab-on-a-chip devices using mobile polymer monoliths, Anal. Chem. **74**, 4913–4918 (2002)
16.296 M. Yamada, M. Seki: Nanoliter-sized liquid dispenser array for multiple biochemical analysis in microfluidic devices, Anal. Chem. **76**, 895–899 (2004)
16.297 H. Andersson, W. van der Wijngaart, P. Griss, F. Niklaus, G. Stemme: Hydrophobic valves of plasma deposited octafluorocyclobutane in DRIE channels, Sens. Actuat B: Chem. **75**, 136–141 (2001)
16.298 K. Yanagisawa, H. Kuwano, A. Tago: Electromagetically driven microvalve, Microsyst. Technol.

2, 22–25 (1995)
16.299 B. Byunghoon, K. Nakhoon, K. Hongseok, K. Seon-Ho, L. Yeon, L. Sangho, P. Kyihwan: Feasibility test of an electromagnetically driven valve actuator for glaucoma treatment, J. Microelectromech. Syst. **11**, 344–354 (2002)
16.300 K.W. Oh, R. Rong, C.H. Ahn: Miniaturization of pinch-type valves and pumps for practical micro total analysis system integration, J. Micromech. Microeng. **15**(12), 2449 (2005)
16.301 J. Schaible, J. Vollmer, R. Zengerle, H. Sandmaier, T. Strobelt: Electrostatic microvalves in silicon with 2-way-function for industrial applications. In: *Transducers' 01 Eurosensors XV*, ed. by E. Obermeier (Springer, Berlin, Heidelberg 2001) pp. 900–903
16.302 T. Rogge, Z. Rummler, W.K. Schomburg: Polymer micro valve with a hydraulic piezo-drive fabricated by the Amanda process, Sens. Actuat. A: Phys. **110**, 206–212 (2004)
16.303 P.W. Barth: Silicon microvalves for gas flow control. In: *The 8th Int. Conf. Solid-State Sensors and Actuators, 1995 and Eurosensors IX Transducers* (1995) pp. 276–279
16.304 C.A. Rich, K.D. Wise: A high-flow thermopneumatic microvalve with improved efficiency and integrated state sensing, J. Microelectromech. Syst. **12**, 201–208 (2003)
16.305 M. Kohl, D. Dittmann, E. Quandt, B. Winzek, S. Miyazaki, D.M. Allen: Shape memory microvalves based on thin films or rolled sheets, Mater. Sci. Eng. A **273–275**, 784–788 (1999)
16.306 C.R. Neagu, J.G.E. Gardeniers, M. Elwenspoek, J.J. Kelly: Electrochemical microsystem technologiesan electrochemical active valve, Electrochim. Acta **42**, 3367–3373 (1997)
16.307 R.H. Liu, J. Bonanno, J. Yang, R. Lenigk, P. Grodzinski: Single-use, thermally actuated paraffin valves for microfluidic applications, Sens. Actuat. B: Chem. **98**, 328–336 (2004)
16.308 B.J. Kirby, T.J. Shepodd, E.F. Hasselbrink Jr.: Voltage-addressable on–off microvalves for high-pressure microchip separations, J. Chromatogr. A **979**, 147–154 (2002)
16.309 J. Döpper, M. Clemens, W. Ehrfeld, S. Jung, K.P. Kämper, H. Lehr: Micro gear pumps for dosing of viscous fluids, J. Micromech. Microeng. **7**, 230 (1997)
16.310 R. Zengerle, J. Ulrich, S. Kluge, M. Richter, A. Richter: A bidirectional silicon micropump, Sens. Actuat. A: Phys. **50**, 81–86 (1995)
16.311 P. Dario, N. Croce, M.C. Carrozza, G. Varallo: A fluid handling system for a chemical microanalyzer, J. Micromech. Microeng. **6**, 95 (1996)
16.312 H. Andersson, W. van der Wijngaart, P. Nilsson, P. Enoksson, G. Stemme: A valve-less diffuser micropump for microfluidic analytical systems, Sens. Actuat. B: Chem. **72**, 259–265 (2001)
16.313 H.T.G. van Lintel, F.C.M. van De Pol, S. Bouwstra: A piezoelectric micropump based on micromachining of silicon, Sens. Actuat. **15**, 153–167 (1988)
16.314 E. Stemme, G. Stemme: A valveless diffuser–nozzle-based fluid pump, Sens. Actuat. A: Phys. **39**, 159–167 (1993)
16.315 F.K. Forster, R.L. Bardell, M.A. Afromowitz, N.R. Sharma, A. Blanchard: Design, fabrication and testing of fixed-valve micro-pumps, ASME-Publications-FED **234**, 39–44 (1995)
16.316 T. Gerlach, H. Wurmus: Working principle and performance of the dynamic micropump, Sens. Actuat. A: Phys. **50**, 135–140 (1995)
16.317 O. Anders, E. Peter, S. Göran, S. Erik: A valve-less planar pump isotropically etched in silicon, J. Micromech. Microeng. **6**, 87 (1996)
16.318 M. Koch, N. Harris, A.G.R. Evans, N.M. White, A. Brunnschweiler: A novel micromachined pump based on thick-film piezoelectric actuation, Sens. Actuat. A: Phys. **70**, 98–103 (1998)
16.319 L. Weber: Micropump and method for the production thereof, US Patent (Application) 7771176 (2010)
16.320 H.T. Van Lintel, D. Maillefer, S. Gamper: Micromachined fluidic device and method for making same, US Patent (Application) CA241030611 (2001)
16.321 J.-H. Tsai, L. Lin: A thermal-bubble-actuated micronozzle-diffuser pump, J. Microelectromech. Syst. **11**, 665–671 (2002)
16.322 S. Böhm, B. Timmer, W. Olthuis, P. Bergveld: A closed-loop controlled electrochemically actuated micro-dosing system, J. Micromech. Microeng. **10**(4), 498 (2000)
16.323 G. Fuhr, T. Schnelle, B. Wagner: Travelling wave-driven microfabricated electrohydrodynamic pumps for liquids, J. Micromech. Microeng. **4**, 217 (1994)
16.324 J. Jang, S.S. Lee: Theoretical and experimental study of MHD (magnetohydrodynamic) micropump, Sens. Actuat. A: Phys. **80**, 84–89 (2000)
16.325 S. Debesset, C. Hayden, C. Dalton, J. Eijkel, A. Manz: An AC electroosmotic micropump for circular chromatographic applications, Lab Chip **4**, 396–400 (2004)
16.326 R. Huter: *Zeta Potential in Colloid Science* (Academic, London 1981)
16.327 P. Luginbuhl, S.D. Collins, G.A. Racine, M.A. Gretillat, N.F.D. Rooij, K.G. Brooks, N. Setter: Microfabricated Lamb wave device based on PZT sol-gel thin film for mechanical transport of solid particles and liquids, J. Microelectromech. Syst. **6**, 337–346 (1997)
16.328 K.S. Yun, I.J. Cho, J.U. Bu, C.J. Kim, E. Yoon: A surface-tension driven micropump for low-voltage and low-power operations, J. Microelectromech. Syst. **11**(5), 454–461 (2002)
16.329 T. Wang, Q. Ni, N. Crane, R. Guldiken: Surface acoustic wave based pumping in a microchannel, Microsyst. Technol. **23**, 1–8 (2016)
16.330 A.E. Kamholz, B.H. Weigl, B.A. Finlayson, P. Yager: Quantitative analysis of molecular interaction in a microfluidic channel: The T-sensor, Anal. Chem. **71**, 5340–5347 (1999)
16.331 R.F. Ismagilov, A.D. Stroock, P.J.A. Kenis, G. Whitesides, H.A. Stone: Experimental and theoretical scaling laws for transverse diffusive broadening in

two-phase laminar flows in microchannels, Appl. Phys. Lett. **76**, 2376–2378 (2000)
16.332 S.H. Wong, M.C.L. Ward, C.W. Wharton: Micro T-mixer as a rapid mixing micromixer, Sens. Actuat. B: Chem. **100**, 359–379 (2004)
16.333 M. Yi, H.H. Bau: The kinematics of bend-induced mixing in micro-conduits, Int. J. Heat Fluid Flow **24**, 645–656 (2003)
16.334 G.F. Bessoth, J.A. deMello, A. Manz: Microstructure for efficient continuous flow mixing, Anal. Commun. **36**, 213–215 (1999)
16.335 J.B. Knight, A. Vishwanath, J.P. Brody, R.H. Austin: Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds, Phys. Rev. Lett. **80**, 3863–3866 (1998)
16.336 R. Miyake, T.S.J. Lammerink, M. Elwenspoek, J.H.J. Fluitman: Micro mixer with fast diffusion. In: *MEMS* (1993) pp. 248–253
16.337 Y. Lin, G.J. Gerfen, D.L. Rousseau, S.-R. Yeh: Ultrafast microfluidic mixer and freeze-quenching device, Anal. Chem. **75**, 5381–5386 (2003)
16.338 C.-C. Hong, J.-W. Choi, C.H. Ahn: A novel in-plane passive microfluidic mixer with modified Tesla structures, Lab on a Chip **4**, 109–113 (2004)
16.339 A.D. Stroock, S.K.W. Dertinger, A. Ajdari, I. Mezić, H.A. Stone, G.M. Whitesides: Chaotic mixer for microchannels, Science **295**, 647–651 (2002)
16.340 K. Hosokawa, T. Fujii, I. Endo: Droplet-based nano–picoliter mixer using hydrophobic microcapillary vent. In: *Proc. MEMS'99* (1999) pp. 388–393
16.341 A.A. Deshmukh, D. Liepmann, A.P. Pisano: Continuous micromixer with pulsatile micropumps. In: *Technical Digest of the IEEE Solid State Sensor and Actuator Workshop, Hilton Head Island* (2000)
16.342 A.O. El Moctar, N. Aubry, J. Batton: Electro-hydrodynamic micro-fluidic mixer, Lab on a Chip **3**, 273–280 (2003)
16.343 J. Deval, P. Tabeling, C.-M. Ho: A dielectrophoretic chaotic mixer. In: *Proc. MEMS'02* (2002) pp. 36–39
16.344 Z. Tang, S. Hong, D. Djukic, V. Modi, A.C. West, J. Yardley, R.M. Osgood: Electrokinetic flow control for composition modulation in a microchannel, J. Micromech. Microeng. **12**(6), 870 (2002)
16.345 H.H. Bau, J. Zhong, M. Yi: A minute magneto hydro dynamic (MHD) mixer, Sens. Actuat. B: Chem. **79**, 207–215 (2001)
16.346 V. Vivek, E.S. Kim: Novel acoustic-wave micromixer. In: *Proc. MEMS* (2000) pp. 668–673
16.347 J.-H. Tsai, L. Lin: A thermal-bubble-actuated micronozzle-diffuser pump, J. Microelectromech. Syst. **11**(6), 665–671 (2002)
16.348 K.S. Elvira, X. Casadevall i Solvas, R.C. Wootton, A.J. deMello: The past, present and potential for microfluidic reactor technology in chemical synthesis, Nat. Chem. **5**, 905–915 (2013)
16.349 T. Rodrigues, P. Schneider, G. Schneider: Accessing new chemical entities through microfluidic systems, Angew. Chem. Int. Ed. Engl. **53**, 5750–5758 (2014)
16.350 S. Das, V.C. Srivastava: Microfluidic-based photocatalytic microreactor for environmental application: A review of fabrication substrates and techniques, and operating parameters, Photochem. Photobiol. Sci. **15**, 714–730 (2016)
16.351 Y. Asano, S. Togashi, H. Tsudome, S. Murakami: Microreactor technology: Innovations in production processes, Pharm. Eng. **30**, 32 (2010)
16.352 D. Roberge: Lonza–hazardous flow chemistry for streamlined large scale synthesis, Green Process. Synth. **1**, 129–130 (2012)
16.353 T. Frank: Fabrication and assembling of microreactors made from glass and silicon. In: *Microreactors in Organic Synthesis and Catalysis*, ed. by T. Wirth (Wiley, Weinheim 2008) pp. 19–41
16.354 C.-C. Lee, G. Sui, A. Elizarov, C.J. Shu, Y.-S. Shin, A.N. Dooley, J. Huang, A. Daridon, P. Wyatt, D. Stout: Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics, Science **310**, 1793–1796 (2005)
16.355 J.S. Moore, K.F. Jensen: Automated multitrajectory method for reaction optimization in a microfluidic system using online IR analysis, Organ. Process. Res. Dev. **16**, 1409–1415 (2012)
16.356 P.R. Selvaganapathy, E.T. Carlen, C.H. Mastrangelo: Recent progress in microfluidic devices for nucleic acid and antibody assays, Proc. IEEE **91**, 954–975 (2003)
16.357 Y. Zhang, H.R. Jiang: A review on continuous-flow microfluidic PCR in droplets: Advances, challenges and future, Anal. Chim. Acta **914**, 7–16 (2016)
16.358 C. Zhang, J. Xu, W. Ma, W. Zheng: PCR microfluidic devices for DNA amplification, Biotechnol. Adv. **24**, 243–284 (2006)
16.359 N.R. Beer, B.J. Hindson, E.K. Wheeler, S.B. Hall, K.A. Rose, I.M. Kennedy, B.W. Colston: On-chip, real-time, single-copy polymerase chain reaction in picoliter droplets, Anal. Chem. **79**, 8471–8475 (2007)
16.360 J. Daniel, S. Iqbal, R. Millington, D. Moore, C. Lowe, D. Leslie, M. Lee, M. Pearce: Silicon microchambers for DNA amplification, Sens. Actuat A: Phys. **71**, 81–88 (1998)
16.361 J. Khandurina, T.E. McKnight, S.C. Jacobson, L.C. Waters, R.S. Foote, J.M. Ramsey: Integrated system for rapid PCR-based DNA analysis in microfluidic devices, Anal. Chem. **72**, 2995–3000 (2000)
16.362 C.G. Koh, W. Tan, M.-Q. Zhao, A.J. Ricco, Z.H. Fan: Integrating polymerase chain reaction, valving, and electrophoresis in a plastic device for bacterial detection, Anal. Chem. **75**, 4591–4598 (2003)
16.363 C.T. Wittwer, G.C. Fillmore, D. Hillyard: Automated polymerase chain reaction in capillary tubes with hot air, Nucl. Acids Res. **17**, 4353–4357 (1989)
16.364 R. Oda, M. Strausbauch, A. Huhmer, N. Borson, S. Jurrens, J. Craighead, P.J. Wettstein, B. Eckloff, B. Kline, J. Landers: Infrared-mediated thermocycling for ultrafast polymerase chain reaction amplification of DNA, Anal. Chem. **70**, 4361–4368 (1998)

16.365 Y. Tanaka, M.N. Slyadnev, A. Hibara, M. Tokeshi, T. Kitamori: Non-contact photothermal control of enzyme reactions on a microchip by using a compact diode laser, J. Chromatogr. A **894**, 45–51 (2000)
16.366 C. Fermér, P. Nilsson, M. Larhed: Microwave-assisted high-speed PCR, Eur. J. Pharm. Sci. **18**, 129–132 (2003)
16.367 P. Belgrader, W. Benett, D. Hadley, J. Richards, P. Stratton, R. Mariella, F. Milanovich: PCR detection of bacteria in 7 minutes, Science **284**, 449–450 (1999)
16.368 J. Liu, M. Enzelberger, S. Quake: A nanoliter rotary device for polymerase chain reaction, Electrophoresis **23**, 1531–1536 (2002)
16.369 M. Bu, T. Melvin, G. Ensell, J.S. Wilkinson, A.G. Evans: Design and theoretical evaluation of a novel microfluidic device to be used for PCR, J. Micromech. Microeng. **13**, S125 (2003)
16.370 K. Perez-Toralla, D. Pekin, J.-F. Bartolo, F. Garlan, P. Nizard, P. Laurent-Puig, J.-C. Baret, V. Taly: PCR digitale en micro-compartiments, Med. Sci. **31**, 84–92 (2015)
16.371 Z. Hua, J.L. Rouse, A.E. Eckhardt, V. Srinivasan, V.K. Pamula, W.A. Schell, J.L. Benton, T.G. Mitchell, M.G. Pollack: Multiplexed real-time polymerase chain reaction on a digital microfluidic platform, Anal. Chem. **82**, 2310–2316 (2010)
16.372 D. Pekin, Y. Skhiri, J.-C. Baret, D. Le Corre, L. Mazutis, C.B. Salem, F. Millot, A. El Harrak, J.B. Hutchison, J.W. Larson: Quantitative and sensitive detection of rare mutations using droplet-based microfluidics, Lab on a Chip **11**, 2156–2166 (2011)
16.373 M. Leman, F. Abouakil, A.D. Griffiths, P. Tabeling: Droplet-based microfluidics at the femtolitre scale, Lab Chip **15**, 753–765 (2015)
16.374 MarketsandMarkets: Digital PCR and qPCR market by product, application, technology and end user-2021, http://www.marketsandmarkets.com/Market-Reports/digital-pcr-market-174151204.html (2016)
16.375 C.-W. Kan, C.P. Fredlake, E.A.S. Doherty, A.E. Barron: DNA sequencing and genotyping in miniaturized electrophoresis systems, Electrophoresis **25**, 3564–3588 (2004)
16.376 Y. Zhang, P. Ozdemir: Microfluidic DNA amplification – A review, Anal. Chim. Acta **638**, 115–125 (2009)
16.377 B.M. Paegel, R.G. Blazej, R.A. Mathies: Microfluidic devices for DNA sequencing: sample preparation and electrophoretic analysis, Curr. Opin. Biotechnol. **14**, 42–50 (2003)
16.378 M.L. Metzker: Emerging technologies in DNA sequencing, Genome Res **15**, 1767–1776 (2005)
16.379 S. Goodwin, J.D. McPherson, W.R. McCombie: Coming of age: Ten years of next-generation sequencing technologies, Nat. Rev. Genet. **17**, 333–351 (2016)
16.380 B.P. Hodkinson, E.A. Grice: Next-generation sequencing: A review of technologies and tools for wound microbiome research, Adv. Wound Care (New Rochelle) **4**, 50–58 (2015)
16.381 E.R. Mardis: Next-generation sequencing platforms, Annu. Rev. Anal. Chem. (Palo Alto Calif.) **6**, 287–303 (2013)
16.382 H.P. Buermans, J.T. den Dunnen: Next generation sequencing technology: Advances and applications, Biochim. Biophys. Acta **1842**, 1932–1941 (2014)
16.383 M. Kircher, J. Kelso: High-throughput DNA sequencing – Concepts and limitations, BioEssays **32**, 524–536 (2010)
16.384 M.L. Metzker: Sequencing technologies the next generation, Nat. Rev. Genet. **11**, 31–46 (2010)
16.385 L. Wang, P.C.H. Li: Microfluidic DNA microarray analysis: A review, Anal. Chim. Acta **687**, 12–27 (2011)
16.386 F. Sanger, S. Nicklen, A.R. Coulson: DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. **74**, 5463–5467 (1977)
16.387 C.S. Effenhauser, A. Paulus, A. Manz, H.M. Widmer: High-speed separation of antisense oligonucleotides on a micromachined capillary electrophoresis device, Anal. Chem. **66**, 2949–2953 (1994)
16.388 A.T. Woolley, R.A. Mathies: Ultra-high-speed DNA sequencing using capillary electrophoresis chips, Anal. Chem. **67**, 3676–3680 (1995)
16.389 R.G. Blazej, P. Kumaresan, R.A. Mathies: Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing, Proc. Natl. Acad. Sci. **103**, 7240–7245 (2006)
16.390 M. Knapp, J. Baker, A. Chow, A. Kopf-Sill, M. Spaid: Single molecule amplification and detection of DNA length, US Patent (Application) 20050042639 A1 (2004)
16.391 D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, J.M. Boutell, J. Bryant, R.J. Carter, R.K. Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, N.A. Gormley, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S.V. Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, P.A. Baybayan, V.A. Benoit, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, J.A. Bridgham, R.C. Brown, A.A. Brown, D.H. Buermann, A.A. Bundu, J.C. Burrows, N.P. Carter, N. Castillo, E.C.M. Chiara, S. Chang: R. Neil Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K.V. Fuentes Fajardo, W. Scott Furey, D. George, K.J. Gietzen, C.P. Goddard, G.S. Golda, P.A. Granieri, D.E. Green, D.L. Gustafson, N.F. Hansen, K. Harnish,

C.D. Haudenschild, N.I. Heyer, M.M. Hims, J.T. Ho, A.M. Horgan, K. Hoschler, S. Hurwitz, D.V. Ivanov, M.Q. Johnson, T. James, T.A. Huw Jones, G.D. Kang, T.H. Kerelska, A.D. Kersey, I. Khrebtukova, A.P. Kindwall, Z. Kingsbury, P.I. Kokko-Gonzales, A. Kumar, M.A. Laurent, C.T. Lawley, S.E. Lee, X. Lee, A.K. Liao, J.A. Loch, M. Lok, S. Luo, R.M. Mammen, J.W. Martin, P.G. McCauley, P. McNitt, P. Mehta, K.W. Moon, J.W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S.M. Novo, M.J. O'Neill, M.A. Osborne, A. Osnowski, O. Ostadan, L.L. Paraschos, L. Pickering, A.C. Pike, A.C. Pike, D. Chris Pinkard, D.P. Pliskin, J. Podhasky, V.J. Quijano, C. Raczy, V.H. Rae, S.R. Rawlings, A. Chiva Rodriguez, P.M. Roe, J. Rogers, M.C. Rogert Bacigalupo, N. Romanov, A. Romieu, R.K. Roth, N.J. Rourke, S.T. Ruediger, E. Rusman, R.M. Sanches-Kuiper, M.R. Schenker, J.M. Seoane, R.J. Shaw, M.K. Shiver, S.W. Short, N.L. Sizto, J.P. Sluis, M.A. Smith, J. Ernest Sohna Sohna, E.J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C.L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S.M. Virk, S. Wakelin, G.C. Walcott, J. Wang, G.J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J.C. Mullikin, M.E. Hurles, N.J. McCooke, J.S. West, F.L. Oaks, P.L. Lundberg, D. Klenerman, R. Durbin, A.J. Smith: Accurate whole human genome sequencing using reversible terminator chemistry, Nature **456**, 53–59 (2008)

16.392 M. Lebl, D.L. Heiner, C. Zhao, D. Barker, L. David: Microfabrication methods for the optimal patterning of substrates, US Patent (Application) 20130096034 A1 (2013)

16.393 N. Rusk: Torrents of sequence, Nat. Methods **8**, 44–44 (2011)

16.394 M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb: Zero-Mode waveguides for single-molecule analysis at high concentrations, Science **299**, 682–686 (2003)

16.395 D. Stoddart, A.J. Heron, E. Mikhailova, G. Maglia, H. Bayley: Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore, Proc. Natl. Acad. Sci. U.S.A. **106**, 7702–7707 (2009)

16.396 E.Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martersteck: Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, Cell **161**, 1202–1214 (2015)

16.397 E.M. Southern, U. Maskos: Support-bound oligonucleotides, US Patent (Application) 5436327 (1995)

16.398 R.J. Lipshutz, S.P. Fodor, T.R. Gingeras, D.J. Lockhart: High density synthetic oligonucleotide arrays, Nature Genet **21**, 20–24 (1999)

16.399 M. Schena, D. Shalon, R.W. Davis, P.O. Brown: Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science **270**, 467 (1995)

16.400 M. Schena: Microarray biochip technology eaton publishing, Pharmacogenomics **1**, 289–307 (2000)

16.401 K. Dill, A. McShea: Recent advances in microarrays, Drug Discov. Today: Technol. **2**, 261–266 (2005)

16.402 D. Gershon: DNA microarrays: More than gene expression, Nature **437**, 1195–1198 (2005)

16.403 R.T. Kelly, A.T. Woolley: Microfluidic systems for integrated, high-throughput DNA analysis, Anal. Chem. **77**, 96A–102A (2005)

16.404 R.H. Liu, K. Dill, H.S. Fuji, A. McShea: Integrated microfluidic biochips for DNA microarray analysis, Expert Rev. Mol. Diagn. **6**, 253–261 (2006)

16.405 MarketsandMarkets: Gene expression analysis market by applications, technology, consumables and services-2020, http://www.marketsandmarkets.com/Market-Reports/gene-expression-analysis-market-156613968.html (2016)

16.406 D. Huh, B.D. Matthews, A. Mammoto, M. Montoya-Zavala, H.Y. Hsin, D.E. Ingber: Reconstituting organ-level lung functions on a chip, Science **328**, 1662–1668 (2010)

16.407 J.W. Song, S.P. Cavnar, A.C. Walker, K.E. Luker, M. Gupta, Y.C. Tung, G.D. Luker, S. Takayama: Microfluidic endothelium for studying the intravascular adhesion of metastatic breast cancer cells, PLoS One **4**, e5756 (2009)

16.408 A.R. Aref, R.Y. Huang, W. Yu, K.N. Chua, W. Sun, T.Y. Tu, J. Bai, W.J. Sim, I.K. Zervantonakis, J.P. Thiery, R.D. Kamm: Screening therapeutic EMT blocking agents in a three-dimensional microenvironment, Integr. Biol. (Camb.) **5**, 381–389 (2013)

16.409 H.J. Kim, D. Huh, G. Hamilton, D.E. Ingber: Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow, Lab Chip **12**, 2165–2174 (2012)

16.410 Y. Kim, M.E. Lobatto, T. Kawahara, B.L. Chung, A.J. Mieszawska, B.L. Sanchez-Gaytan, F. Fay, M.L. Senders, C. Calcagno, J. Becraft, M. Tun Saung, R.E. Gordon, E.S. Stroes, M. Ma, O.C. Farokhzad, Z.A. Fayad, W.J. Mulder, R. Langer: Probing nanoparticle translocation across the permeable endothelium in experimental atherosclerosis, Proc. Natl. Acad. Sci. A **111**, 1078–1083 (2014)

16.411 S.A. Sundberg: High-throughput and ultra-high-throughput screening: Solution-and cell-based approaches, Curr. Opin. Biotechnol. **11**, 47–53 (2000)

16.412 J. Hong, J.B. Edel: Micro- and nanofluidic systems for high-throughput biological screening, Drug Discov. Today **14**, 134–146 (2009)

16.413 R. Ghaemi, P.R. Selvaganapathy: Microfluidic devices for automation of assays on drosophila melanogaster for applications in drug discovery and biological studies, Curr. Pharm. Biotechnol. **17**, 822–836 (2016)

16.414 I. Barbulovic-Nad, A.R. Wheeler: Cell assays in microfluidics. In: *Encyclopedia of Microfluidics and Nanofluidics*, ed. by D. Li (Springer, New York 2008) pp. 209–216

16.415 V.I. Chin, P. Taupin, S. Sanga, J. Scheel, F.H. Gage, S.N. Bhatia: Microfabricated platform for studying stem cell fates, Biotechnol. Bioeng. **88**, 399–415 (2004)
16.416 E. Delamarche, D. Juncker, H. Schmid: Microfluidics for processing surfaces and miniaturizing biological assays, Adv. Mater. **17**, 2911–2933 (2005)
16.417 R. Ghaemi, P. Rezai, B.G. Iyengar, P.R. Selvaganapathy: Microfluidic devices for imaging neurological response of Drosophila melanogaster larva to auditory stimulus, Lab on a Chip **15**, 1116–1122 (2015)
16.418 P. Rezai, S. Salam, P. Selvaganapathy, B.P. Gupta: Microfluidic systems to study the biology of human diseases and identify potential therapeutic targets in C. elegans. In: *Integrated Microsystems*, ed. by K. Iniewski (CRC Press, Boca Raton 2011) pp. 581–608
16.419 P. Rezai, S. Salam, P.R. Selvaganapathy, B.P. Gupta: Electrical sorting of caenorhabditis elegans, Lab Chip **12**, 1831–1840 (2012)
16.420 C. Mandrycky, Z. Wang, K. Kim, D.H. Kim: 3-D bioprinting for engineering complex tissues, Biotechnol. Adv. **34**, 422–434 (2016)
16.421 M. Safdar, J. Janis, S. Sanchez: Microfluidic fuel cells for energy generation, Lab Chip **16**, 2754–2758 (2016)
16.422 H. Lee, S. Choi: A micro-sized bio-solar cell for self-sustaining power generation, Lab Chip **15**, 391–398 (2015)
16.423 L. Li, G. Wang, R. Chen, X. Zhu, H. Wang, Q. Liao, Y. Yu: Optofluidics based micro-photocatalytic fuel cell for efficient wastewater treatment and electricity generation, Lab Chip **14**, 3368–3375 (2014)
16.424 J.P. Esquivel, M. Castellarnau, T. Senn, B. Lochel, J. Samitier, N. Sabate: Fuel cell-powered microfluidic platform for lab-on-a-chip applications, Lab Chip **12**, 74–79 (2012)

17. Microfluidic Micro/Nano Droplets

Gopakumar Kamalakshakurup, Derek Vallejo, Abraham Lee

Microfluidic droplet technology has evolved rapidly since the first microfluidic droplet generator was reported over a decade ago. It has subsequently branched out and emerged as a practical solution to enhance the capabilities of many other fields, including, but not limited to: high-throughput screening, biosensing, drug delivery and synthetic biology. In this chapter, we will report on recent advancements in droplet microfluidic technologies that have emerged since *Teh* et al.'s comprehensive 2007 review. We begin with a brief history of droplet microfluidics and introduce methods of droplet production, manipulation, and sensing methodologies. The remainder of the chapter is dedicated to design considerations for various droplet production configurations, concluding with a discussion on applications, trends and the general direction that the field is headed.

17.1 **Introduction to Micro/Nano Droplet Microfluidic Technologies** 537
17.1.1 Droplet Generation 538
17.1.2 Droplet Merging 541
17.2 **Overview of Current Trends in Droplet Microfluidic Technologies** 546
17.3 **Fundamental Designs and Techniques for Microfluidic Generation of Droplets** 547
17.3.1 Passive Droplet Generation Techniques 548
17.4 **Microfluidic Micro-/Nanodroplet Applications** 550
17.4.1 Chemical Reactions in Droplets 550
17.4.2 Biomolecule Synthesis 552
17.4.3 Drug Discovery 553
17.5 **Conclusion** 556
References 556

17.1 Introduction to Micro/Nano Droplet Microfluidic Technologies

High-throughput screening (HTS) is a scientific method to conduct millions of chemical and biological reactions. A typical example of HTS is the discovery of drugs in the pharmaceutical industry, where millions of drug compounds are tested against the target samples to analyze the activity. Since early 1960s, screening has been performed using microplate technology and it is still a convenient platform used by many HTS industries. If we take an example of drug screening in pharmaceutical industry, the process flow occurs in three steps as shown in Fig. 17.1:

1. Biological target samples are pipetted into the well plates
2. Drug compounds are added to the wells
3. The fluorescent activity is captured using a detector that scans the wells.

The limitations of this technology includes:

1. Reagent and sample consumption. Each well consumes milliliter volumes of reagents, which increases the cost of screening. To reduce the reagent consumption, attempts have made to incorporate a higher number of miniaturized wells (Moore's law of miniaturization). In general, 96 well plates are considered standard, but companies like DuPont have come up with 9600 well plates. This leads to another limitation of liquid handling.
2. Using 96 well plates, one can use pipettes to manually load the samples, but 9600 well plates require robotic pipetting, which again increases the cost of screening.
3. Evaporation and contamination. Reagents/samples can evaporate from the wells and can also cause contamination issues.
4. Low throughput. To perform millions of reactions, the microplate platform takes several days. These limitations forced the scientific community to come up with an alternate solution to microplate technology.

Advancements in microfabrication technology revolutionized the process of sensing and detection of chemical and biological samples [17.1]. The samples

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_17

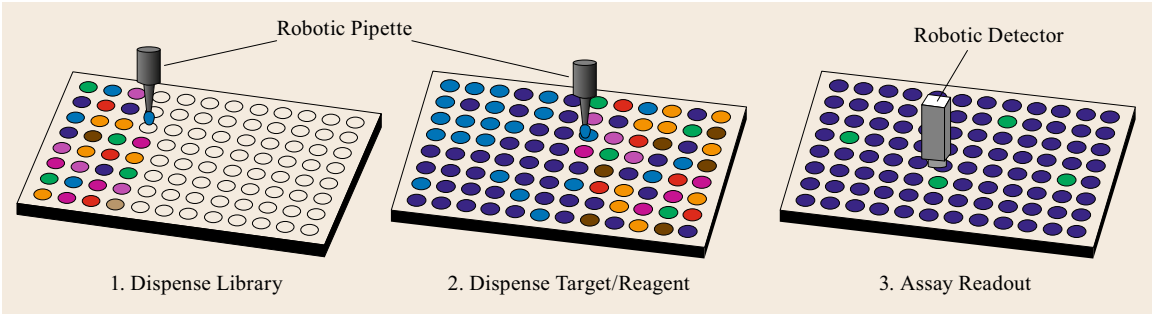


Fig. 17.1 Drug screening process in microplate format: 1. Biological sample is pipetted into the wells 2. The reagent or drug compound is dispensed into the wells. 3. The aftereffect of the reaction is detected using a robotic scanner

are manipulated and analyzed in miniaturized microfluidic chips often referred to as *lab on chip* devices. Depending on the liquid manipulation methodology, we can classify microfluidics into two groups: single-phase (or continuous) microfluidics and multiphase (or droplet) microfluidics. Single-phase microfluidics deals with manipulation of miscible liquids inside a microfluidic chip. Using single-phase technology, numerous chemical or biological reactions can be performed by integrating thousands of micromechanical valves and control components [17.2] (Fig. 17.2). These systems are referred to as microfluidic large scale integration [17.3]. However, single-phase microfluidics has limitations:

1. Samples directly interact with the channel walls and cause cross-contamination of fluid constituents
2. Scaling-up of the screening process require additional chambers and valves, leading to an increase in overall chip size
3. Reagent consumption, though microliters (μL) in volume, could become significant in large-scale screening.

Attempts to overcome the limitations of single-phase microfluidics led to the evolution of the second category of microfluidics referred to as multiphase microfluidics. The most popular form of multiphase microfluidics is droplet microfluidics. Droplet microfluidics deals with the manipulation of nanoliter (nl) to femtoliter (fl) droplets inside microfluidic channels [17.4]. Each droplet can act as an isolated reaction container for biochemical assays while offering several advantages over conventional single-phase methods:

1. Droplets can have volumes as little as femtoliters, leading to a reduction in reagent consumption and therefore the cost of biological screening.
2. The thin layer of carrier phase around the droplet prevents evaporation or contamination; droplets can be generated and detected with ultrafast assay throughputs (rates up to 10 KHz).
3. In droplet format, we can perform millions of chemical or biological reactions in minutes. One of the most important advantages is that droplet microfluidic technology can perform operations that can emulate laboratory benchtop processes, such as fluid dispensing, mixing, aliquot and sorting/separation.
4. Fluid dispensing is analogous to droplet generation; mixing, aliquot and separation is equivalent to droplet merging, droplet splitting and droplet sorting, respectively.

These unit operations make the droplet technology feasible for HTS of chemical or biological assays. Here we describe and discuss the droplet unit operations in detail.

17.1.1 Droplet Generation

One critical aspect of droplet microfluidics is the ability to generate stable, uniformly sized droplets. Precise control over the parameters, including channel geometry, the flowrate of two phases (shear force), the viscosity of the phases and the interfacial tension between the phases, are critical to achieve reliable droplet production. The two basic (and most commonly adopted) device configurations to generate droplets are T-junction and flow-focusing geometries.

T-Junction Configuration

In this configuration, the dispersed phase and the continuous stream intersect perpendicular to each other [17.5]. The continuous phase exerts unidirectional shear on the dispersed phase (Fig. 17.3). As the dispersed phase gradually fills the main channel, it im-

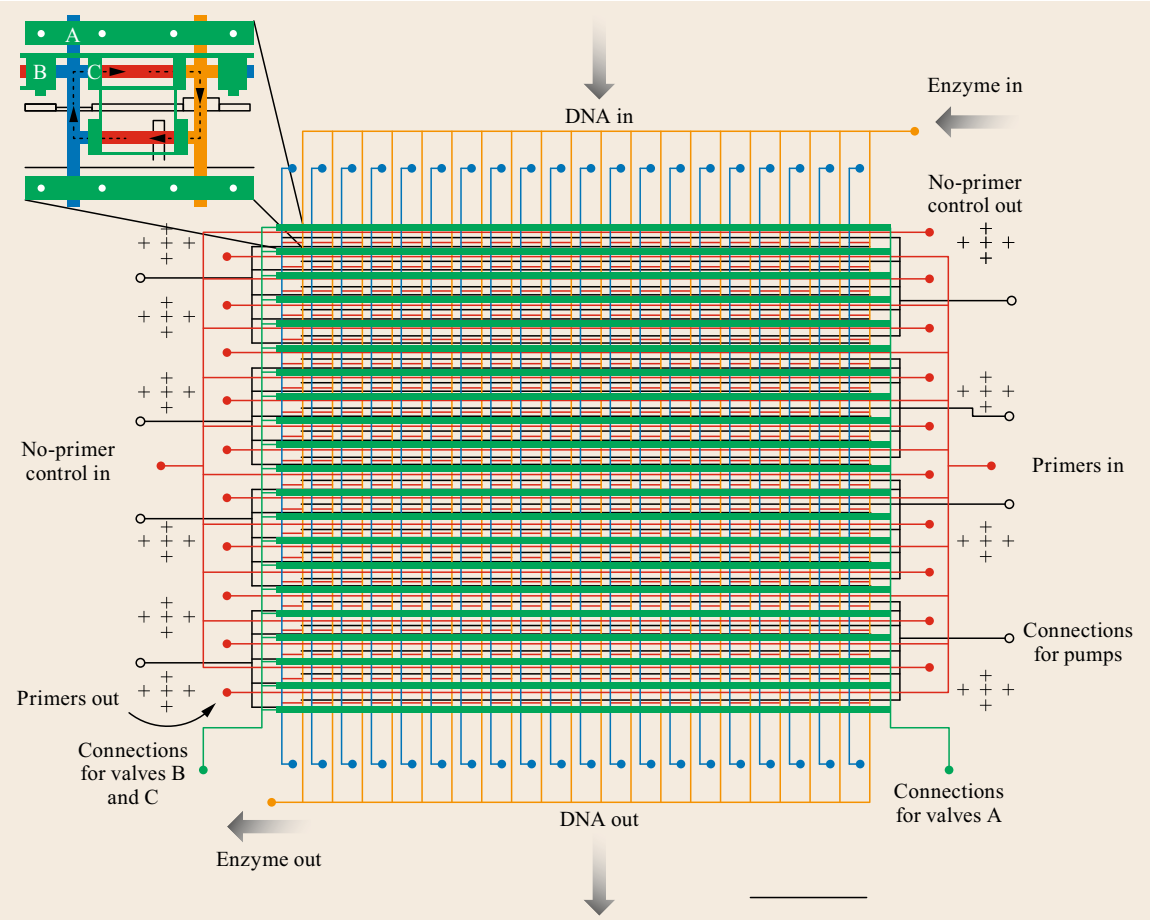


Fig. 17.2 Schematic of a microfluidic large-scale integration device. It consists of 25×40 chambers and each chamber can hold 250 picoliters of sample volume (after [17.3])

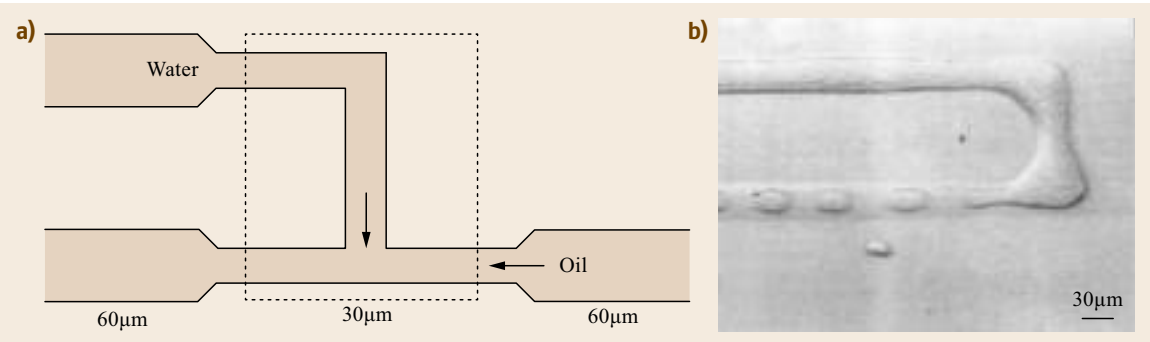


Fig. 17.3 **(a)** Schematic of first T-junction geometry reported in 2001. T-junction microfluidic droplet generator. **(b)** Reprinted with permission from [17.5], copyright 2001 by the American Physical Society

pedes the continuous phase flow, resulting in a pressure buildup in the main channel and leading to the droplet pinch-off. To perform more sophisticated assays, alternate droplet generators are implemented with multiple T-junctions enabling simultaneous introduction of multiple reagents and sample solutions (Fig. 17.4) [17.6].

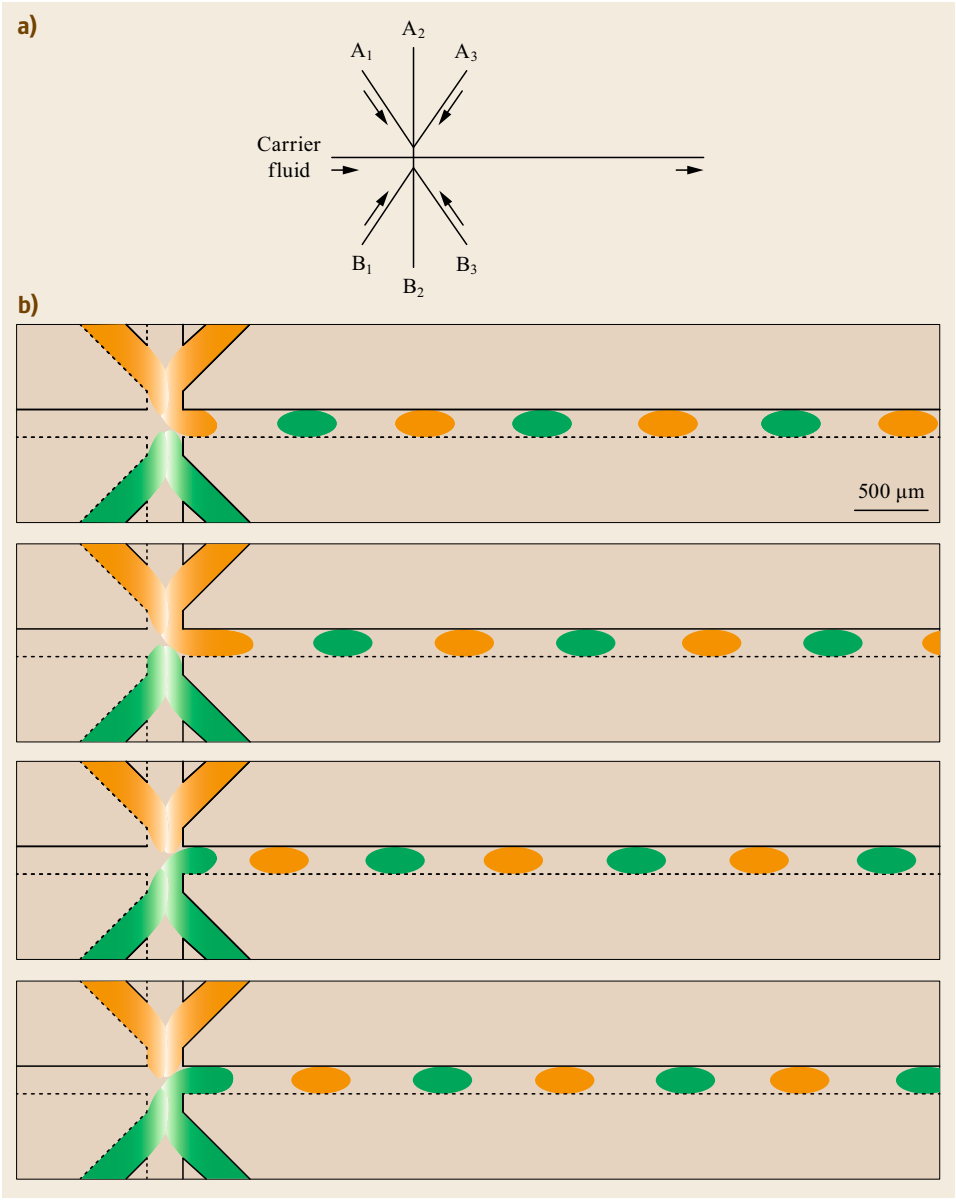


Fig. 17.4a,b Alternating T-junction droplet generator. **(a)** Schematic of the microfluidic chip **(b)** Sequential images illustrating the detailed process of alternate droplet generation

Flow-Focusing Configuration

In this configuration, both the dispersed and continuous phases are forced to pass through a narrow orifice in the microfluidic device in a sheath flow fashion (dispersed phase sandwiched by two continuous phase streams) (Fig. 17.5 and 17.6) [17.7]. The flow-focusing droplet generation can be explained by a simplified model. When the dispersed phase enters the junction, it blocks the flow of continuous phase resulting in a pressure buildup in the continuous phase inlets. The subsequent shear, as a result of the pressure buildup, pinches off the dispersed phase to form droplets. This process of droplet breakup is valid only for geometry-controlled regime, and not applicable for all droplet generation regimes. The physical principles of droplet generation and techniques are described in Sect. 17.3. Recently, it was reported that 10–50 fl droplets can be generated at frequencies ranging from 200 to 1.3 MHz, enabling droplet technology as an ultrahigh-throughput screening platform [17.8]. Parallel droplet generation is an alternative method to increase the throughput by interconnecting networks of individual droplet generators. This can achieve ultrahigh throughput albeit the complexity of the system increases with multiplexing.

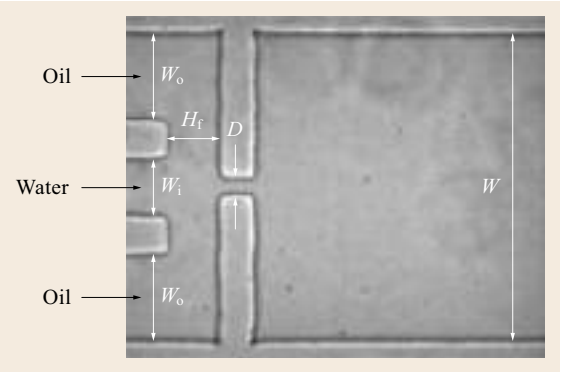


Fig. 17.5 Microfluidic flow-focusing droplet generator device. Reprinted from [17.7] with the permission of AIP Publishing

An example of parallelization is a three-dimensional monolithic elastomer device with an array of 20×50 flow-focusing generators capable of generating droplets at 8.7 MHz, the highest droplet generation rate reported to date (Fig. 17.7) [17.9]. In most cases, the microfluidic devices are manufactured using polydimethyl siloxane (PDMS), which is an inexpensive, flexible and transparent elastomeric polymer [17.10].

17.1.2 Droplet Merging

Automated droplet fusion is also a critical unit operation, where the independently generated droplets of reagents and samples are subjected to controlled merging and the activity of the merged droplets is monitored using droplet detection technologies.

Passive Merging

Passive techniques utilize microfluidic channel designs to ensure controlled merging. There are several ways to perform passive droplet fusion. However, the two eminent techniques are based on channel expansion [17.11] and pillar designs [17.12]. When a droplet encounters a channel expansion, the decrease in velocity slows down the droplet ensuring merging with the preceding droplet (Fig. 17.8). In the pillar-induced technique, the continuous phase near the droplet drains through the gap between the pillars. This slows down the droplet and eventually enables merging with the successive droplets until the hydrostatic pressure exceeds the capillary pressure, pushing the merged droplet out of the chamber (Fig. 17.9). The process of droplet merging depends on the pillar geometry, interfacial tension, flowrate ratio and viscosity of the continuous and dispersed streams.

Active Merging

Active methods employ on-chip structures and control electronics to achieve droplet merging. It includes electrical, thermal and optical methods. Electrical-based methods include dielectrophoresis (DEP) [17.13] and electrowetting-on-dielectric (EWOD) [17.14] force fields where the actuation of electrodes induces the merging. Droplets are stacked up one by one over a series of electrodes and the controlled actuation of a pair of adjacent electrodes coalesce the packed droplets (Fig. 17.10). It is also possible to actuate the adjacent electrodes to move the droplets to the desired location and fuse them. Digital microfluidics is a field that controls droplet transport by the EWOD principle. Droplet merging can also be initiated by temperature. Temperature changes correlate to changes in the viscosity and drainage rate of the continuous phase resulting in droplet fusion [17.15]. Optical droplet merging techniques rely on the principle of optical tweezers but it is less attractive because of the lower throughput and lower actuation force [17.16].

Droplet Splitting

Droplet splitting is a process of breaking the parent droplet into two daughter droplets. It is an important

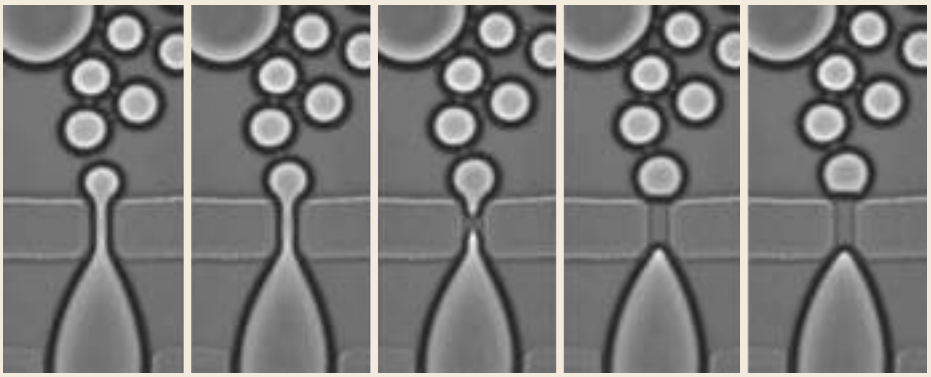


Fig. 17.6a,b Droplet generation in flow-focusing geometry. Reprinted from [17.7] with the permission of AIP Publishing

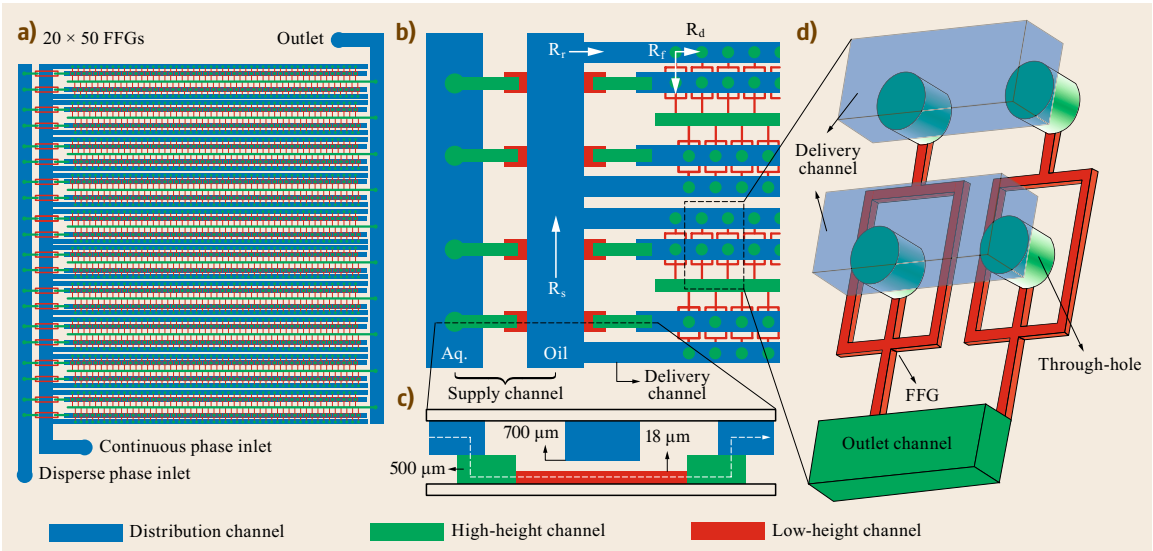


Fig. 17.7 Schematic of a 3D monolithic elastomer device (MED). **(a)** Design consists of 20 rows by 50 columns of flow focusing generators. **(b)** Ladder geometry and fluidic resistance modeling to ensure uniform distribution of solution in each flow focusing geometry. **(c)** Cross-sectional view of the device. **(d)** Fluid from the delivery channel enter the flow focusing geometry through the intermediate through-hole. Using this device, the droplet production rate can go up to 8.7 MHz



Fig. 17.8 Channel expansion design for droplet merging (after [17.11])

operation unit that increases throughput by parallelization. Droplet splitting is achieved using both passive and active methods. Passive techniques split the parent drop into two daughter droplets by shear force induced at the channel bifurcation point [17.11, 17, 18]. Given a constant flowrate, the size of the daughter droplet is determined by the hydraulic resistance of the channels (Fig. 17.11). Bifurcating channels with the same hydraulic resistances results in same shear forces on either half of the droplets, leading to symmetric splitting of the parent droplet (Fig. 17.12). Unequal channel resistances result in uneven droplet splitting such that the daughter droplet in the low-resistance channel is relatively larger than the one in the high resistance channel. Active splitting primarily uses electrical methods to perform droplet splitting. The EWOD technique can be used for active fission in digital microfluidics. The two electrodes on either end of the droplet are actuated while the center electrode is grounded. The electrodes on front and rear end of the droplet pulls the droplet in opposite directions in reference to the ground electrode, resulting in droplet splitting [17.19].

Assay Readout/Droplet Detection Technique

Sensing and detecting droplets is a critical operation unit in HTS of chemical or biological assays. It identifies the positive droplets from the population prior to droplet sorting for postprocessing. Light-induced fluorescent (LIF) detection via chemical labeling of biomolecules is a widely adopted detection technique [17.20]. In LIF, the excitation light source with a suitable wavelength is focused into the microfluidic chip. The excitation light can be either scattered or absorbed by particles, cells or biomolecules. If any of the molecules are stained by fluorophores, the resulting emission wavelength is captured by a detector, which can be a high-speed camera, confocal microscope, complementary metal-oxide-semiconductor (CMOS) sensors, or photomultiplier tubes (PMT). Recently, a combination of microlens array, reflective channel surface and high-speed camera was used for high-speed detection of droplets. This setup can screen droplets at a rate of 2 kHz per microlens [17.21]. Another detection technique relies on a CMOS sensor that can detect 40 pl droplets at 254 kHz [17.22] in the microfluidic chip. An integrated comprehensive droplet digital detection (IC3D) technique was reported in 2014, which can detect one positive droplet from 40 million droplets with 77% confidence [17.23]. Using

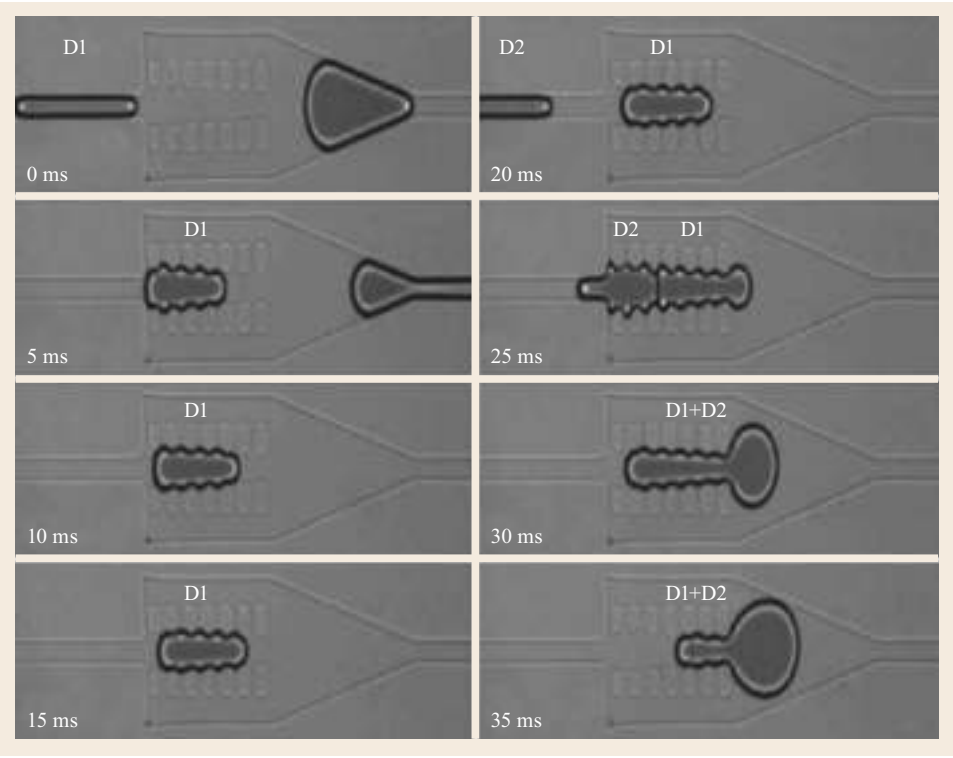


Fig. 17.9 Sequence of images illustrating the merging of two droplets induced by pillar structures. Reproduced from [17.12] with the permission of the Royal Society of Chemistry

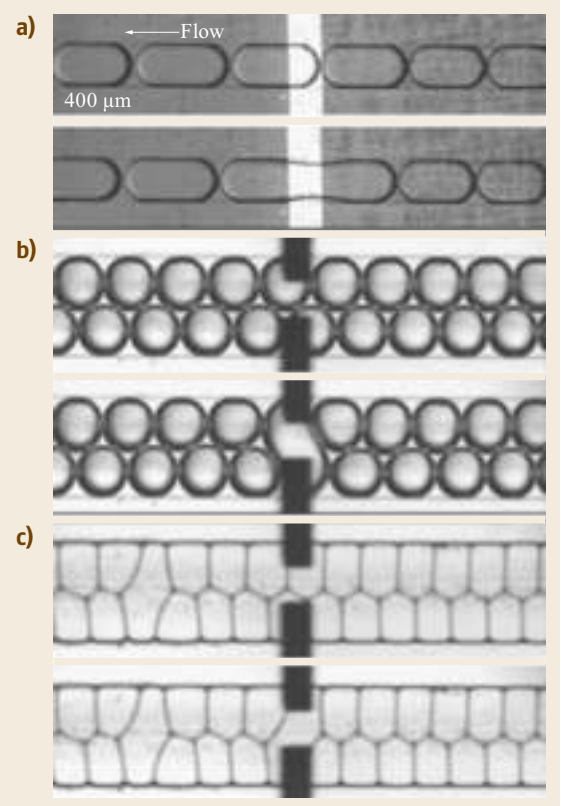


Fig. 17.10a–c Electro-coalescence of the droplets in a microchannel. When a low voltage pulse is applied, pairs of droplets coalesce at the gap between the electrodes. Coalescence of droplets within a **(a)** *bamboo* structure, **(b)** *zigzag* structure and **(c)** targeting of single lamellae in a foamlike zigzag structure. Reprinted from [17.14] with the permission of AIP Publishing ◄

this technique, the droplets collected in a glass cuvette are subjected to vertical and rotational translation and are detected by confocal microscopy (Fig. 17.13). In addition to LIF detection, there are other techniques such as electrochemical detection [17.24], mass spectroscopy [17.25], high-pressure liquid chromatography (HPLC) [17.26], surface enhanced Raman spectroscopy or the combinations of the above-mentioned methods [17.27].

Droplet Sorting

Droplet sorting enables the separation of individual droplets from a population as a function of its size or chemical payload. Droplet sorting can be grouped into two categories: passive and active.

Passive Sorting. As the name suggests, passive sorting techniques do not require any active components such as on-chip electrodes or additional instrumentation. Passive techniques are used to sort the droplets

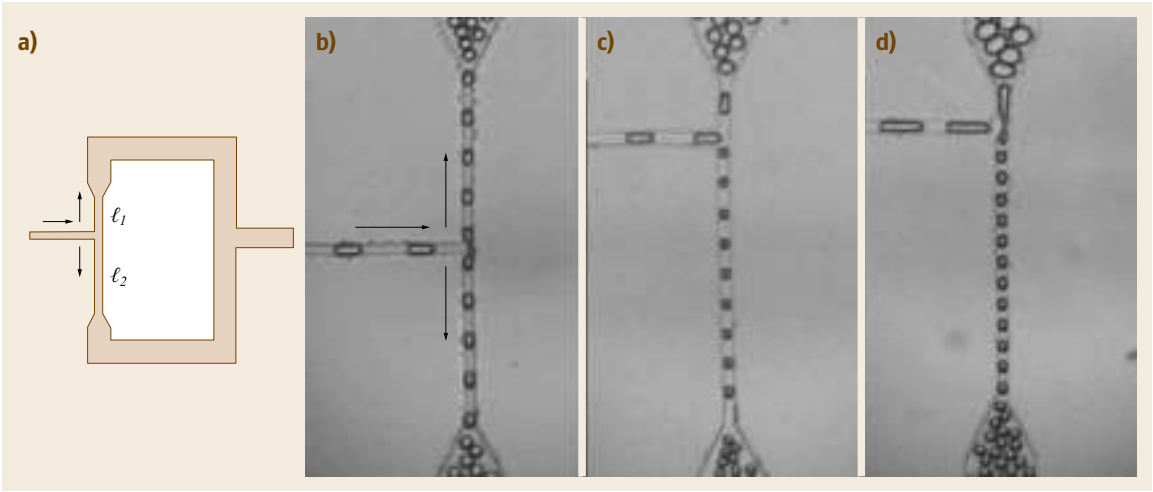


Fig. 17.11a–d Passive droplet splitting using a T-junction geometry. **(a)** Ratio of the two side arms of length ℓ_1 and ℓ_2 determines the size of daughter droplets. **(b)**, **(c)**, and **(d)**; the ratios of the arm lengths are 1:1, 1:5.2 and 1:8.1 respectively

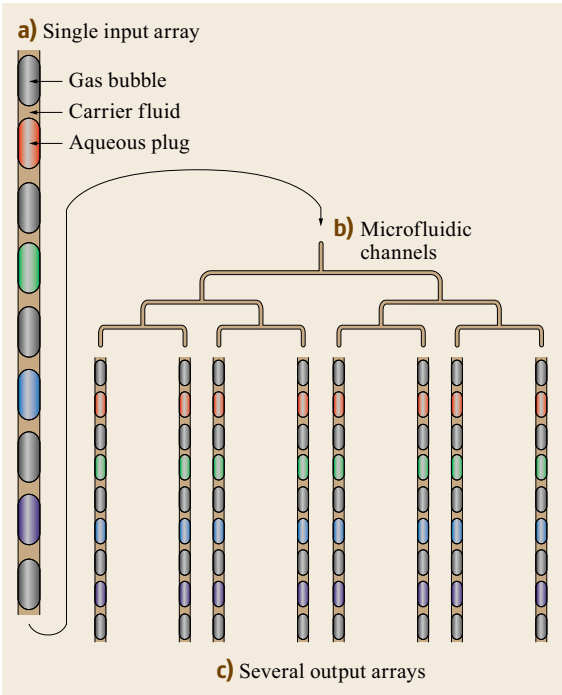


Fig. 17.12a–c Symmetric T-junction droplet splitter design for HTS (after [17.18])

based on their size. In 2007, *Tan* et al. demonstrated a hydrodynamic droplet sorting technique that utilized the channel geometry to alter the flow velocity that resulted in sorting [17.28]. Different sized droplets were successfully sorted into the respective outlets determined by the shear force ratio, which depends on the area of projection and shear rate imposed across the droplet (Fig. 17.14). In 2008, *Chabert* et al. presented a hydrodynamic technique for cell encapsulation and subsequent sorting of the droplets at the channel bifurcation [17.29]. The cell-encapsulated droplets are larger than the empty droplets and thus gets sorted into the *left* outlet by the combined effect of lateral shear induced drift generated by the asymmetric oil flowrates and the steric interactions between the droplets at the bifurcation. In 2011, *Joensson* et al. demonstrated a deterministic lateral displacement (DLD) pillar array that could successfully sort 11 µm droplets from the 30 µm droplets at a frequency of 12 000 droplets/s [17.30]. In 2013, *Hatch* et al. reported a passive hydrodynamic droplet sorting technique by tuning the viscosity ratio (γ), which is defined as the ratio of the viscosity of the dispersed phase to the continuous phase [17.31]. If γ is between 0.5–10, the droplet migrates toward the region of high shear gradient (channel wall). For all other values of γ, the droplet migrates toward the low shear region (channel center).

Active Sorting. Active sorting techniques can sort droplets by their size and chemical contents, but require labeling and on-chip electrodes. In fluorescence-activated droplet sorting (FADS) [17.32] the fluorescent actuation of the on-chip electrodes redirects the droplets at the channel bifurcation via DEP forces [17.33]. This is an efficient technique capable of performing selective sorting of droplets based on their chemical contents (Fig. 17.15). Some of the other active techniques include piezoelectric actuated surface acoustic waves (SAW) [17.34], electrically actuated membrane valves [17.35], which hydrodynamically sort droplets, or laser-induced thermocapillary valves, which can

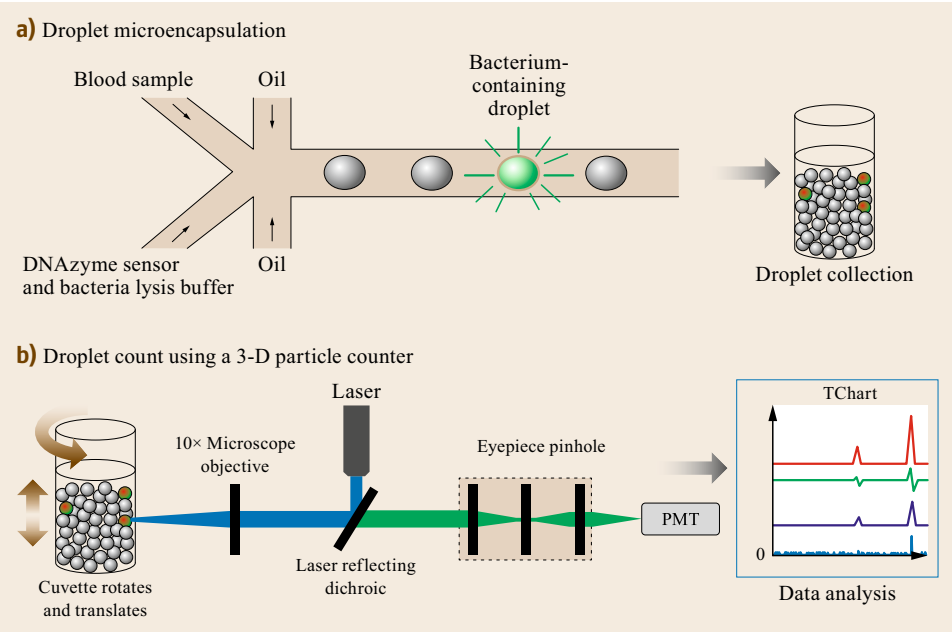


Fig. 17.13a,b Integrated Comprehensive Droplet Digital Detection (IC 3D) technique **(a)** Blood samples, DNAzyme sensors and bacteria lysis buffer is encapsulated in millions of droplets. DNAzyme sensor generates instantaneous signal in the droplets that contains bacterium. **(b)** Droplet are collected and analysed in a high-throughput 3D particle counter (after [17.23])

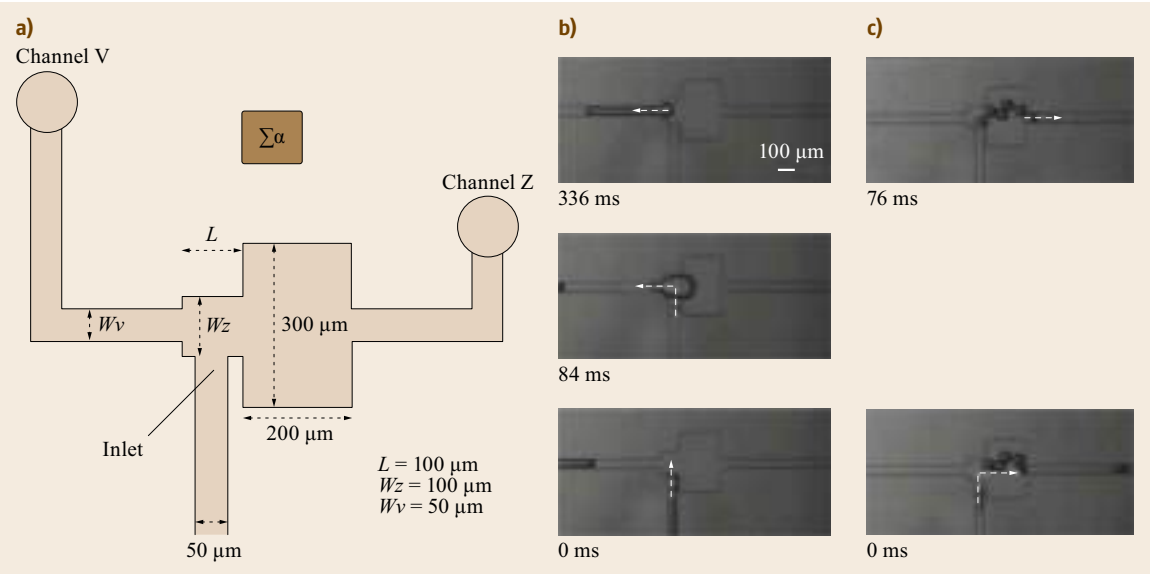


Fig. 17.14a–c Size based droplet sorting in a bifurcating microchannel **(a)** schematic of the device **(b)** large droplets are sorted into the channel V and **(c)** small droplets are sorted into channel Z. Sorting is based on the shear force ratio which depends on the area of projectionand shear rate imposed across the droplet (after [17.28])

block the advancement of droplets [17.36]. For a comprehensive understanding on various droplet unit operations, the reader is directed to an excellent review article [17.37].

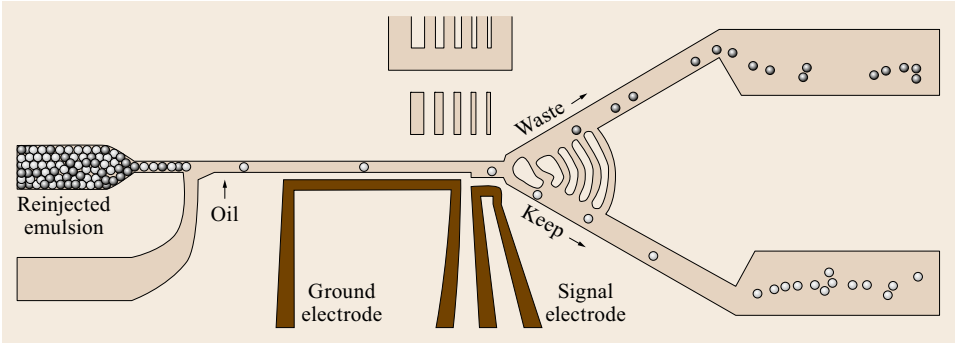


Fig. 17.15 Fluorescence-activated droplet sorting by dielectrophoresis (DEP). Reprinted from [17.20]

17.2 Overview of Current Trends in Droplet Microfluidic Technologies

The field of microfluidics originally emerged as an alternative to traditional, expensive, and bulky laboratory techniques. Costly reagents could be used in even smaller quantities, meaning that reaction time were shortened, translating to comparatively high resolution and sensitive detection. Microfluidic unit operations can emulate most basic laboratory operations, such as fluid dispensing, sample aliquot, sample mixing, and assay readout, at a fraction of the cost and time required with traditional methods. This enabled droplet microfluidic technology to rapidly emerge as a powerful platform to perform numerous applications in chemistry and biology. Academic interest in microfluidic technology first grew tremendously in the 1990s when the U.S. Department of Defense saw an opportunity for microfluidic systems to serve as portable detection units for biological and chemical threats, and in turn, developed a series of programs to stimulate microfluidic research in academic institutions.

The first microfluidic droplet generator was reported in 2001 by *Thorsen* et al. [17.5]. Since then, the technological developments in chemistry, biology and medicine have garnered substantial interest in applications enabled by droplet microfluidics. Further advancements in the microfabrication technologies using various substrate materials (e.g., glass, polymer) has led to the realization of many microfluidic droplet generator designs for control of droplet size, generation rate, and chemical composition [17.38]. The droplet based biomolecular assay was the first commercial success story of droplet technology. Bio-Rad Laboratories and Rain Dance Technologies successfully commercialized the droplet digital polymerase chain reaction (ddPCR) system in 2012 [17.39]. However, the droplet generation, thermocycling, and droplet readout are performed in three different instruments and the assay kit is still very expensive (> $ 75 000). Several startups including Intuitive Biosciences (Madison, WI) are working on ways to overcome the limitations and reduce costs. A droplet-based, fully integrated next-generation sequencing platform by GnuBIO Technologies is currently in the development stage [17.40]. The manufacturer claims that this instrument can perform scalable and accurate DNA sequencing that includes the additional steps of target enrichment and DNA amplification. If successful, this will be the first example of a commercial droplet-sequencing platform.

Another highly pursued and top-priority application in the field of droplet microfluidics is single-cell analysis [17.41]. Since the droplet microfluidic technology can easily generate cell-sized aqueous compartments, large populations of cells can be compartmentalized into independent monodisperse droplets. This enables the researchers to query the heterogeneity in cell populations at the single-cell level by performing sequencing or gene expression profiling in droplets, or to screen prospective drug variants against the target sample for cancer therapeutics [17.42, 43]. Droplet microfluidics has also emerged as a powerful tool for applications in directed evolution and synthetic biology. To engineer proteins and enzymes with novel activity, millions of genetic variants can be tested in individual microenvironments, evaluated, and isolated with high-throughput optical sorting methods [17.44].

Even with the tremendous progress of droplet microfluidics in recent years, there are still many technical challenges that researchers are working on to advance the field. One of the primary challenges pertains to high-throughput, sensitive detection and sorting of droplets. At present, there exists many efficient detection technologies, such as fluorescence detection, surface enhanced Raman spectroscopy, electrochemical detection, mass spectroscopy, HPLC and capillary elec-

trophoresis, but how well can these techniques work for high- throughput sensing and detection? Can we extract complete information from the picoliter volume reactors? As the miniaturization effort is still ongoing, droplet microfluidics will require further future dimensional scale-down in terms of the device size, droplet size and reagent volume, which would seriously challenge existing detection systems that have been designed for cell-sized structures.

Another area that requires improvement is the material used in the fabrication of microfluidic devices. In most research applications, polydimethylsiloxane (PDMS) is used to make the microfluidic chips. PDMS is transparent, flexible, biocompatible and permeable to gases; features that are suitable for culturing biomolecules and cells, but PDMS is also auto-fluorescent. If the signal-to-noise ratio is too low, fluorescent measurements can be compromised. PDMS also has low thermal conductivity, absorbs certain dyes, and swells in certain organic solvents (e.g., silicone oil). PDMS microfluidic devices are also typically polymerized in photolithographically patterned molds, which make them difficult to bulk manufacture and commercialize. An alternative to PDMS are injection-molded thermoplastic materials such as cyclic olefin copolymer (COC) chips. The comparatively low contact angle (82°) of COC, however, makes it difficult to generate water-in-oil droplets. The research is ongoing to improve the contact angle by appropriate surface treatments.

Another challenge for the future of droplet microfluidics is in the synthesis of ideal surfactants to perform assays within microfluidic channels. Although there are many commercially available organic solvents, proper incubation of the droplet emulsions requires standardized surfactants that limit the choice of organic solvents. A major problem encountered while incubating droplets is imperfect partitioning, leading to inter-droplet cross talk, which limits the stability of the emulsions, causing droplet shrinkage or coalescence. Since this is a relatively new and rapidly developing field, more work is ongoing to resolve these issues and challenges. As the field advances, one could foresee the development of integrated droplet microfluidic platforms that are portable, low-cost and autonomous. These platforms would be able to perform many complex biological processes in shorter times, yet with high efficiency and precision.

17.3 Fundamental Designs and Techniques for Microfluidic Generation of Droplets

Emulsions are liquid droplets suspended in another immiscible liquid. There are numerous ways to generate emulsions, such as atomization, chaotic mixing of two immiscible liquids, sprays and membrane emulsification. The consequence of this stochastic generation of emulsions leads to significant droplet size variation. Over the years, microfluidic technology emerged as a powerful tool to generate uniform droplet emulsions of femtoliter to nanoliter volumes at an ultrahigh throughput. As described earlier in the chapter, microfluidic droplet generators can be categorized into active and passive methods. Active droplet generators require additional instrumentation including on-chip devices and control electronics whereas passive methods rely on the interacting flow fields and interfacial instabilities to generate droplets. Passive droplet generation can be grouped into three categories:

1. Coflowing geometry
2. Cross-flowing or T-junction geometry
3. Flow-focusing geometry.

The schematic of all three geometries are shown in Fig. 17.16. In all three cases, the dispersed phase and the continuous phase are introduced separately into the microchannels by a pressure-driven flow that is driven either at constant volumetric flowrates (using syringe pumps) or constant pressures (by pressure regulator and solenoid valves). The two phases meet at a channel junction and eventually the dispersed phase pinches off by free surface instability. The droplet pinch-off process is a result of the competition between the local fluid stress (viscous force), which causes the interface to deform while the capillary pressure (interfacial tension force) resists the deformation. The ratio of viscous stress to the capillary stress is defined by a dimensionless number called the capillary number (Ca)

$$Ca = \mu_c \frac{V}{\sigma} ,$$

where μ_c is the viscosity of the continuous phase, V is the velocity of the continuous phase and σ is the interfacial tension between continuous phase and dispersed phase. Another two dimensionless parameters that directly influence droplet breakup are the volumetric flowrate ratio (φ)

$$\varphi = \frac{Q_d}{Q_c} ,$$

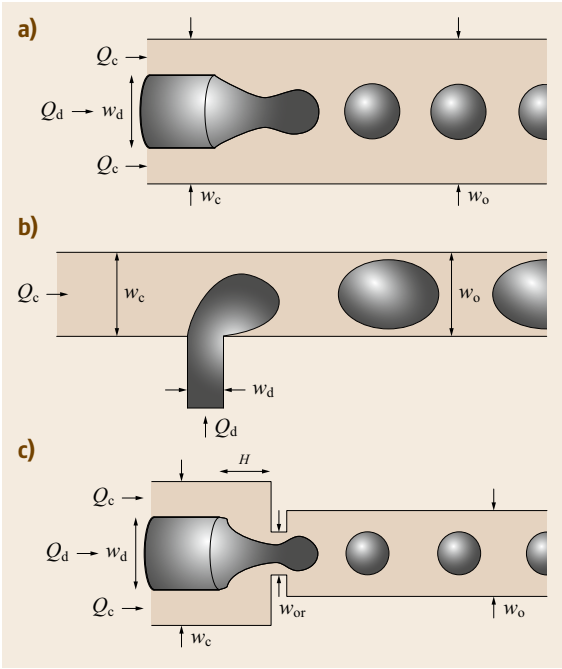


Fig. 17.16a–c Schematic of three major passive droplet generation geometries **(a)** coflowing geometry **(b)** cross-flowing or T-junction geometry **(c)** flow-focusing geometry. (After [17.38])

where Q_d is the volumetric flowrate of the dispersed phase and Q_c is the volumetric flowrate of the continuous phase, and the viscosity ratio (γ)

$$\gamma = \frac{\mu_d}{\mu_c} ,$$

where μ_d and μ_c are the viscosity of the dispersed and continuous phases, respectively.

17.3.1 Passive Droplet Generation Techniques

Coflowing Geometry

This is the simplest geometry to generate uniform droplets. The geometry consists of a set of concentric channels, where the dispersed phase is introduced through the inner concentric channel while the continuous phase flows through the exterior [17.45]. The configuration is illustrated in Fig. 17.1a. Two droplet generation regimes are observed in this configuration, dripping and jetting. The transition from dripping to jetting is observed if the continuous phase increases above a critical velocity [17.46]. The critical velocity is a function of dispersed phase velocity, interfacial tension and viscosity ratio (γ). The critical velocity decreases with increasing dispersed phase velocity and decreasing interfacial tension. Droplet size is a function of the volumetric flowrate (φ) and interfacial tension (σ). An increase in φ results in an increase in droplet size. Higher continuous flowrates exert more shear stress on the dispersed phase interface that result in smaller droplets. Similarly, an increase in σ makes the interface weaker and hence the dispersed phase can easily break off into smaller droplets. It is shown that the viscosity ratio (γ) has no significant effect on the droplet size over a wide range of continuous phase flowrates. In this regime, the predicted droplet size is given by the nondimensional equation [17.45]

$$d_1^3 - \left(1 + \frac{1}{3Ca}\right) d_1^2 - \frac{\varphi}{\alpha} d_1 + \frac{\varphi}{\alpha} = 0 ,$$

where d_1 is the ratio of droplet diameter (d) to the diameter of the inner capillary tube and $\alpha = A_d/A_c$ is the ratio of the two cross-sectional areas of the inner and outer capillary tube. If the dispersed phase flowrate is slow enough, the above equation simplifies to

$$d_1 \approx 1 + \frac{1}{3Ca} .$$

The droplet breakup and the size depends on the viscous stress and the capillary pressures. Using this geometry, highly monodisperse droplets with diameters in the range of few hundred micrometers are produced with polydispersity values ranging from 1 to 2%.

Notably, the wettability of the channel wall is crucial in determining which phase is dispersed. If the channel wall is hydrophobic (e.g., PDMS), the oil phase will wet the wall and form the continuous phase. In that case, the aqueous phase will form the droplets. Conversely, if the channel wall is hydrophilic (e.g., silicon), then the aqueous phase will form the continuous phase while the oil phase will form the droplets.

Cross-Flow or T-Junction Geometry

This is one of the simplest and common microfluidic geometries to generate monodisperse droplets [17.4, 47]. Depending on how the droplet phase fills the junction, the T-Junction geometry can be grouped into two categories, unconfined geometry and confined geometry.

Unconfined Geometry. In the unconfined geometry, the width of the continuous phase channel is at least five times larger than that of dispersed phase channel [17.48]. As the name suggests, the emerging droplets do not interact with the continuous phase channel walls and remain unconfined upon exiting the dispersed phase

inlet. The droplets generated from this geometry do not significantly disrupt the main channel flow, hence the droplet size is determined by the local shear stress at the junction. At a fixed dispersed phase flowrate, the size of the droplets decreases with increasing continuous phase flowrate. There are a few studies that show the droplet pinch-off occurring when the drag force exerted by the continuous phase flow overcomes the interfacial tension resisting the droplet deformation and the dispersed phase flowrate has little influence on the droplet breakup and the size [17.48, 49]. The droplet size in this geometry is given by

$$d_1^4 - 2d_1^3 + \frac{\epsilon}{2}d_1^2 + \frac{\alpha}{2C_\gamma Ca} = 0\,,$$

where $d_1 = d/D_c$ is the dimensionless number, which is the ratio of droplet diameter (d) to the hydraulic diameter (D_c) of the continuous phase channel

$$D_c = \frac{2w_c h}{w_c + h}\,,$$

where w_c and h are the width of the continuous phase channel and height of the channel respectively

$$\alpha = \left(\frac{w_d}{D_c}\right)^2\,,$$

where w_d is the width of the dispersed phase channel,

$$C_\gamma = \frac{\left(3 + \frac{2}{\gamma}\right)}{\left(1 + \frac{1}{\gamma}\right)}$$

is the correction factor for viscosity ratio γ and ϵ is the ratio of droplet velocity to the average continuous flow velocity. The value ranges from $0 < \epsilon < 2$ and it depends on capillary number.

Confined Geometry. In the confined geometry, the width of the dispersed phase channel is comparable to that of the continuous phase [17.50]. The droplet phase completely obstructs the continuous phase channel (except for a thin Bretherton's film between the droplet and the wall), leading to an increase in the upstream pressure driving the breakup of the interface. Under these conditions, plugs are often generated instead of droplets. Plugs can be thought of an elongated droplet or the droplets with length in the flow direction greater than its width. The neck of the plug squeezes at a rate proportional to the average velocity of the continuous phase fluid and the plug fills at a rate proportional to the dispersed phase flowrate [17.51, 52]. The final length of the plug is given by the expression

$$L_1 = \frac{L}{w_c} = 1 + \alpha\varphi\,,$$

where α is the constant that depends on the width of the thinning neck. The equation shows that at low capillary number, the length of the plug depends only on the volumetric flowrate ratio. Above a critical capillary number, the viscosity ratio plays an important role in the droplet breakup, in contrast to the unconfined geometry. In the confined geometry, the droplet formation occurs within a specific range of flow conditions. If φ approaches unity, the dispersed-phase flowrate and the continuous-phase flowrate become equal, and it is observed that two phases flow parallel to each other without breakup.

Flow-Focusing Geometry. In this configuration the dispersed phase and the continuous phase are forced to pass through a small and narrow region in the microfluidic device [17.7]. The shear applied by the continuous phase over the dispersed phase results in droplet pinch-off. Three regimes are observed in this configuration:

1. Squeezing regime or geometrically controlled regime
2. Dripping regime
3. Jetting regime.

The transition between these regimes depends on the capillary numberand the flowrate ratio. In general, microfluidic droplet generation occurs at a capillary number within the range of 10^{-3}–10^{1}. The squeezing regime occurs at low capillary numbers ($\approx 10^{-3}$). When the droplet phase enters the junction, it blocks the flow of the continuous phase and results in pressure buildup in the continuous phase inlets. The subsequent bidirectional shear results in the pressure buildup that pinches off the dispersed phase to form droplets. Notably, the pinch-off occurs right at the orifice. The droplet diameter in this regime is observed to be comparable to the orifice size and is larger than the other two regimes. As the capillary number increases, transition from squeezing regime to dripping regime is observed. In the dripping regime, the droplet breakup occurs within one characteristic diameter of the orifice [17.53, 54]. The interface after the breakup remains at a static location within the orifice. The droplets generated in this regime are smaller than the orifice diameter and are highly monodisperse (polydispersity $< 2\%$). At a fixed orifice size, the droplet diameter

decreases with increasing Ca and decreasing φ. The relationship between the droplet generation frequency (f), droplet spacing (s) and droplet velocity (U) is given by [17.55]

$$U = fs\,.$$

Zhou et al. [17.56] describe the droplet pinch-off mechanism in the dripping regime as a combination of a capillary instability, including Rayleigh capillary instability and end-pinching mechanism, shared with viscous drag exerted by the continuous phase on the droplet phase. The transition from the dripping to jetting regime occurs at a higher capillary number. In the jetting regime, as the name suggests, the dispersed phase protrudes like a long jet and it extends at least three orifice diameters outside the orifice exit. The droplets generated in this regime are polydisperse with a diameter larger than the dripping regime. The transition from the dripping to jetting regime occurs at a critical capillary number given by [17.54]

$$Ca_{\mathrm{cr}} = \frac{t_{\mathrm{pinch}}}{t_{\mathrm{growth}}} = \frac{\mu_{\mathrm{c}} Q_{\mathrm{d}}}{\sigma R_{\mathrm{jet}}^2} \approx 1\,,$$

when $Ca < Ca_{\mathrm{cr}}$. The dripping regime is observed because of the interface instability, which will not allow the growth of a long jet, whereas for $Ca > Ca_{\mathrm{cr}}$, the jetting regime is observed because the viscous stress will suppress the interface instability and allow the growth of a long jet. Droplet generation in coflowing, cross-flowing and flow-focusing geometries are illustrated in Fig. 17.17.

Passive droplet generators are widely used because of its ease of use and simplicity. Active droplets, though more complex, allow for better control over the droplet generation and production (ink jet printing technology). Active droplet generation employs piezoelectric transducers [17.57], membrane valves [17.58], electric fields [17.59] to initiate the droplet breakup.

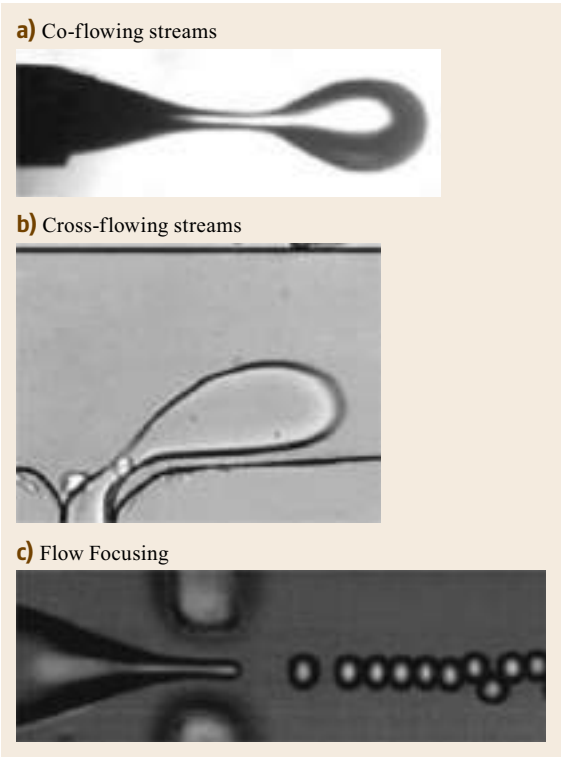


Fig. 17.17a–c Droplet generation in all three geometries. **(a)** Coflowing geometry. **(b)** Cross-junction/T-junction geometry. **(c)** Flow-focusing geometry from [17.38], copyright IOP Publishing. Reproduced with permission all rights reserved

17.4 Microfluidic Micro-/Nanodroplet Applications

Droplet microfluidics offers an excellent platform to conduct a diverse set of applications from chemical synthesis, particle manipulation, bioassays and drug discovery to point-of-care diagnostic chips [17.37, 60–62]. This section discusses the current and potential applications of droplet microfluidics.

17.4.1 Chemical Reactions in Droplets

Droplets act as isolated micro-/nanocontainers to perform chemical reactions. The exquisite control and manipulation of these droplet reactors enables one to conduct innumerable chemical reactions within the droplets [17.4]. Droplet technology offers several advantages while performing the chemical reactions:

1. Reduced consumption of reagents
2. Less exposure to hazardous chemicals
3. Large surface-to-volume ratio enables better diffusion and mass flowrates
4. Recirculation within the droplets ensures proper mixing of the reactants
5. Scale up of the process by parallelization.

The hydrolysis reaction between *p*-nitro phenyl acetate and sodium hydroxide was performed in droplets.

Sodium hydroxide forms the droplet phase and the continuous phase contains *p*-nitro phenyl acetate. The reaction occurred as result of the mass transfer of *p*-nitro phenyl acetate into the droplet phase containing sodium hydroxide. Similar work has been reported to demonstrate the increased yield in droplet-based microreactor platforms [17.65]. The neutralization reaction can also be performed in droplets [17.66]. The acetic acid present in the continuous phase diffuse into the KOH/NaOH droplets and the color variation of the pH indicator is used to monitor the reaction and calculate the diffusion rates. It has been found that the diffusion rate is faster in small droplets than plugs. Other chemical reactions include precipitation reactions [17.67], anticoagulants [17.68], organic phase reactions [17.69] and the generation of cadmium selenide (CdSe) nanoparticles inside the droplets [17.63]. Separate droplets with different reagents were merged in a controlled environment to induce reactions inside the fused droplet to produce monodispersed nanoparticles (Fig. 17.18).

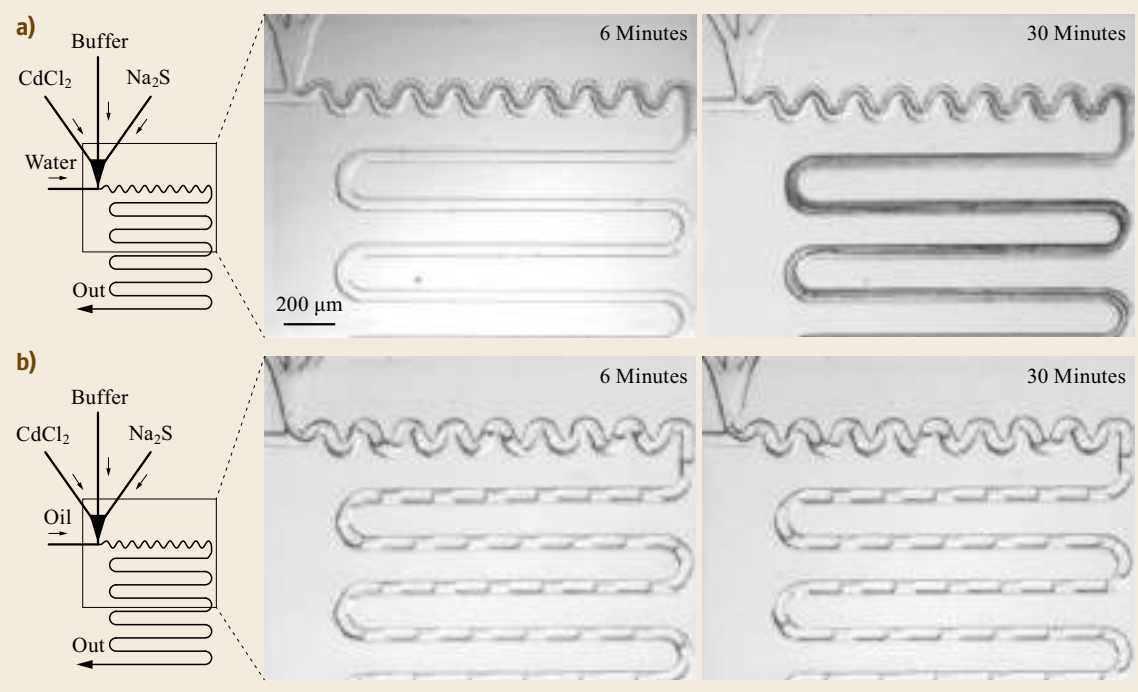


Fig. 17.18a,b Comparison of the nanoparticle (CdS) synthesis in **(a)** single-phase flow and **(b)** droplet phase. Unlike the droplet system, the single-phase system results in nanoparticle accumulation inside the channel and eventually leads to clogging. Reproduced from [17.63] with the permission of the Royal Society of Chemistry

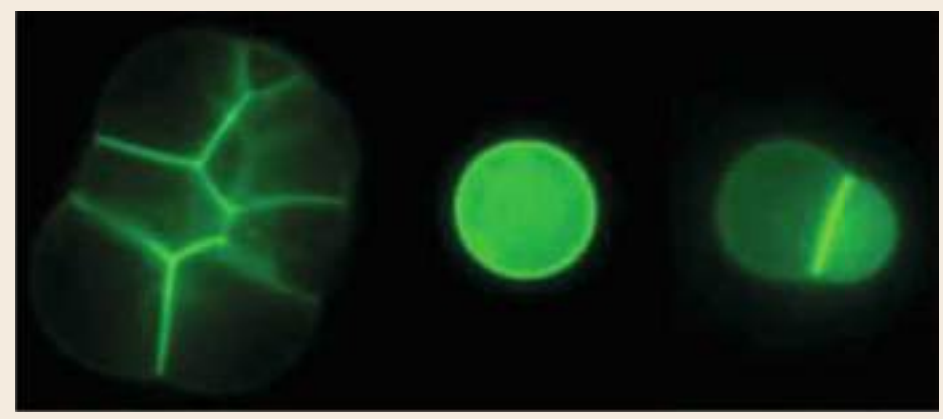


Fig. 17.19 A cell-like bioreactor produced by encapsulating the *E. coli* cell's free expression system in a phospholipid vesicle (after [17.64]), copyright 2004 National Academy of Sciences, USA

17.4.2 Biomolecule Synthesis

Droplet microfluidics is an ideal platform to carry out controlled synthesis of a wide range of biomolecules including high- throughput directed evolution of proteins and polymerization chain reaction (PCR). The ability of droplet microfluidics to generate uniform and homogeneous droplets in a controlled environment is the fundamental requirement for the synthesis of artificial cells [17.71]. The formation of an artificial cell helps one study life's fundamental reactions. A cell-like bioreactor was produced by encapsulating the *E. coli* cell's free expression system in a phospholipid vesicle [17.64]. The phospholipid bilayer acts as a semipermeable membrane that helps in the exchange of nutrients in and out of the cell (Fig. 17.19). *Huebner* et al. demonstrated protein expression by encapsulating single cells inside droplets and detecting the expression of fluorescent proteins within the cell [17.70]. This in-vitro protein expression in droplet vesicles offers potential applications in the high-throughput directed evolution of proteins (Fig. 17.20). In genomics, droplet microfluidics offer a compelling platform to perform PCR in droplets. Compared to the conventional approach, droplet-based PCR technology offers advantages in making the process quantitative and

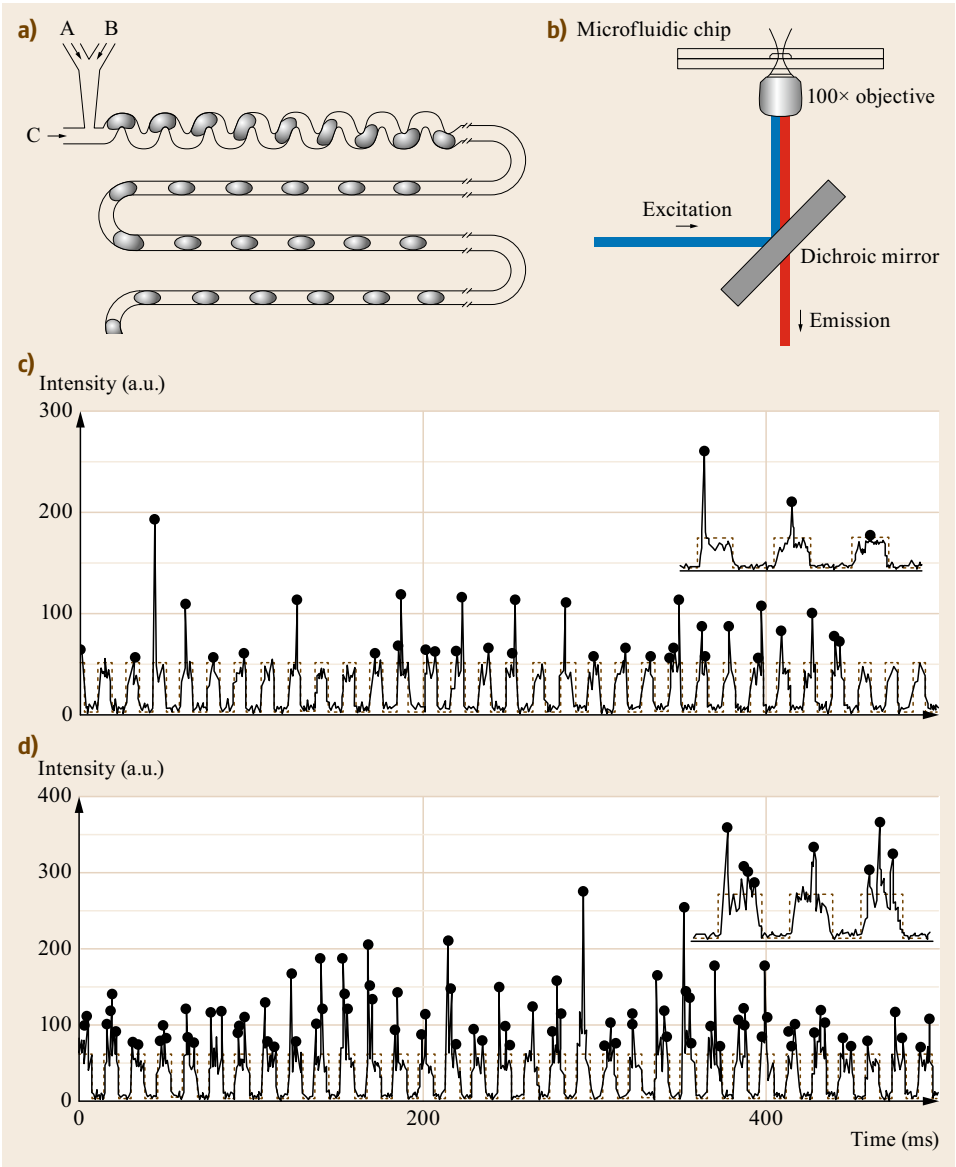


Fig. 17.20a–d Protein expression studies by encapsulating a single cell inside the droplet and detecting the expression of fluorescent proteins within the cell. Droplet-containing cells are distinguished by a vertical spike arising from the expressed fluorescent protein. Reproduced from [17.70] with the permission of the Royal Society of Chemistry

sensitive [17.72]. Droplet-based PCR also increases the amplification efficiency by minimizing the reagent dispersion and surface adsorption. Generally, droplet-based PCR is performed in two different ways. The first is a continuous flow-based approach. A droplet-based micro-oscillation flow PCR chip was designed and implemented by *Wang* et al. [17.73]. Injected droplets containing a PCR mixture, pass through the three distinct temperature zones (denaturation, extension and annealing) in the microfluidic channel to attain the required temperature transition. They successfully amplified human papillomavirus (HPV)-DNA samples and the PCR products were analyzed by gel electrophoresis. The group has also shown that the whole process takes less than 15 min, compared to 2 h for the conventional process. In the second method, microfluidic chips are used only for the generation and detection process [17.74]. This is the method that most commercial products utilize (Fig. 17.21). The process flow occurs in three steps:

1. Uniformly sized droplets with template DNA were created by a microfluidic drop generation device.
2. These droplets are transformed manually into 96-well plates and are placed in a thermocycling device that is programmed to control the temperature required for the DNA amplification.
3. Thermocycled droplets with amplicons are automatically transferred to the detection device, which has a microfluidic chip built into it.

The LIF detection system helps to quantitatively estimate the DNAexpression based on the fluorescence associated with each droplet.

17.4.3 Drug Discovery

In conjunction with high-speed and sensitive-detection technologies, such as LIF detection, mass spectroscopy, capillary electrophoresis, x-ray crystallography and HPLC, droplet microfluidics has the capability to perform HTS of target molecules against potential drugs. For example, to study the activity of a certain drug against the disease causing cells or proteins, both the labeled cells/proteins and drugs are encapsulated separately in the form of droplets using suitable droplet-generation configurations. These two specific droplet groups are subjected to droplet merging/fusion using appropriate passive or active fusion techniques. The product resulting from the reactions inside the fused droplet is picked up by fluorescence-detection techniques. Since the droplet generation frequency can reach 10 KHz, one can screen up to 10 000 droplets in one second, a speed that could revolutionize the high-throughput industry. As a result, the whole screening process has been transformed from conventional well plates to droplet technology. Another tool to perform high-throughput protein analysis is matrix-assisted desorption or ionization (MALDI). MALDI spectroscopy has been integrated with droplet technology to achieve high-throughput analysis of proteins and chemicals [17.75]. *Wheeler* et al. illustrated an EWOD-

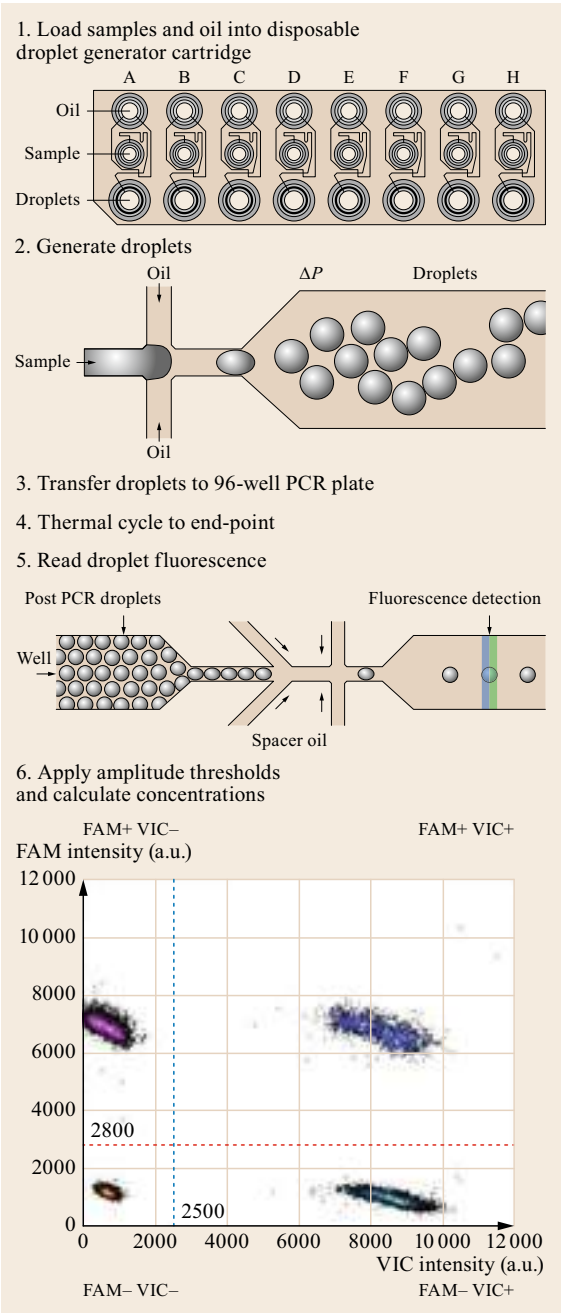


Fig. 17.21 BioRad's droplet PCR workflow

based technique to purify the sample and analyze them using matrix-assisted laser desorption ionization-time of flight mass spectrometer (MALDI-TOF) mass spectroscopy [17.76]. In this platform, the whole process is automated but the common channels still allow for potential surface cross-contamination. Protein characterization can also be done by x-ray diffraction of protein crystals. Proteins (glucose isomerase, thaumatin and ferritin) are subjected to crystallization inside the droplets and were characterized using x-ray crystallography [17.77]. Another technique to characterize a drug compound is to perform kinetic studies to explore the activity of enzymes and chemicals. *Song* and *Ismagilov* have conducted kinetic measurements of ribonuclease A (RNase A) with millisecond temporal resolution [17.78]. The enzymatic activity was determined by the fluorescence resulting from the cleavage of a fluorogenic substrate by RNase A (Fig. 17.22). The intensity of the fluorescence is used to figure out the duration of the reaction process. In general, droplet microfluidics offers a standardized platform to perform HTS of biomolecules. As such, the droplet microfluidics does not alter the format of the drug screening but it adds functionality and performance to the existing platform.

Directed evolution is another important application of droplet-based technology. Copies of genes or cells are distributed into each droplet compartment and are assayed for a specific phenotype. The positive droplets (the droplets containing the mutant genes) were sorted out and broken to recover the genotype for characterization [17.79]. Reverse-transcription PCR (RT-PCR) has emerged as a powerful technique to characterize the transcriptome of single cells in droplets [17.80]. This study will provide insights into the heterogeneity of the cells, their response to drugs or pathogens and the cause of various diseases. The single-cell transcriptomes are captured using the hydrogel microbeads, which are encapsulated in the droplets along with the cell. Entire transcriptomes of the cells were barcoded for identification purposes and this technique has reported a 13% mRNA capture efficiency (Fig. 17.23) [17.81].

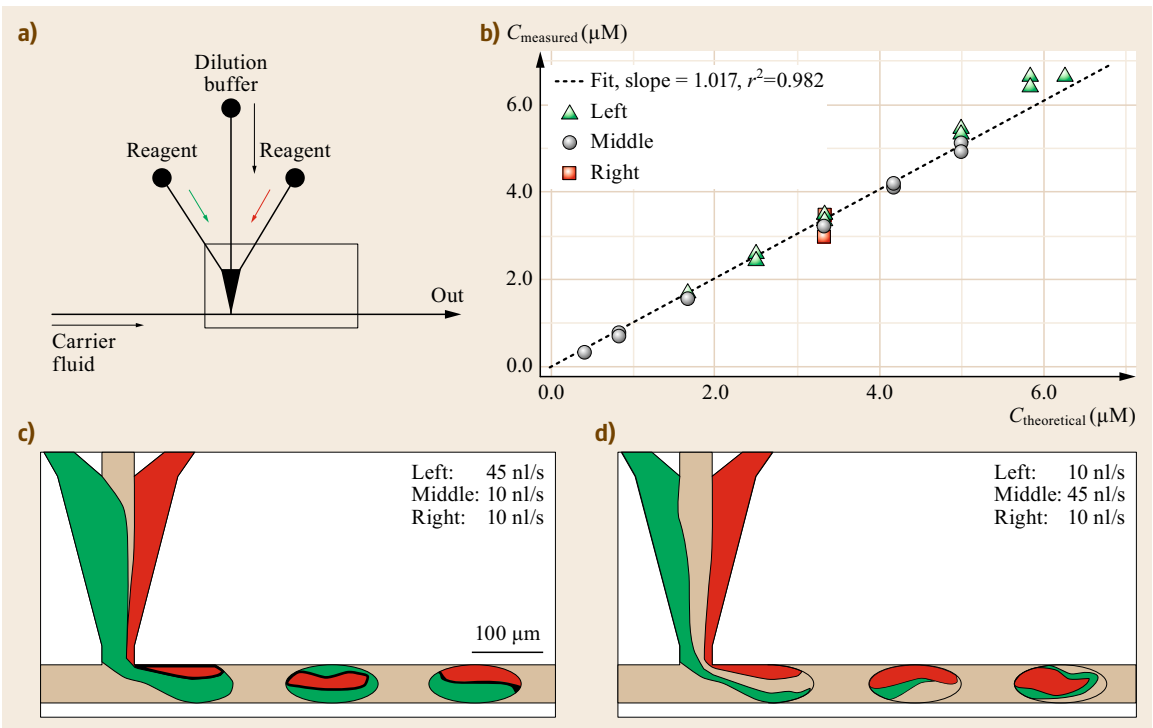


Fig. 17.22a–d Drug discovery in droplets. Demonstration of kinetic studies to explore the activity of enzymes and chemicals (after [17.78])

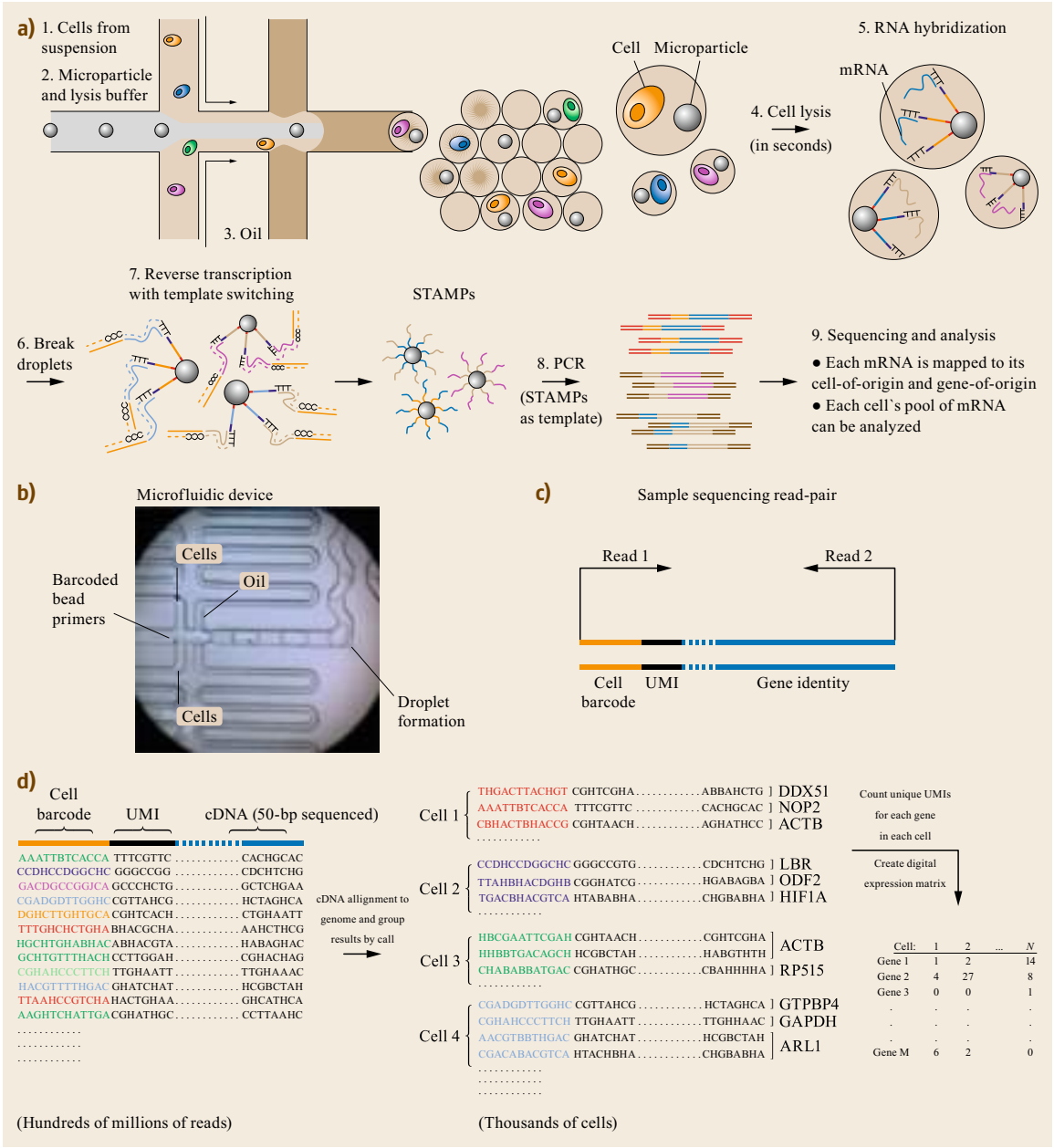


Fig. 17.23a–d The single-cell transcriptomes are captured using the hydrogel microbeads, which are encapsulated in the droplets along with the cell. Entire transcriptomes of the cells were barcoded for identification purposes and this technique has reported a 13% mRNA capture efficiency. Reprinted from [17.81] with permission from Elsevier

17.5 Conclusion

In this chapter, we introduced droplet microfluidic technology and its ability to manipulate fluids on the micro/nanoscale. The chapter also covered droplet generation techniques, current trends in droplet microfluidics and generic applications. The field of droplet microfluidics has matured over the last decade, as droplet microfluidic technologies are now being applied to a wide variety of commercial applications. There remain technical challenges to enable large networks of microfluidic operations in integrated devices that researchers are working on. However, the future of this promising technology is bright and it is predicted that droplet microfluidics will become an indispensable tool for a wide range of fluidic diagnostic platforms.

References

17.1 G.M. Whitesides: The origins and the future of microfluidics, Nature **442**(7101), 368–373 (2006)
17.2 T. Thorsen, S.J. Maerkl, S.R. Quake: Microfluidic large-scale integration, Science **298**(5593), 580–584 (2002)
17.3 J. Melin, S.R. Quake: Microfluidic large-scale integration: The evolution of design rules for biological automation, Annu. Rev. Biophys. Biomol. Struct. **36**(1), 213–231 (2007)
17.4 H. Song, D.L. Chen, R.F. Ismagilov: Reactions in droplets in microfluidic channels, Angew. Chem. Int. Ed. **45**(44), 7336–7356 (2006)
17.5 T. Thorsen, R.W. Roberts, F.H. Arnold, S.R. Quake: Dynamic pattern formation in a vesicle-generating microfluidic device, Phys. Rev. Lett. **86**(18), 4163–4166 (2001)
17.6 B. Zheng, J.D. Tice, R.F. Ismagilov: Formation of droplets of alternating composition in microfluidic channels and applications to indexing of concentrations in droplet-based assays, Anal. Chem. **76**(17), 4977–4982 (2004)
17.7 S.L. Anna, N. Bontoux, H.A. Stone: Formation of dispersions using flow focusing in microchannels, Appl. Phys. Lett. **82**(3), 364 (2003)
17.8 J. Shim, R.T. Ranasinghe, C.A. Smith, S.M. Ibrahim, F. Hollfelder, W.T.S. Huck, D. Klenerman, C. Abell: Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays, ACS Nano **7**(7), 5955–5964 (2013)
17.9 H.-H. Jeong, V.R. Yelleswarapu, S. Yadavali, D. Issadore, D. Lee: Kilo-scale droplet generation in three-dimensional monolithic elastomer device (3-D MED), Lab. Chip **15**(23), 4387–4392 (2015)
17.10 Y. Xia, G.M. Whitesides: Soft lithography, Annu. Rev. Mater. Sci. **28**(1), 153–184 (1998)
17.11 Y.-C. Tan, J.S. Fisher, A.I. Lee, V. Cristini, A.P. Lee: Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting, Lab. Chip **4**(4), 292 (2004)
17.12 X. Niu, S. Gulati, J.B. Edel, A.J. deMello: Pillar-induced droplet merging in microfluidic circuits, Lab. Chip **8**(11), 1837 (2008)
17.13 P. Singh, N. Aubry: Transport and deformation of droplets in a microdevice using dielectrophoresis, Electrophoresis **28**(4), 644–657 (2007)
17.14 C. Priest, S. Herminghaus, R. Seemann: Controlled electrocoalescence in microfluidics: Targeting a single lamella, Appl. Phys. Lett. **89**(13), 134101 (2006)
17.15 J. Köhler, T. Henkel, A. Grodrian, T. Kirner, M. Roth, K. Martin, J. Metze: Digital reaction technology by micro segmented flow–components, concepts and applications, Chem. Eng. J. **101**(1–3), 201–216 (2004)
17.16 R.M. Lorenz, J.S. Edgar, G.D.M. Jeffries, D.T. Chiu: Microfluidic and optical systems for the on-demand generation and manipulation of single femtoliter-volume aqueous droplets, Anal. Chem. **78**(18), 6433–6439 (2006)
17.17 D. Link, S. Anna, D. Weitz, H. Stone: Geometrically mediated breakup of drops in microfluidic devices, Phys. Rev. Lett. (2004) doi:10.1103/PhysRevLett.92.054503
17.18 D.N. Adamson, D. Mustafi, J.X.J. Zhang, B. Zheng, R.F. Ismagilov: Production of arrays of chemically distinct nanolitre plugs via repeated splitting in microfluidic devices, Lab. Chip **6**(9), 1178 (2006)
17.19 S.K. Cho, H. Moon, C.-J. Kim: Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits, J. Microelectromech. Syst. **12**(1), 70–80 (2003)
17.20 J.J. Agresti, E. Antipov, A.R. Abate, K. Ahn, A.C. Rowat, J.-C. Baret, M. Marquez, A.M. Klibanov, A.D. Griffiths, D.A. Weitz: Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, Proc. Natl. Acad. Sci. **107**(9), 4004–4009 (2010)
17.21 J. Lim, P. Gruner, M. Konrad, J.-C. Baret: Micro-optical lens array for fluorescence detection in droplet-based microfluidics, Lab. Chip **13**(8), 1472 (2013)
17.22 M. Kim, M. Pan, Y. Gai, S. Pang, C. Han, C. Yang, S.K.Y. Tang: Optofluidic ultrahigh-throughput detection of fluorescent drops, Lab. Chip **15**(6), 1417–1423 (2015)
17.23 D.-K. Kang, M.M. Ali, K. Zhang, S.S. Huang, E. Peterson, M.A. Digman, E. Gratton, W. Zhao: Rapid detection of single bacteria in unprocessed blood using integrated comprehensive droplet digital detection, Nat. Commun. **5**, 5427 (2014)
17.24 S. Liu, Y. Gu, R.B. Le Roux, S.M. Matthews, D. Bratton, K. Yunus, A.C. Fisher, W.T.S. Huck: The electrochemical detection of droplets in microfluidic devices, Lab. Chip **8**(11), 1937 (2008)

17.25 L.M. Fidalgo, G. Whyte, B.T. Ruotolo, J.L.P. Benesch, F. Stengel, C. Abell, C.V. Robinson, W.T.S. Huck: Coupling microdroplet microreactors with mass spectrometry: Reading the contents of single droplets online, Angew. Chem. Int. Ed. **48**(20), 3665–3668 (2009)

17.26 X.Z. Niu, B. Zhang, R.T. Marszalek, O. Ces, J.B. Edel, D.R. Klug, A.J. deMello: Droplet-based compartmentalization of chemically separated components in two-dimensional separations, Chem. Commun. (2009) doi:10.1039/b918100h

17.27 M.P. Cecchini, J. Hong, C. Lim, J. Choo, T. Albrecht, A.J. deMello, J.B. Edel: Ultrafast surface enhanced resonance raman scattering detection in droplet-based microfluidic systems, Anal. Chem. **83**(8), 3076–3081 (2011)

17.28 Y.-C. Tan, Y.L. Ho, A.P. Lee: Microfluidic sorting of droplets by size, Microfluid. Nanofluidics **4**(4), 343–348 (2007)

17.29 M. Chabert, J.-L. Viovy: Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells, Proc. Natl. Acad. Sci. **105**(9), 3191–3196 (2008)

17.30 H.N. Joensson, M. Uhlén, H.A. Svahn: Droplet size based separation by deterministic lateral displacement–separating droplets by cell-induced shrinking, Lab. Chip **11**(7), 1305 (2011)

17.31 A.C. Hatch, A. Patel, N.R. Beer, A.P. Lee: Passive droplet sorting using viscoelastic flow focusing, Lab. Chip **13**(7), 1308 (2013)

17.32 J.-C. Baret, O.J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M.L. Samuels, J.B. Hutchison, J.J. Agresti, D.R. Link, D.A. Weitz, A.D. Griffiths: Fluorescence-activated droplet sorting (FADS): Efficient microfluidic cell sorting based on enzymatic activity, Lab. Chip **9**(13), 1850 (2009)

17.33 K. Ahn, C. Kerbage, T.P. Hunt, R.M. Westervelt, D.R. Link, D.A. Weitz: Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices, Appl. Phys. Lett. **88**(2), 24104 (2006)

17.34 T. Franke, A.R. Abate, D.A. Weitz, A. Wixforth: Surface acoustic wave (SAW) directed droplet flow in microfluidics for PDMS devices, Lab. Chip **9**(18), 2625 (2009)

17.35 A.R. Abate, J.J. Agresti, D.A. Weitz: Microfluidic sorting with high-speed single-layer membrane valves, Appl. Phys. Lett. **96**(20), 203509 (2010)

17.36 C.N. Baroud, M.R. de Saint Vincent, J.-P. Delville: An optical toolbox for total control of droplet microfluidics, Lab. Chip **7**(8), 1029 (2007)

17.37 S.-Y. Teh, R. Lin, L.-H. Hung, A.P. Lee: Droplet microfluidics, Lab. Chip **8**(2), 198 (2008)

17.38 G.F. Christopher, S.L. Anna: Microfluidic methods for generating continuous droplet streams, J. Phys. Appl. Phys. **40**(19), R319–R336 (2007)

17.39 M. Baker: Digital PCR hits its stride, Nat. Methods **9**(6), 541–544 (2012)

17.40 A.R. Abate, T. Hung, R.A. Sperling, P. Mary, A. Rotem, J.J. Agresti, M.A. Weiner, D.A. Weitz: DNA sequence analysis with droplet-based microfluidics, Lab. Chip **13**(24), 4864 (2013)

17.41 S. Abalde-Cela, A. Gould, X. Liu, E. Kazamia, A.G. Smith, C. Abell: High-throughput detection of ethanol-producing cyanobacteria in a microdroplet platform, J. R. Soc. Interface **12**(106), 20150216–20150216 (2015)

17.42 P.S. Dittrich, A. Manz: Lab-on-a-chip: Microfluidics in drug discovery, Nat. Rev. Drug Discov. **5**(3), 210–218 (2006)

17.43 E. Brouzes, M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J.B. Hutchison, J.M. Rothberg, D.R. Link, N. Perrimon, M.L. Samuels: Droplet microfluidic technology for single-cell high-throughput screening, Proc. Natl. Acad. Sci. **106**(34), 14195–14200 (2009)

17.44 A.C. Larsen, M.R. Dunn, A. Hatch, S.P. Sau, C. Youngbull, J.C. Chaput: A general strategy for expanding polymerase function by droplet microfluidics, Nat. Commun. **7**, 11235 (2016)

17.45 P.B. Umbanhowar, V. Prasad, D.A. Weitz: Monodisperse emulsion generation via drop break off in a coflowing stream, Langmuir **16**(2), 347–351 (2000)

17.46 C. Cramer, P. Fischer, E.J. Windhab: Drop formation in a co-flowing ambient fluid, Chem. Eng. Sci. **59**(15), 3045–3058 (2004)

17.47 G.F. Christopher, N.N. Noharuddin, J.A. Taylor, S.L. Anna: Experimental observations of the squeezing-to-dripping transition in T-shaped microfluidic junctions, Phys. Rev. E **78**(3), 36317 (2008)

17.48 J. Husny, J.J. Cooper-White: The effect of elasticity on drop creation in T-shaped microchannels, J. Non-Newton. Fluid Mech. **137**(1–3), 121–136 (2006)

17.49 J. Xu, G. Luo, G. Chen, J. Wang: Experimental and theoretical approaches on droplet formation from a micrometer screen hole, J. Membr. Sci. **266**(1–2), 121–131 (2005)

17.50 J.D. Tice, A.D. Lyon, R.F. Ismagilov: Effects of viscosity on droplet formation and mixing in microfluidic channels, Anal. Chim. Acta **507**(1), 73–77 (2004)

17.51 P. Garstecki, H. Stone, G. Whitesides: Mechanism for flow-rate controlled breakup in confined geometries: A route to monodisperse emulsions, Phys. Rev. Lett. (2005) doi:10.1103/PhysRevLett.94.164501

17.52 P. Garstecki, M.J. Fuerstman, H.A. Stone, G.M. Whitesides: Formation of droplets and bubbles in a microfluidic T-junction–scaling and mechanism of break-up, Lab. Chip **6**(3), 437 (2006)

17.53 S.L. Anna, H.C. Mayer: Microscale tipstreaming in a microfluidic flow focusing device, Phys. Fluids **18**(12), 121512 (2006)

17.54 A.S. Utada: Monodisperse double emulsions generated from a microcapillary device, Science **308**(5721), 537–541 (2005)

17.55 T. Ward, M. Faivre, M. Abkarian, H.A. Stone: Microfluidic flow focusing: Drop size and scaling in pressureversus flow-rate-driven pumping, Electrophoresis **26**(19), 3716–3724 (2005)

17.56 C. Zhou, P. Yue, J.J. Feng: Formation of simple and compound drops in microfluidic devices, Phys. Fluids **18**(9), 92105 (2006)

17.57 B. Beulen, J. de Jong, H. Reinten, M. van den Berg, H. Wijshoff, R. van Dongen: Flows on the nozzle

plate of an inkjet printhead, Exp. Fluids **42**(2), 217–224 (2007)
17.58 H. Willaime, V. Barbier, L. Kloul, S. Maine, P. Tabeling: Arnold tongues in a microfluidic drop emitter, Phys. Rev. Lett. (2006) doi:10.1103/PhysRevLett.96.054501
17.59 O. Ozen, N. Aubry, D.T. Papageorgiou, P.G. Petropoulos: Monodisperse drop formation in square microchannels, Phys. Rev. Lett. (2006) doi:10.1103/PhysRevLett.96.144501
17.60 D.T. Chiu, R.M. Lorenz, G.D.M. Jeffries: Droplets for ultrasmall-volume analysis, Anal. Chem. **81**(13), 5111–5118 (2009) doi:10.1021/ac900306q
17.61 A. Huebner, S. Sharma, M. Srisa-Art, F. Hollfelder, J.B. Edel, A.J. deMello: Microdroplets: A sea of applications?, Lab. Chip **8**(8), 1244 (2008)
17.62 A.K. Price, B.M. Paegel: Discovery in droplets, Anal. Chem. **88**(1), 339–353 (2016)
17.63 I. Shestopalov, J.D. Tice, R.F. Ismagilov: Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system, Lab. Chip **4**(4), 316 (2004)
17.64 V. Noireaux, A. Libchaber: A vesicle bioreactor as a step toward an artificial cell assembly, Proc. Natl. Acad. Sci. **101**(51), 17669–17674 (2004)
17.65 B. Ahmed, D. Barrow, T. Wirth: Enhancement of reaction rates by segmented fluid flow in capillary scale reactors, Adv. Synth. Catal. **348**(9), 1043–1048 (2006)
17.66 J.R. Burns, C. Ramshaw: The intensification of rapid reactions in multiphase systems using slug flow in capillaries, Lab. Chip **1**(1), 10 (2001)
17.67 K.-I. Sotowa, K. Irie, T. Fukumori, K. Kusakabe, S. Sugiyama: Droplet formation by the collision of two aqueous solutions in a microchannel and application to particle synthesis, Chem. Eng. Technol. **30**(3), 383–388 (2007)
17.68 H. Song, H.-W. Li, M.S. Munson, T.G. Van Ha, R.F. Ismagilov: On-chip titration of an anticoagulant argatroban and determination of the clotting time within whole blood or plasma using a plug-based microfluidic system, Anal. Chem. **78**(14), 4839–4849 (2006)
17.69 Z.T. Cygan, J.T. Cabral, K.L. Beers, E.J. Amis: Microfluidic platform for the generation of organic-phase microreactors, Langmuir **21**(8), 3629–3634 (2005)
17.70 A. Huebner, M. Srisa-Art, D. Holt, C. Abell, F. Hollfelder, A.J. deMello, J.B. Edel: Quantitative detection of protein expression in single cells using droplet microfluidics, Chem. Commun. (2007) doi:10.1039/b618570c
17.71 M.S. Long, C.D. Jones, M.R. Helfrich, L.K. Mangeney-Slavin, C.D. Keating: Dynamic microcompartmentation in synthetic cells, Proc. Natl. Acad. Sci. **102**(17), 5920–5925 (2005)
17.72 V. Taly, D. Pekin, A.E. Abed, P. Laurent-Puig: Detecting biomarkers with microdroplet technology, Trends Mol. Med. **18**(7), 405–416 (2012)
17.73 W. Wang, Z.-X. Li, R. Luo, S.-H. Lü, A.-D. Xu, Y.-J. Yang: Droplet-based micro oscillating-flow PCR chip, J. Micromech. Microeng. **15**(8), 1369–1377 (2005)
17.74 B.J. Hindson, K.D. Ness, D.A. Masquelier, P. Belgrader, N.J. Heredia, A.J. Makarewicz, I.J. Bright, M.Y. Lucero, A.L. Hiddessen, T.C. Legler, T.K. Kitano, M.R. Hodel, J.F. Petersen, P.W. Wyatt, E.R. Steenblock, P.H. Shah, L.J. Bousse, C.B. Troup, J.C. Mellen, D.K. Wittmann, N.G. Erndt, T.H. Cauley, R.T. Koehler, A.P. So, S. Dube, K.A. Rose, L. Montesclaros, S. Wang, D.P. Stumbo, S.P. Hodges, S. Romine, F.P. Milanovich, H.E. White, J.F. Regan, G.A. Karlin-Neumann, C.M. Hindson, S. Saxonov, B.W. Colston: High-throughput droplet digital PCR system for absolute quantitation of DNA copy number, Anal. Chem. **83**(22), 8604–8610 (2011)
17.75 T. Hatakeyama, D.L. Chen, R.F. Ismagilov: Microgram-scale testing of reaction conditions in solution using nanoliter plugs in microfluidics with detection by MALDI-MS, J. Am. Chem. Soc. **128**(8), 2518–2519 (2006)
17.76 A.R. Wheeler, H. Moon, C.A. Bird, R.R.O. Loo, C.-J. Kim, J.A. Loo, R.L. Garrell: Digital microfluidics with in-line sample purification for proteomics analyses with MALDI-MS, Anal. Chem. **77**(2), 534–540 (2005)
17.77 B.T.C. Lau, C.A. Baitz, X.P. Dong, C.L. Hansen: A complete microfluidic screening platform for rational protein crystallization, J. Am. Chem. Soc. **129**(3), 454–455 (2007)
17.78 H. Song, R.F. Ismagilov: Millisecond kinetics on a microfluidic chip using nanoliters of reagents, J. Am. Chem. Soc. **125**(47), 14613–14619 (2003)
17.79 B. Kintses, C. Hein, M.F. Mohamed, M. Fischlechner, F. Courtois, C. Lainé, F. Hollfelder: Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution, Chem. Biol. **19**(8), 1001–1009 (2012)
17.80 D.J. Eastburn, A. Sciambi, A.R. Abate: Ultrahigh-throughput mammalian single-cell reverse-transcriptase polymerase chain reaction in microfluidic drops, Anal. Chem. **85**(16), 8016–8021 (2013)
17.81 E.Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martersteck, J.J. Trombetta, D.A. Weitz, J.R. Sanes, A.K. Shalek, A. Regev, S.A. McCarroll: Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, Cell **161**(5), 1202–1214 (2015)

18. Nanorobotics

Bradley J. Nelson, Lixin Dong

Nanorobotics is the study of robotics at the nanometer scale, and includes robots that are nanoscale in size and large robots capable of manipulating objects that have dimensions in the nanoscale range with nanometer resolution. With the ability to position and orient nanometer-scale objects, nanorobotic manipulation is a promising way to enable the assembly of nanosystems including nanorobots.

This chapter overviews the state of the art of nanorobotics, outlines nanoactuation, and focuses on nanorobotic manipulation systems and their application in nanoassembly, biotechnology and the construction and characterization of nanoelectromechanical systems (NEMS) through a hybrid approach.

Because of their exceptional properties and unique structures, carbon nanotubes (CNTs) and SiGe/Si nanocoils are used to show basic processes of nanorobotic manipulation, structuring and assembly, and for the fabrication of NEMS including nano tools, sensors and actuators.

A series of processes of nanorobotic manipulation, structuring and assembly has been demonstrated experimentally. Manipulation of individual CNTs in three-dimensional (3-D) free space has been shown by grasping using dielectrophoresis and placing with both position and orientation control for mechanical and electrical property characterization and assembly of nanostructures and devices. A variety of material property investigations can be performed, including bending, buckling, and pulling to investigate elasticity as well as strength and tribological characterization. Structuring of CNTs can be performed including shape modification, the exposure of nested cores, and connecting CNTs by van der Waals forces, electron-beam-induced deposition and mechanochemical bonding. Nanorobotics provides novel techniques for exploring the biodomain by manipulation and characterization of nanoscale objects such as cellular membranes, DNA and other biomolecules. Nano tools, sensors and actuators can provide measurements and/or movements that are calculated in nanometers, gigahertz, piconewtons, femtograms, etc., and are promising for molecular machines and bio- and nanorobotics applications. Efforts are focused on developing enabling technologies for nanotubes and other nanomaterials and structures for NEMS and nanorobotics. By combining bottom-up nanorobotic manipulation and top-down nanofabrication processes, a hybrid approach is demonstrated for creating complex 3-D nanodevices. Nanomaterial science, bionanotechnology, and nanoelectronics will benefit from advances in nanorobotics.

18.1 **Overview of Nanorobotics** 560

18.2 **Actuation at Nanoscales** 561
18.2.1 Electrostatics 562
18.2.2 Electromagnetics 562
18.2.3 Piezoelectrics 562
18.2.4 Other Techniques 563

18.3 **Nanorobotic Manipulation Systems** 563
18.3.1 Overview 563
18.3.2 Nanorobotic Manipulation Systems 565

18.4 **Nanorobotic Assembly** 568
18.4.1 Overview 568
18.4.2 Carbon Nanotubes 569
18.4.3 Nanocoils 574

18.5 **Applications** 576
18.5.1 Robotic Biomanipulation 576
18.5.2 Nanorobotic Devices 578

References 580

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_18

18.1 Overview of Nanorobotics

Progress in robotics over the past years has dramatically extended our ability to explore the world from perception, cognition and manipulation perspectives at a variety of scales extending from the edges of the solar system down to individual atoms (Fig. 18.1). At the bottom of this scale, technology has been moving toward greater control of the structure of matter, suggesting the feasibility of achieving thorough control of the molecular structure of matter atom by atom as *Richard Feynman* first proposed in 1959 in his prophetic article on miniaturization [18.1]:

> *What I want to talk about is the problem of manipulating and controlling things on a small scale... I am not afraid to consider the final question as to whether, ultimately – in the great future – we can arrange the atoms the way we want: the very atoms, all the way down!*

He asserted that:

> *At the atomic level, we have new kinds of forces and new kinds of possibilities, new kinds of effects. The problems of manufacture and reproduction of materials will be quite different. The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom.*

This technology is now labeled *nanotechnology*.

The *great future* of Feynman began to be realized in the 1980s. Some of the capabilities he dreamed of have been demonstrated, while others are being developed. Nanorobotics represents the next stage in miniaturization for maneuvering nanoscale objects. Nanorobotics is the study of robotics at the nanometer scale, and includes robots that are nanoscale in size, i. e., nanorobots, and large robots capable of manipulating objects that have dimensions in the nanoscale range with nanometer resolution, i. e., nanorobotic manipulators. The field of nanorobotics brings together several disciplines, including nanofabrication processes used for producing nanoscale robots, nanoactuators, nanosensors, and physical modeling at nanoscales. Nanorobotic manipulation technologies, including the assembly of nanometer-sized parts, the manipulation of biological cells or molecules, and the types of robots used to perform these types of tasks also form a component of nanorobotics.

As the 21st century unfolds, the impact of nanotechnology on the health, wealth, and security of humankind is expected to be at least as significant as the combined influences in the 20th century of antibiotics, the integrated circuit, and human-made polymers. For example, *Lane* stated in 1998 [18.2]:

> *If I were asked for an area of science and engineering that will most likely produce the breakthroughs*

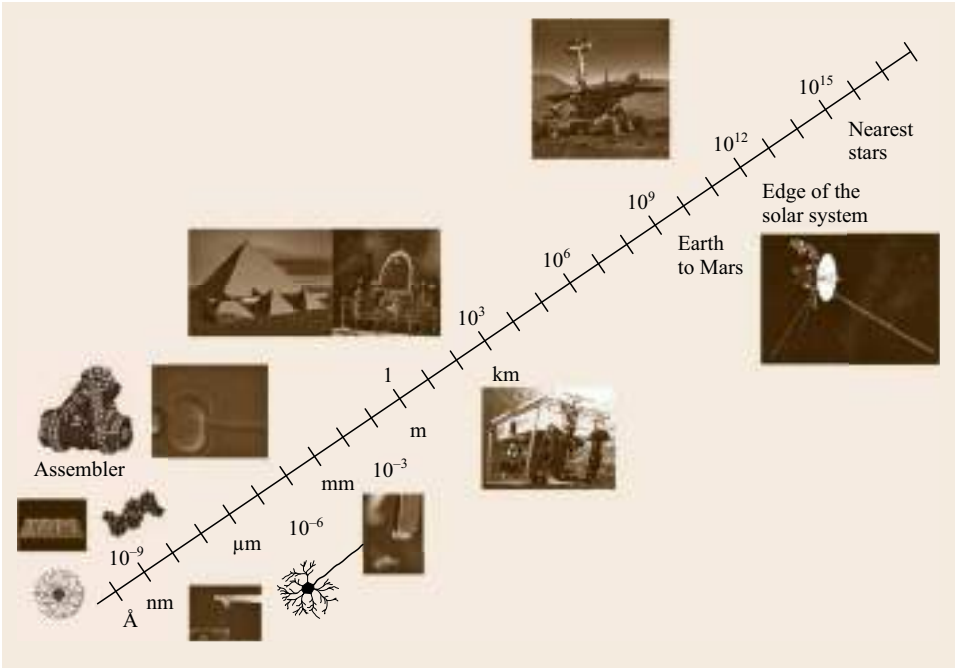


Fig. 18.1 Robotic exploration

of tomorrow, I would point to nanoscale science and engineering.

The great scientific and technological opportunities nanotechnology presents have stimulated extensive exploration of the nanoworld and initiated an exciting worldwide competition, which has been accelerated by the publication of the *National Nanotechnology Initiative* by the US government in 2000 [18.3]. Nanorobotics will play a significant role as an enabling nanotechnology and could ultimately be a core part of nanotechnology if *Drexler*'s machine-phase nanosystems based on self-replicative molecular assemblers via mechanosynthesis can be realized [18.4].

By the early 1980s, scanning tunneling microscopes (STMs) [18.5] radically changed the ways in which we interacted with and even regarded single atoms and molecules. The very nature of proximal probe methods encourages exploration of the nanoworld beyond conventional microscopic imaging. Scanned probes now allow us to perform engineering operations on single molecules, atoms, and bonds, thereby providing a tool that operates at the ultimate limits of fabrication. They have also enabled exploration of molecular properties on an individual nonstatistical basis.

STMs and other nanomanipulators are nonmolecular machines but use bottom-up strategies. Although performing only one molecular reaction at a time is obviously impractical for making large amounts of a product, it is a promising way to provide the next generation of nanomanipulators. Most importantly, it is possible to realize the directed assembly of molecules or supermolecules to build larger nanostructures through nanomanipulation. The products produced by nanomanipulation could be the first step of a bottom-up strategy in which these assembled products are used to self-assemble into nanomachines.

One of the most important applications of nanorobotic manipulation will be nanorobotic assembly. However, it appears that until assemblers capable of replication can be built, the parallelism of chemical synthesis and self-assembly are necessary when starting from atoms; groups of molecules can self-assemble quickly due to their thermal motion, enabling them to *explore* their environments and find (and bind to) complementary molecules. Given their key role in natural molecular machines, proteins are obvious candidates for early work in self-assembling artificial molecular systems. *Degrado* [18.6] demonstrated the feasibility of designing protein chains that predictably fold into solid molecular objects. Progress is also being made in artificial enzymes and other relatively small molecules that perform functions like those of natural proteins; the 1987 Nobel prize for chemistry went to *Cram* and *Lehn* for such work on supramolecular chemistry [18.7]. Several bottom-up strategies using self-assembly appear feasible [18.8]. *Fujita* et al.'s pioneering work has shown that self-assembly can be directed by adroitly exploiting the chemical and electrical bonds that hold natural molecules together, and hence get molecules to form desired nanometer-scale structures [18.9]. Chemical synthesis, self assembly, and supramolecular chemistry make it possible to provide building blocks at relatively large sizes beginning from the nanometer scale. Nanorobotic manipulation serves as the base for a hybrid approach to construct nanodevices by structuring these materials to obtain building blocks and assembling them into more complex systems.

This chapter focuses on nanorobotics including actuation, manipulation and assembly at the nanoscale. The main goal of nanorobotics is to provide an effective technology for the experimental exploration of the nanoworld, and to push the boundaries of this exploration from a robotics research perspective.

18.2 Actuation at Nanoscales

The positioning of nanorobots and nanorobotic manipulators depends largely on nanoactuators. While nanosized actuators for nanorobots are still under exploration and relatively far from implementation, microelectromechanical system (MEMS)-based efforts are focused on shrinking their sizes [18.10]. Nanometer-resolution motion has been extensively investigated and can be generated using various actuation principles. Electrostatics, electromagnetics, and piezoelectrics are the most common ways to realize actuation at nanoscales. For nanorobotic manipulation, besides nanoresolution and compact sizes, actuators generating large strokes and high forces are best suited for such applications. The speed criteria are of less importance as long as the actuation speed is in the range of a couple of hertz and above. Table 18.1 provides a small selection of early works on actuators [18.11–16] suitable in actuation principle actuators suitable for nanorobotic applications (partially adapted from [18.10]).

Several extensive reviews on various actuation principles have been published [18.17–21]. During the design of an actuator, the tradeoffs among range of motion, force, speed (actuation frequency), power

Table 18.1 Actuation with MEMS

Actuation principle	Type of motion	Volume (mm^3)	Speed (s^{-1})	Force (N)	Stroke (m)	Resolution (m)	Power density (W/m^3)	Ref.
Electrostatic	Linear	400	5000	1×10^{-7}	6×10^{-6}	n/a	200	[18.12]
Magnetic	Linear	$0.4\times0.4\times0.5$	1000	2.6×10^{-6}	1×10^{-4}	n/a	3000	[18.13]
Piezoelectric	Linear	$25.4\times12.7\times1.6$	4000	350	1×10^{-3}	7×10^{-8}	n/a	[18.14]
Actuation principle	**Type of motion**	**Volume (mm^3)**	**Speed (rad/s)**	**Torque (N m)**	**Stroke (rad)**	**Resolution (rad)**	**Power density (W/m^3)**	**Ref.**
Electrostatic	Rotational	$\pi/4\times0.5^2\times3$	40	2×10^{-7}	2π	n/a	900	[18.15]
Magnetic	Rotational	$2\times3.7\times0.5$	150	1×10^{-6}	2π	$5/36\pi$	3000	[18.16]
Piezoelectric	Rotational	$\pi/4\times1.5^2\times0.5$	30	2×10^{-11}	0.7	n/a	n/a	[18.17]

consumption, control accuracy, system reliability, robustness, load capacity, etc. must be taken into consideration. This section reviews basic actuation technologies and potential applications at nanometer scales.

18.2.1 Electrostatics

Electrostatic charge arises from a build up or deficit of free electrons in a material, which can exert an attractive force on oppositely charged objects, or a repulsive force on similarly charged objects. Since electrostatic fields arise and disappear rapidly, such devices will likewise demonstrate very fast operation speeds and be little affected by ambient temperatures.

Recent investigations have produced many examples of miniature devices using electrostatic force for actuation including silicon micro motors [18.22, 23], microvalves [18.24], and microtweezers [18.25]. This type of actuation is important for achieving nanosized actuation.

Electrostatic fields can exert great forces, but generally across very short distances. When the electric field must act over larger distances, a higher voltage will be required to maintain a given force. The extremely low-current consumption associated with electrostatic devices makes for highly efficient actuation.

18.2.2 Electromagnetics

Electromagnetism arises from electric current moving through a conducting material. Attractive or repulsive forces are generated adjacent to the conductor and proportional to the current flow. Structures can be built which gather and focus electromagnetic forces, and harness these forces to create motion.

Electromagnetic fields arise and disappear rapidly, thus permitting devices with very fast operation speeds. Since electromagnetic fields can exist over a wide range of temperatures, performance is primarily limited by the properties of the materials used in constructing the actuator.

One example of a microfabricated electromagnetic actuator is a microvalve which uses a small electromagnetic coil wrapped around a silicon micromachined valve structure [18.26]. However, the downward scalability of electromagnetic actuators into the micro- and nanorealm may be limited by the difficulty of fabricating small electromagnetic coils. Furthermore, most electromagnetic devices require perpendicularity between the current conductor and the moving element, presenting a difficulty for planar fabrication techniques commonly used to make silicon devices.

An important advantage of electromagnetic devices is their high efficiency in converting electrical energy into mechanical work. This translates into less current consumption from the power source.

18.2.3 Piezoelectrics

Piezoelectric motion arises from the dimensional changes generated in certain crystalline materials when subjected to an electric field or to an electric charge. Structures can be built which gather and focus the force of the dimensional changes, and harness them to create motion. Typical piezoelectric materials include quartz (SiO_2), lead zirconate titanate (PZT), lithium niobate, and polymers such as polyvinyledene fluoride (PVDF).

Piezoelectric materials respond very quickly to changes in voltages and with great repeatability. They can be used to generate precise motions with repeatable oscillations, as in quartz timing crystals used in many electronic devices. Piezo materials can also act as sensors, converting tension or compression strains to voltages.

On the microscale, piezoelectric materials have been used in linear inchworm drive devices [18.27], and micropumps [18.28]. STMs and most nanomanipulators use piezoelectric actuators.

Piezo materials operate with high force and speed, and return to a neutral position when unpowered. They exhibit very small strokes (under 1 %). Alternating electric currents produce oscillations in the piezo

Table 18.2 Comparison of nanoactuators

Method	Efficiency	Speed	Power density
Electrostatic	Very high	Fast	Low
Electromagnetic	High	Fast	High
Piezoelectric	Very high	Fast	High
Thermomechanical	Very high	Medium	Medium
Phase change	Very high	Medium	High
Shape memory	Low	Medium	Very high
Magnetostrictive	Medium	Fast	Very high
Electrorheological	Medium	Medium	Medium
Electrohydrodynamic	Medium	Medium	Low
Diamagnetism	High	Fast	High

material, and operation at the sample's fundamental resonant frequency produces the largest elongation and highest power efficiency [18.29]. Piezo actuators working in the stick–slip mode can provide millimeter to centimeter strokes. Most commercially available nanomanipulators adopt this type of actuators, such as Picomotors from New Focus and Nanomotors from Klock.

18.2.4 Other Techniques

Other techniques include thermomechanical, phase change, shape memory, magnetostrictive, electrorheological, electrohydrodynamic, diamagnetism, magnetohydrodynamic, shape changing, polymers, biological methods (living tissues, muscle cells, etc.) and so on. Table 18.2 lists a comparison of these.

18.3 Nanorobotic Manipulation Systems

18.3.1 Overview

Nanomanipulation, or positional and/or force control at the nanometer scale, is a key enabling technology for nanotechnology by filling the gap between top-down and bottom-up strategies, and may lead to the appearance of replication-based molecular assemblers [18.4]. These types of assemblers have been proposed as general-purpose manufacturing devices for building a wide range of useful products as well as copies of themselves (self-replication).

Presently, nanomanipulation can be applied to the scientific exploration of mesoscopic physical phenomena, biology and the construction of prototype nanodevices. It is a fundamental technology for property characterization of nanomaterials, structures and mechanisms, for the preparation of nanobuilding blocks, and for the assembly of nanodevices such as nanoelectromechanical systems (NEMS).

Nanomanipulation was enabled by the inventions of the STM [18.5], atomic force microscope (AFM) [18.30], and other types of scanning probe microscope (SPM). Besides these, optical tweezers (laser trapping) [18.31] and magnetic tweezers [18.32] are also possible nanomanipulators. Nanorobotic manipulators (NRMs) [18.33, 34] are characterized by the capability of 3-D positioning, orientation control, independently actuated multiple end-effectors, and independent real-time observation systems, and can be integrated with scanning probe microscopes. NRMs largely extend the complexity of nanomanipulation.

A concise comparison of STM, AFM, and NRM technology is shown in Fig. 18.2. With its incomparable imaging resolution, an STM can be applied to particles as small as atoms with atomic resolution. However, limited by its two-dimensional (2-D) positioning and available strategies for manipulations, standard STMs

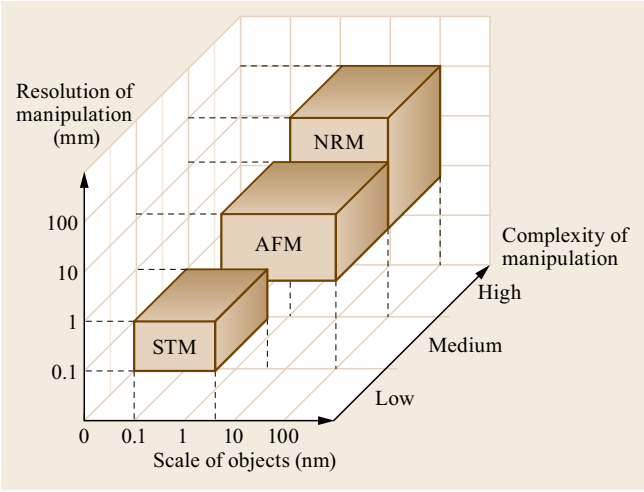


Fig. 18.2 Comparison of nanomanipulators

are ill-suited for complex manipulation and cannot be used in 3-D space. An AFM is another important type of nanomanipulator. There are three imaging modes for AFMs, i. e., contact mode, tapping mode (periodic contact mode), and non-contact mode. The latter two are also called dynamic modes and can attain higher imaging resolution than the contact mode. Atomic resolution is obtainable with non-contact mode. Manipulation with an AFM can be done in either contact or dynamic mode. Generally, manipulation with an AFM involves moving an object by touching it with a tip. A typical manipulation is like this: image a particle first in non-contact mode, then remove the tip oscillation voltage and sweep the tip across the particle in contact with the surface and with the feedback disabled. Mechanical pushing can exert larger forces on objects and, hence, can be applied for the manipulation of relatively larger objects. One-dimensional (1-D) to 3-D objects can be manipulated on a 2-D substrate. However, the manipulation of individual atoms with an AFM remains a challenge. By separating the imaging and manipulation functions, nanorobotic manipulators can have many more degrees of freedom including rotation for orientation control, and, hence, can be used for the manipulation of zero-dimensional (0-D, symmetric spheres) to 3-D objects in 3-D free space. Limited by the relative lower resolution of electron microscopes, NRMs are difficult to use for the manipulation of atoms. However, their general robotic capabilities including 3-D positioning, orientation control, independently actuated multiple end-effectors, separate real-time observation system, and integrations with SPMs inside makes NRMs quite promising for complex nanomanipulation.

The first nanomanipulation experiment was performed by *Eigler* and *Schweizer* in 1990 [18.35]. They used an STM and materials at low temperatures (4 K) to position individual xenon atoms on a single-crystal nickel surface with atomic precision. The manipulation enabled them to fabricate rudimentary structures of their own design, atom by atom. The result is the famous set of images showing how 35 atoms were moved to form the three-letter logo *IBM*, demonstrating that matter could indeed be maneuvered atom by atom as *Feynman* suggested [18.1].

A nanomanipulation system generally includes nanomanipulators as the positioning device, microscopes as *eyes*, various end-effectors including probes and tweezers among others as its *fingers*, and types of sensors (force, displacement, tactile, strain, etc.) to facilitate the manipulation and/or to determine the properties of the objects. Key technologies for nanomanipulation include observation, actuation, measurement, system design and fabrication, calibration and control, communication, and human–machine interface.

Strategies for nanomanipulation are basically determined by the environment – air, liquid or vacuum – which is further decided by the properties and size of the objects and observation methods. Figure 18.3 depicts the microscopes, environments and strategies of nanomanipulation. In order to observe manipulated objects, STMs can provide subangstrom imaging resolution, whereas AFMs can provide atomic resolution. Both can obtain 3-D surface topology. Because AFMs can be used in an ambient environment, they provide a powerful tool for biomanipulation that may require a liquid environment. The resolution of scanning electron microscopes (SEMs) is limited to about 1 nm, whereas field-emission SEMs (FESEMs) can achieve higher resolutions. SEMs/FESEMs can be used for 2-D real-time observation for both the objects and end-effectors of manipulators, and large ultrahigh-vacuum (UHV) sample chambers provide enough space to contain an NRM with many degrees of freedom (DOFs) for 3-D nanomanipulation. However, the 2-D nature of the observation makes positioning along the electron-beam direction difficult. High-resolution transmission electron microscopes (HRTEMs) can provide atomic resolution. However, the narrow UHV specimen chamber makes it difficult to incorporate large manipulators. In principle, optical microscopes (OMs) cannot be used for nanometer-scale (smaller than the wavelength of visible lights) observation because of diffraction limits. Scanning near-field OMs (SNOMs) break this limitation and are promising as a real-time observation

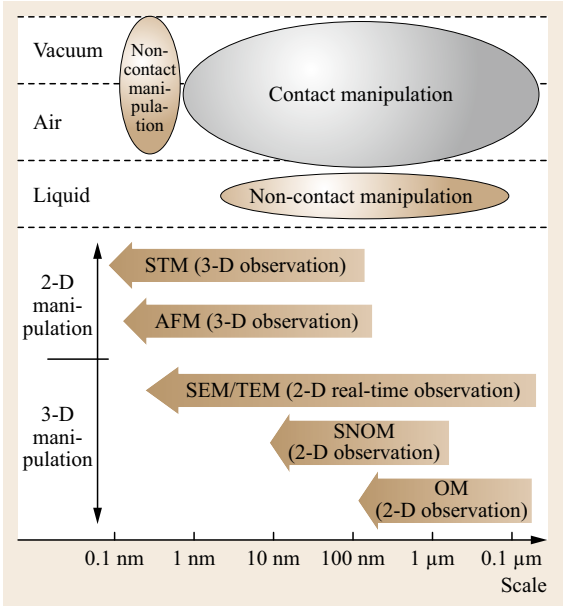


Fig. 18.3 Microscopes, environments and strategies of nanomanipulation

device for nanomanipulation, especially for ambient environments. SNOMs can be combined with AFMs, and potentially with NRMs for nanoscale biomanipulation.

Nanomanipulation processes can be broadly classified into three types: (1) lateral non-contact, (2) lateral contact, and (3) vertical manipulation. Generally, lateral non-contact nanomanipulation is mainly applied for atoms and molecules in UHV with an STM or bio-object in liquid using optical or magnetic tweezers. Contact nanomanipulation can be used in almost any environment, generally with an AFM, but is difficult for atomic manipulation. Vertical manipulation can be performed by NRMs. Figure 18.4 shows the processes of the three basic strategies.

Motion of the lateral noncontact manipulation processes are shown in Fig. 18.4a. Applicable effects [18.36] able to cause the motion include long-range van der Waals (vdW) forces (attractive) generated by the proximity of the tip to the sample [18.37], electric-field-induced fields caused by the voltage bias between the tip and the sample [18.38, 39], tunneling current local heating or inelastic tunneling vibration [18.40, 41]. With these methods, some nanodevices and molecules have been assembled [18.42, 43]. Laser trapping (optical tweezers) and magnetic tweezers are possible for non-contact manipulation of nanoorder biosamples, e.g., DNA [18.44, 45].

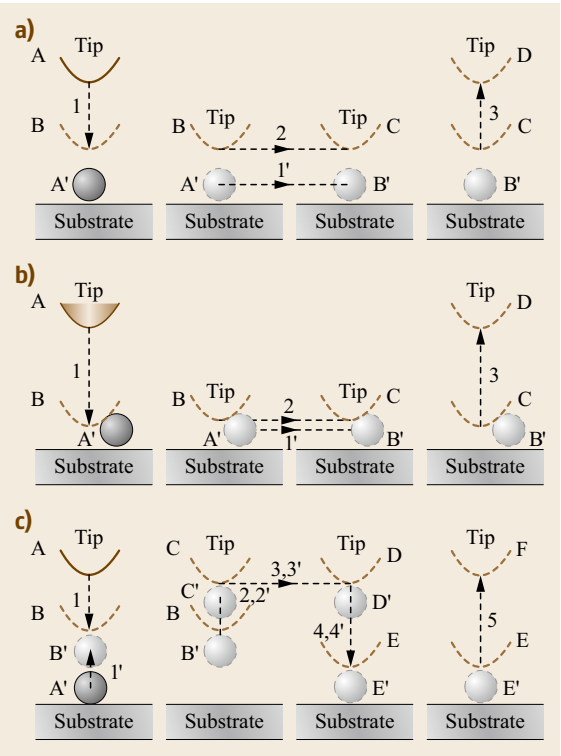


Fig. 18.4a–c Basic strategies of nanomanipulation. In the figure, A, B, C, ... represent the positions of end-effector (e.g., a tip), A′, B′, C′, ... the positions of objects, 1, 2, 3, ... the motions of end-effector, and 1′, 2′, 3′, ... the motions of objects. Tweezers can be used in pick-and-place to facilitate the picking-up, but are generally not necessarily helpful for placing. **(a)** Lateral non-contact nanomanipulation (sliding), **(b)** lateral contact nanomanipulation (pushing/pulling), **(c)** vertical nanomanipulation (picking and placing)

Non-contact manipulation combined with STMs has revealed many possible strategies for manipulating atoms and molecules. However, for the manipulation of CNTs no examples have been demonstrated.

Pushing or pulling nanometer objects on a surface with an AFM is a typical manipulation using this method as shown in Fig. 18.4b. Early work showed the effectiveness of this method for the manipulation of nanoparticles [18.46–49]. This method has also been shown in nanoconstruction [18.50] and biomanipulation [18.51]. A virtual-reality interface facilitates such manipulation [18.52, 53] and may create an opportunity for other types of manipulation. This technique has been used in the manipulation of nanotubes on a surface, and some examples will be introduced later in this chapter.

The pick-and-place task as shown in Fig. 18.4c is especially significant for 3-D nanomanipulation since its main purpose is to assemble prefabricated building blocks into devices. The main difficulty is in achieving sufficient control of the interaction between the tool and object and between the object and the substrate. Two strategies have been presented for micromanipulation [18.54] and have also proven to be effective for nanomanipulation [18.34, 55]. One strategy is to apply a dielectrophoretic force between a tool and an object as a controllable additional external force by applying a bias between the tool and the substrate on which the object is placed. Another strategy is to modify the van der Waals and other intermolecular and surface forces between the object and the substrate. For the former, an AFM cantilever is ideal as one electrode to generate a nonuniform electrical field between the cantilever and the substrate.

18.3.2 Nanorobotic Manipulation Systems

Nanorobotic manipulators are the core components of nanorobotic manipulation systems. The basic requirements for a nanorobotic manipulation system for 3-D manipulation include nanoscale positioning resolution, a relative large working space, enough DOFs including rotational ones for 3-D positioning and orientation

control of the end-effectors, and usually multiple end-effectors for complex operations.

A commercially available nanomanipulator (MM3A from Kleindiek) installed inside a SEM (Carl Zeiss DSM962) is shown in Fig. 18.5. The manipulator has three degrees of freedom, and nanometer to subnanometer-scale resolution (Table 18.3). Calculations show that, when moving/scanning in *A*/*B*-direction by joint q_1/q_2, the additional linear motion in *C* is very small. For example, when the arm length is 50 mm, the additional motion in the *C*-direction is only 0.25–1 nm when moving in the *A*-direction for 5–10 μm; these errors can be ignored or compensated with an additional motion of the prismatic joint p_3, which has a 0.25 nm resolution.

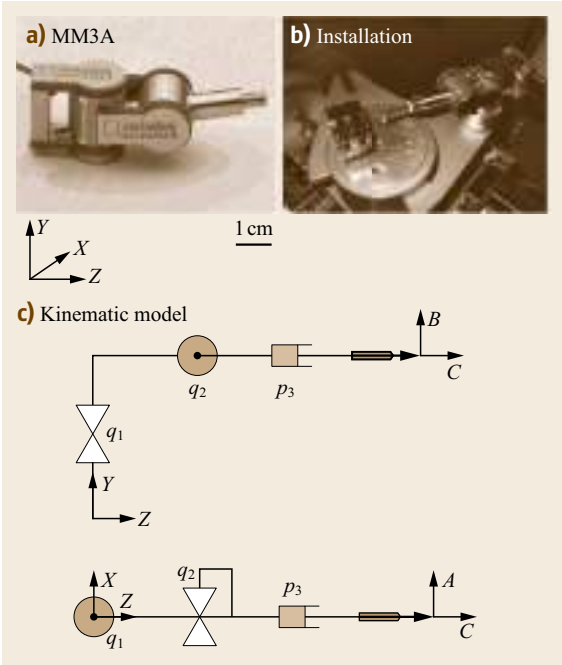


Fig. 18.5a–c Nanomanipulator (MM3A from Kleindiek) inside an SEM

Table 18.3 Specifications of MM3A

Item	Specification
Operating range q_1 and q_2	240°
Operating range *Z*	12 mm
Resolution *A* (horizontal)	10^{-7} rad (5 nm)
Resolution *B* (vertical)	10^{-7} rad (3.5 nm)
Resolution *C* (linear)	0.25 nm
Fine (scan) range *A*	20 μm
Fine (scan) range *B*	15 μm
Fine (scan) range *C*	1 μm
Speed *A*, *B*	10 mm/s
Speed *C*	2 mm/s

Figure 18.6a shows a nanorobotic manipulation system that has 16 DOFs in total and can be equipped with three to four AFM cantilevers as end-effectors for both manipulation and measurement. Table 18.4 lists the specifications of the system. Table 18.5 shows the functions of the nanorobotic manipulation system for nanomanipulation, nanoinstrumentation, nanofabrication and nanoassembly. The positioning resolution is subnanometer order and strokes are centimeter scale. The manipulation system is not only for nanomanipulation, but also for nanoassembly, nanoinstrumentation and nanofabrication. Four-probe semiconductor measurements are perhaps the most complex manipulation this system can perform, because it is necessary to actuate four probes independently by four manipulators. Theoretically, 24 DOFs are needed for four manipulators for general-purpose manipulations, i. e., 6 DOFs for each manipulator for complete control of three linear DOFs and three rotation DOFs. However, 16 DOFs are sufficient for this specific purpose. In general, two manipulators are sufficient for most tasks. More probes

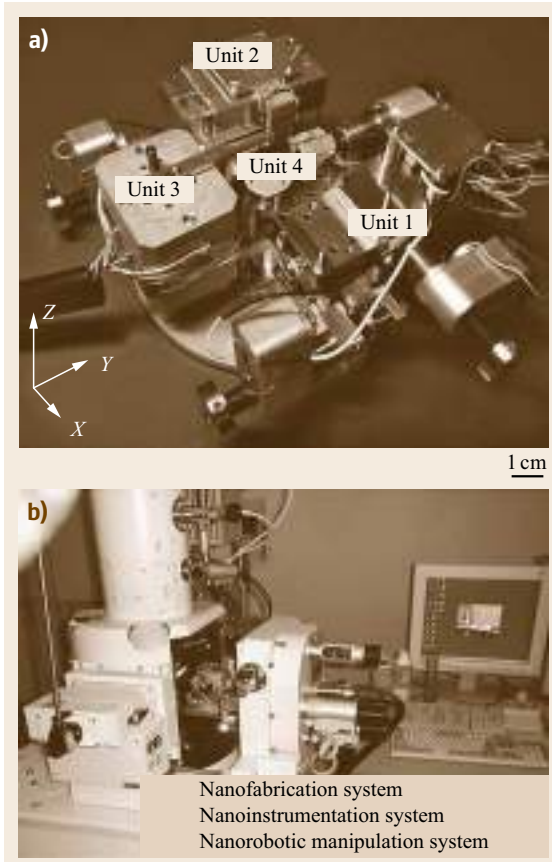


Fig. 18.6a,b Nanorobotic system. **(a)** Nanorobotic manipulators, **(b)** system setup

Table 18.4 Specifications of a nanorobotic manipulation system

Item	Specification
Nanorobotic manipulation system	
DOFs	Total: 16 DOFs
	Unit 1: 3 DOFs (x, y and β; coarse)
	Unit 2: 1 DOF (z; coarse), 3-DOF (x, y and z; fine)
	Unit 3: 6 DOFs (x, y, z, α, β, γ; ultrafine)
	Unit 4: 3 DOFs (z, α, β; fine)
Actuators	4 Picomotors (Units 1 & 2)
	9 PZTs (Units 2 & 3)
	7 Nanomotors (Units 2 & 4)
End-effectors	3 AFM cantilevers + 1 substrate or
	4 AFM cantilevers
Working space	18 mm × 18 mm × 12 mm × 360° (coarse, fine),
	26 μm × 22 μm × 35 μm (ultrafine)
Positioning resolution	30 nm (coarse), 2 mrad (coarse), 2 nm (fine), sub-nm (ultrafine)
Sensing system	FESEM (imaging resolution: nm) and AFM cantilevers
Nanoinstrumentation system	
FESEM	Imaging resolution: 1.5 nm
AFM cantilever	Stiffness constant: 0.03 nN/nm
Nanofabrication system	
EBID	FESEM emitter: T-FE
	CNT emitter

Table 18.5 Functions of a nanorobotic manipulation system

Functions	Manipulations involved
Nanomanipulation	Picking up nanotubes by controlling intermolecular and surface forces, and positioning them together in 3-D space
Nanoinstrumentation	Mechanical properties: buckling or stretching
	Electrical properties: placing between two probes (electrodes)
Nanofabrication	EBID with a CNT emitter and parallel EBID
	Destructive fabrication: breaking
	Shape modification: deforming by bending and buckling, and fixing with EBID
Nanoassembly	Connecting with van der Waals
	Soldering with EBID
	Bonding through mechanochemical synthesis

provide for more potential applications. For example, three manipulators can be used to assemble a nanotube transistor, a third probe can be applied to cut a tube supported on the other two probes, four probes can be used for four-terminal measurements to characterize the electric properties of a nanotube or a nanotube cross-junction. There are many potential applications for the manipulators if all four probes are used together. With the advancement of nanotechnology, one could shrink the size of nanomanipulators and insert more DOFs inside the limited vacuum chamber of a microscope, and, perhaps, the molecular version of manipulators such as that dreamed of by *Drexler* could be realized [18.4].

For the construction of multiwalled carbon nanotubes (MWNT)-based nanostructures, manipulators position and orient nanotubes for the fabrication of nanotube probes and emitters, for performing na-

nosoldering with electron-beam-induced deposition (EBID) [18.56], for the property characterization of single nanotubes for selection purposes and for characterizing junctions to test connection strength.

A nanolaboratory is shown in Fig. 18.6b, and its specifications are listed in Table 18.4. The nanolaboratory integrates a nanorobotic manipulation system with a nano analytical system and a nanofabrication system, and can be applied for manipulating nanomaterials, fabricating nanobuilding blocks, assembling nanodevices, and for in situ analysis of the properties of such materials, building blocks and devices. Nanorobotic manipulation within the nanolaboratory has opened a new path for constructing nanosystems in 3-D space, and will create opportunities for new nanoinstrumentation and nanofabrication processes.

18.4 Nanorobotic Assembly

18.4.1 Overview

Nanomanipulation is a promising strategy for nanoassembly. Key techniques for nanoassembly include the structuring and characterization of nanobuilding blocks, the positioning and orientation control of the building blocks with nanometer-scale resolution, and effective connection techniques. Nanorobotic manipulation, which is characterized by multiple DOFs with both position and orientation controls, independently actuated multi-probes, and a real-time observation system, have been shown to be effective for assembling nanotube-based devices in 3-D space.

The well-defined geometries, exceptional mechanical properties, and extraordinary electric characteristics, among other outstanding physical properties (as listed in Table 18.6), of CNTs [18.57] qualify them for many potential applications (as concisely listed in Table 18.7), especially in nanoelectronics [18.58–60], NEMS, and other nanodevices [18.61]. For NEMS, some of the most important characteristics of nanotubes include their nanometer diameter [18.62], large aspect ratio (10−1000) [18.63, 64], TPa-scale Young's modulus [18.65–71], excellent elasticity [18.33, 72], ultrasmall interlayer friction, excellent capability for field emission, various electric conductivities [18.73–75], high thermal conductivity [18.76], high current-carrying capability with essentially no heating [18.77, 78], sensitivity of conductance to various physical or chemical changes, and charge-induced bond-length change.

Helical 3-D nanostructures, or nanocoils, have been synthesized from various materials, including helical carbon nanotubes [18.79] and zinc oxide nanobelts [18.80]. A new method of creating structures with nanometer-scale dimensions has recently been presented [18.81] and can be fabricated in a controllable way [18.82]. The structures are created through a top-down fabrication process in which a strained nanometer-thick heteroepitaxial bilayer curls up to form 3-D structures with nanoscale features. Helical geometries and tubes with diameters between 10 nm and 10 μm have been achieved. Because of their interesting morphology, mechanical, electrical, and electromagnetic properties, potential applications of these nanostructures in NEMS include nanosprings [18.83], electromechanical sensors [18.84], magnetic-field detectors, chemical or biological sensors, generators of magnetic beams, inductors, actuators, and high-performance electromagnetic-wave absorbers.

Table 18.6 Properties of carbon nanotubes

Property	Item	Data
Geometrical	Layers	Single/multiple
	Aspect ratio	10−1000
	Diameter	≈ 0.4 nm to > 3 nm (SWNTs)
		≈ 1.1 to > 100 nm (MWNTs)
	Length	Several μm (rope up to cm)
Mechanical	Young's modulus	≈ 1 TPa (steel: 0.2 TPa)
	Tensile strength	45 GPa (steel: 2 GPa)
	Density	1.33−1.4 g/cm^3 (Al: 2.7 g/cm^3)
Electronic	Conductivity	Metallic/semiconductivity
	Current carrying	
	Capacity	≈ 1 TA/cm^3 (Cu: 1 GA/cm^3)
	Field emission	Activate phosphorus at 1−3 V
Thermal	Heat transmission	> 3 kW/(m K) (diamond: 2 kW/(m K))

Table 18.7 Applications of carbon nanotubes

State	Device	Main properties applied
Bulk/array	Composite	High strength, conductivity, etc.
	Field emission devices: flat display, lamp, gas discharge tube, x-ray source, microwave generator, etc.	Field emission: stable emission, long lifetimes, and low emission threshold potentials, high current densities
	Electrochemical devices: supercapacitor, battery cathode, electromechanical actuator, etc.	Large surface area conductivity, high strength, high reversible component of storage capacity
	Fuel cell, hydrogen storage, etc.	Large surface area
Individual	Nanoelectronics: wire, diode, transistor, switch memory, etc.	Small sizes, semiconducting/metallic
	NEMS: probe, tweezers, scissors, sensor, actuator, bearing, gear, etc.	Well-defined geometrics, exceptional mechanical and electronic properties

NEMS based on individual single- or multi-walled carbon nanotubes (SWNTs [18.85, 86] or MWNTs [18.57]) and nanocoils are of increasing interest, indicating that capabilities for incorporating these individual building blocks at specific locations on a device must be developed. Random spreading [18.87], direct growth [18.88], self-assembly [18.89], dielectrophoretic assembly [18.90] and nanomanipulation [18.91] have been demonstrated for positioning as-grown nanotubes on electrodes for the construction of these devices. However, for nanotube-based structures, nanorobotic assembly is still the only technique capable of in situ structuring, characterization and assembly. Because the as-fabricated nanocoils are not free-standing from their substrate, nanorobotic assembly is virtually the only way to incorporate them into devices at present.

18.4.2 Carbon Nanotubes

Nanotube manipulation in two dimensions on a surface was first performed with an AFM by contact pushing on a substrate. Figure 18.7 shows the typical methods for 2-D pushing. Although similar to that shown in Fig. 18.4b, the same manipulation caused various results because nanotubes cannot be regarded as 0-D points. The first demonstration was given by *Lieber* and coworkers for measuring the mechanical properties of a nanotube [18.66]. They used the method shown in Fig. 18.7b, i. e., bending a nanotube by pushing on one end of it and fixing the other end. The same strategy was used for the investigation of the behaviors of nanotubes under large strain [18.92]. *Dekker* and coworkers applied the strategies shown in Fig. 18.7c,d to get a kinked junction and crossed nanotubes [18.93, 94]. *Avouris* and coworkers combined this technique with an inverse process, namely straightening by pushing along a bent tube, and realized the translation of the tube to another location [18.95] and between two electrodes to measure the conductivity [18.96], and to form a field-effect transistor (FET) [18.97]. This technique was also used to place a tube on another tube to form a single electron transistor (SET) with cross-junction of nanotubes [18.98]. Pushing-induced breaking (Fig. 18.7d) has also been demonstrated for an adsorbed nanotube [18.95] and a freely suspended SWNT rope [18.72]. The simple assembly of two bent tubes and a straight one formed the Greek letter θ.

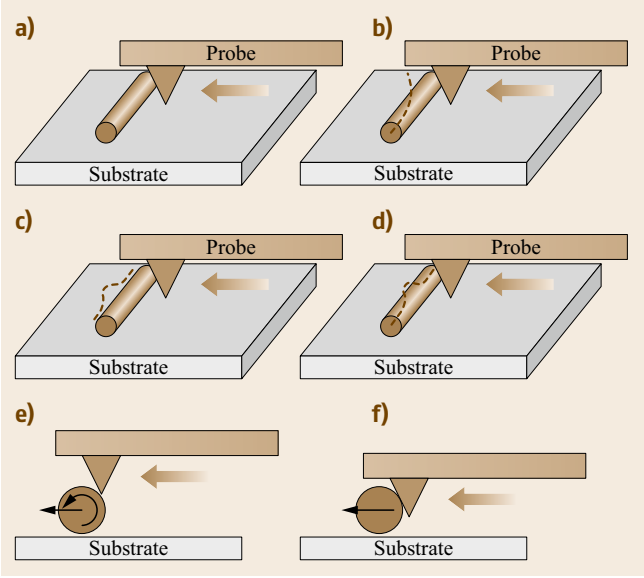


Fig. 18.7a–f 2-D manipulation of CNTs. Starting from the original state shown in **(a)**, pushing the tube at a different site with a different force may cause the tube to deform as in **(b)** and **(c)**, to break as in **(d)**, or to move as in **(e)** and **(f)**. **(a)** Original state, **(b)** bending, **(c)** kinking, **(d)** breaking, **(e)** rolling, **(f)** sliding

To investigate the dynamics of rolling at the atomic level, rolling and sliding of a nanotube (as shown in Fig. 18.7e,f) are performed on graphite surfaces using an AFM [18.99, 100].

Manipulation of CNTs in 3-D space is important for assembling CNTs into structures and devices. The basic techniques for the nanorobotic manipulation of carbon nanotubes are shown in Fig. 18.8 [18.101]. These serve as the basis for handling, structuring, characterizing and assembling NEMS.

The basic procedure is to pick up a single tube from nanotube soot, Fig. 18.8a. This has been shown first by using dielectrophoresis [18.34] through nanorobotic manipulation (Fig. 18.8b). By applying a bias between a sharp tip and a plane substrate, a nonuniform electric field can be generated between the tip and the substrate with the strongest field near the tip. This field can cause a tube to orient along the field or further *jump* to the tip by electrophoresis or dielectrophoresis (determined by the conductivity of the objective tubes). Removing the bias, the tube can be placed at other locations at will. This method can be used for free-standing tubes on nanotube soot or on a rough surface on which surface van der Waals forces are generally weak. A tube strongly rooted in CNT soot or lying on a flat surface cannot be picked up in this way. The interaction be-

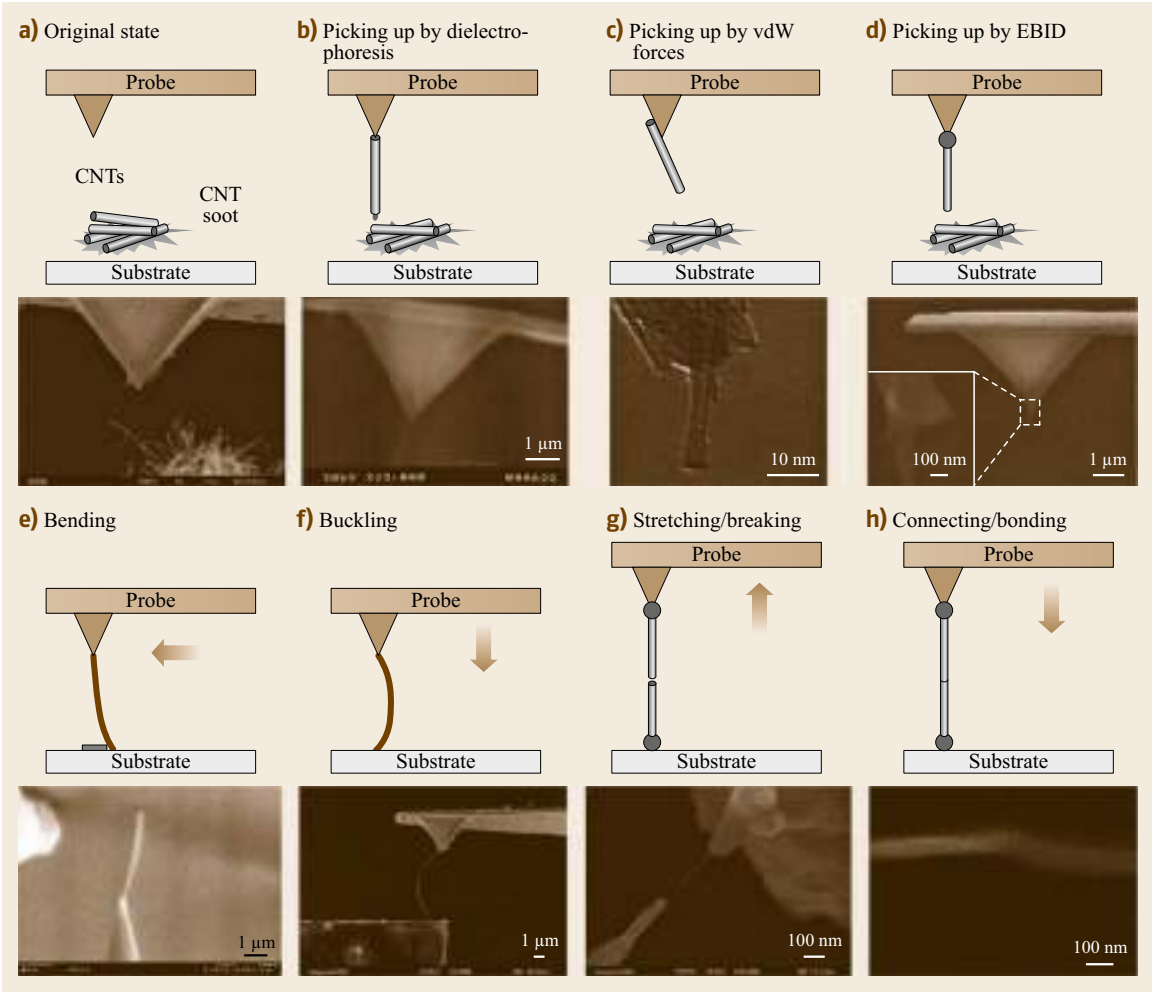


Fig. 18.8a–h Nanorobotic manipulation of CNTs. The basic technique is to pick up an individual tube from CNT soot **(a)** or from an oriented array; **(b)** shows a free-standing nanotube picked up by dielectrophoresis generated by a nonuniform electric field between the probe and substrate, **(c)** (after [18.102]) and **(d)** show the same manipulation by contacting a tube with the probe surface or fixing (e.g., with EBID) a tube to the tip (*inset* shows the EBID deposit). Vertical manipulation of nanotubes includes bending **(e)**, buckling **(f)**, stretching/breaking **(g)**, and connecting/bonding **(h)**. All examples with the exception of **(c)** are from the authors' work

tween a tube and the atomically flat surface of the AFM cantilever tip has been shown to be strong enough to pick up a tube with the tip [18.102] (Fig. 18.8c). By using EBID, it is possible to pick up and fix a nanotube onto a probe [18.103] (Fig. 18.8d). For handling a tube, a weak connection between the tube and the probe is desired.

Bending and buckling a CNT as shown in Fig. 18.8e,f are important for in situ property characterization of a nanotube [18.104, 105], which is a simple way to get the Young's modulus of a nanotube without damaging the tube (if performed within its elastic range) and, hence, can be used for the selection of a tube with desired properties. The process is shown in Fig. 18.9a. The left figure shows an individual MWNT, whereas the right four show a bundle of MWNTs being buckled. Figure 18.9b depicts the property curve of the elastic and plastic deformations of the MWNT (∅133 nm × 6.055 µm), where d and F are the axial deformation and buckling force, as shown in Fig. 18.9a (top). By using the model and analysis method presented in [18.55], the flexural rigidity of the MWNT bundle shown in Fig. 18.9a (bottom) is found to be $EI = 2.086 \times 10^{-19}$ [N m^2]. This result suggests that the diameter of the nanotube is 46.4 nm if the theoretical value of the Young's modulus of the nanotube $E = 1.26$ TPa is used. The SEM image shows that the bundle of nanotubes includes at least three single ones with diameters of 31, 34, and 41 nm. Hence, it can be determined that there must be damaged parts in the bundle because the stiffness is too low, and it is necessary to select another one without defects. By buckling an MWNT over its elastic limit, a kinked structure can be obtained. After three loading/releasing rounds, as shown in Fig. 18.9b, a kinked structure is obtained, as shown in Fig. 18.10a [18.106]. To obtain any desired angle for a kinked junction it is possible to fix the shape of a buckled nanotube within its elastic limit by using EBID (Fig. 18.10b) [18.107]. For a CNT, the maximum angular displacement will appear at the fixed left end under pure bending or at the middle point under pure buckling. A combination of these two kinds of loads will achieve a controllable position of the kink point and a desired kink angle θ. If the deformation is within the elastic limit of the nanotube, it will recover as the load is released. To avoid this, EBID can be applied at the kink point to fix the shape.

Stretching a nanotube between two probes or a probe and a substrate has generated several interesting results (Fig. 18.8g). The first demonstration of 3-D nanomanipulation of nanotubes took this as an example to show the breaking mechanism of an MWNT [18.33], and to measure the tensile strength of CNTs [18.71]. By breaking an MWNT in a controlled manner, interesting nanodevices have been fabricated. This technique – destructive fabrication – has been presented to get sharpened and layered structures of nanotubes and to improve control of the length of nanotubes [18.108]. Typically, a layered and a sharpened structure can be obtained from this process, similar to that achieved from electric pulses [18.109]. Bearing motion has also been observed in an incompletely broken MWNT (Fig. 18.11). As shown in Fig. 18.11a, an MWNT is supported between a substrate (left end) and an AFM cantilever (right end). Figure 18.11b shows a zoomed image of the centrally blocked part of Fig. 18.11a, and the inset shows its structure schematically. It can be found that the nanotube has a thinner neck (part B in Fig. 18.11b) that was formed by destructive fabrication, i. e., by moving the cantilever to the right. To move it further in the same direction, a motion like a linear bearing is observed, as shown in Fig. 18.11c and schematically by the inset. By comparing Fig. 18.11b,c, we find that part B remained

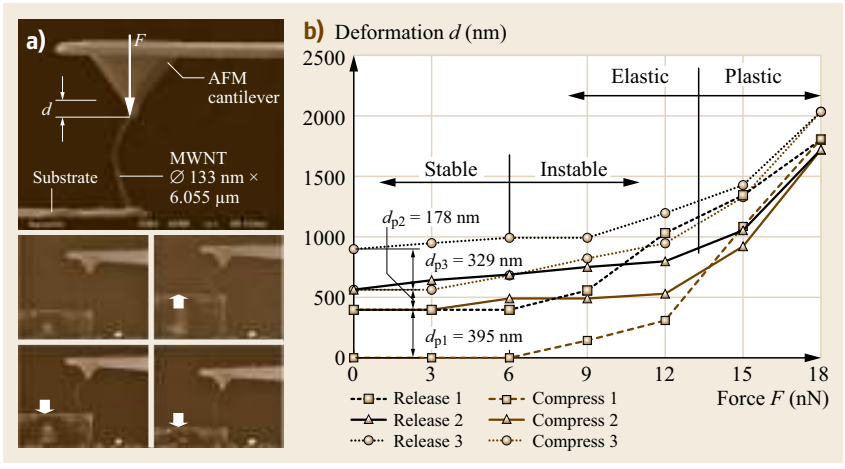


Fig. 18.9a,b In-situ mechanical property characterization of a nanotube by buckling it (scale bars: 1 µm). **(a)** Process of buckling, **(b)** elastic and plastic properties of an MWNT

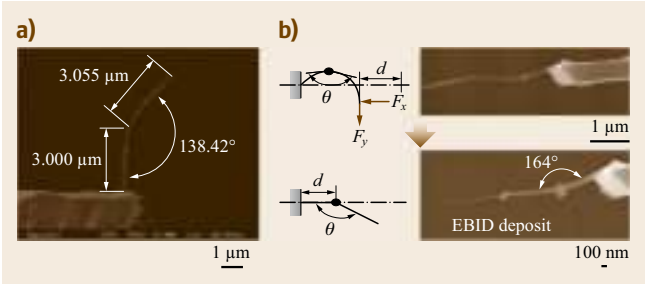


Fig. 18.10a,b Shape modification. **(a)** Kinked structure of a MWNT through plastic deformation, **(b)** shape modifications of a MWNT by elastic bending and buckling deformation and shape fixing through EBID

unchanged in its length and diameter, while its two ends brought out two new parts I and II from parts A and B, respectively. Part II has uniform diameter (∅ 22 nm), while part I is a tapered structure with a smallest diameter of ∅ 25 nm. The interlayer friction has been predicted to be very small [18.110], but direct measurement of the friction remains a challenging problem.

The reverse process, namely the connection of broken tubes (Fig. 18.8h), has been demonstrated recently, and the mechanism is revealed as rebonding of unclosed dangling bonds at the ends of broken tubes [18.111]. Based on this interesting phenomenon, mechanochemical nanorobotic assembly has been performed [18.112].

Assembly of nanotubes is a fundamental technology for enabling nanodevices. The most important tasks include the connection of nanotubes and placing of nanotubes onto electrodes. Pure nanotube circuits [18.113–115] created by interconnecting nanotubes of different diameters and chirality could lead to further size reductions in devices. Nanotube intermolecular and intramolecular junctions are basic elements for such systems. An intramolecular kink junction behaving like a rectifying diode has been reported [18.116]. Room-temperature (RT) SETs [18.117] have been shown with a short (≈ 20 nm) nanotube section that is created by inducing local barriers into the tube with an AFM, and Coulomb charging has been observed. With a cross-junction of two SWNTs (semiconducting/metallic), three- and four-terminal electronic devices have been

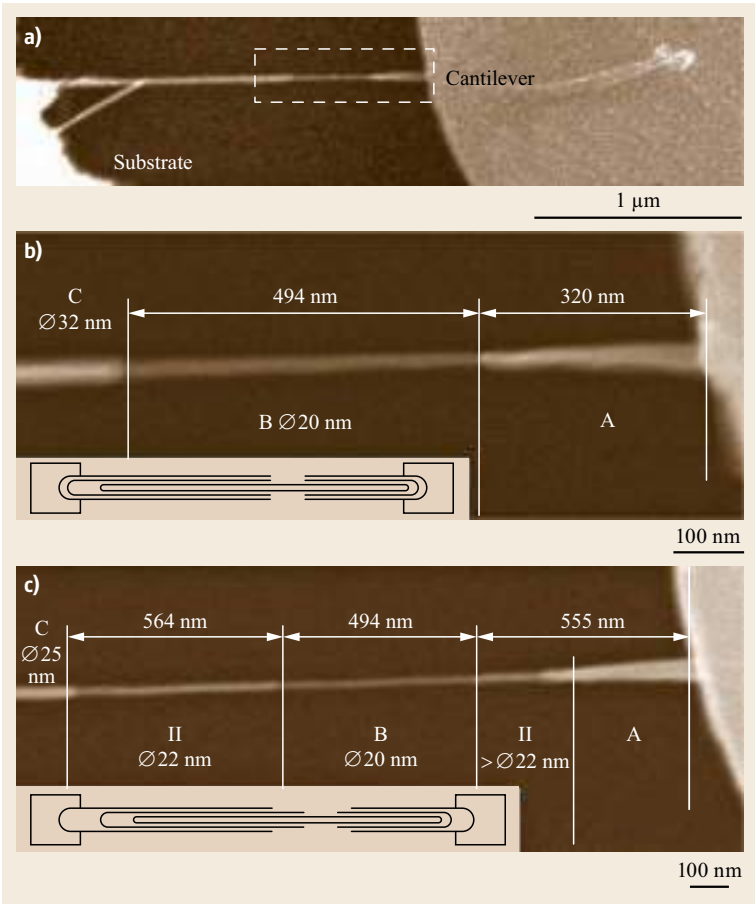


Fig. 18.11a–c Destructive fabrication of a MWNT and its bearing-like motion

made [18.118]. A suspended cross-junction can function as electromechanical nonvolatile memory [18.119].

Although some kinds of junctions have been synthesized with chemical methods, there is no evidence yet showing that a self-assembly-based approach can provide more complex structures. SPMs were also used to fabricate junctions, but they are limited to a 2-D plane. We have presented 3-D nanorobotic manipulation-based nanoassembly, which is a promising strategy, both for the fabrication of nanotube junctions and for the construction of more complex nanodevices with such junctions as well.

Nanotube junctions can be classified into different types by: the kind of components – SWNTs or MWNTs; geometric configuration – V (kink), I, X- (cross), T-, Y- (branch), and 3-D junctions; conductivity – metallic or semiconducting; and connection methods – intermolecular (connected with van der Waals force, EBID, etc.) or intramolecular (connected with chemical bonds) junctions. Here we show the fabrication of several kinds of MWNT junctions by emphasizing the connection methods. These methods will also be effective for SWNT junctions. Figure 18.12 shows CNT junctions constructed by connecting with van der Waals forces (a), joining by electron-beam-induced deposition (b), and bonding through mechanochemistry (c).

MWNT junctions connected with van der Waals forces are the basic forms of junctions. To fabricate such junctions, the main process is to position two or more nanotubes together with nanometer resolution; they will then be connected naturally by intermolecular van der Waals forces. Such junctions are mainly for structures where contact rather than strength is emphasized. Placing them onto a surface can make them more stable. In some cases, when lateral movement along the surface of nanotubes is desired while keeping them in contact, van der Waals-type connections are the only ones that are suitable.

Figure 18.12a shows a T-junction connected with van der Waals forces, which is fabricated by positioning the tip of an MWNT onto another MWNT until they form a bond. The contact is checked by measuring the shear connection force.

EBID provides a soldering method to obtain stronger nanotube junctions than those connected through van der Waals forces. Hence, if the strength of nanostructures is emphasized, EBID can be applied. Figure 18.12b shows an MWNT junction connected through EBID, in which the upper MWNT is a single one with a diameter of 20 nm and the lower one is a bundle of MWNTs with an extruded single CNT with ∅30 nm. The development of conventional EBID has been limited by the expensive electron filament used and low productivity. We have presented a parallel EBID system by using CNTs as emitters because of their excellent field-emission properties [18.120, 121]. The feasibility of parallel EBID is presented. It is a promising strategy for large-scale fabrications of nanotube junctions. Similar to its macro counterpart, welding, EBID works by adding material to obtain stronger connections, but in some cases, added material might influence normal functions for nanosystems. So, EBID is mainly applied to nanostructures rather than nanomechanisms.

To construct stronger junctions without adding additional material, mechanochemical nanorobotic assembly is an important strategy. Mechanochemical nanorobotic assembly is based on solid-phase chemical reactions, or mechanosynthesis, which is defined as chemical synthesis controlled by mechanical systems operating with atomic-scale precision, enabling direct positional selection of reaction sites [18.4]. By picking up atoms with dangling bonds rather than natural atoms only, it is easier to form primary bonds, which provides a simple but strong connection. Destructive fabrication provides a way to form dangling bonds at the ends of broken tubes. Some of the dangling bonds may close with neighboring atoms, but generally a few bonds will remain dangling. A nanotube with dangling bonds at its end will bind more easily to another

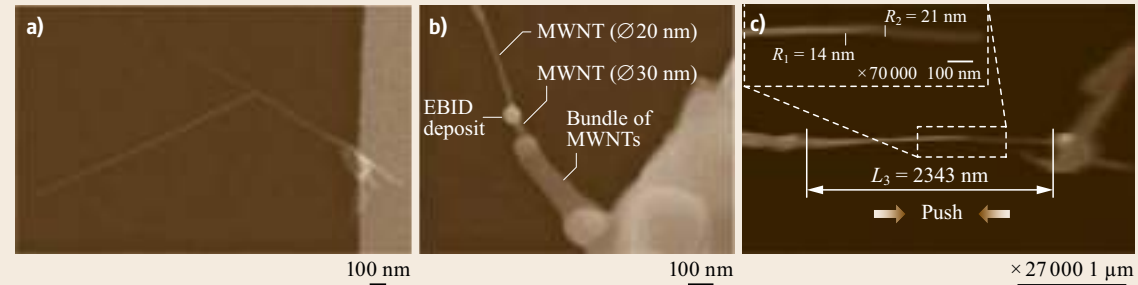


Fig. 18.12a–c MWNT junctions. **(a)** MWNTs connected with van der Waals force. **(b)** MWNTs joined with EBID. **(c)** MWNTs bonded with a mechanochemical reaction

to form intramolecular junctions. Figure 18.12c shows such a junction. An MWNT (length $L_1 = 1329$ nm, diameter $D_1 = 42$ nm) is placed between a substrate and an AFM cantilever with a CNT tip, and the two ends are fixed. By pulling the two ends of the MWNT, it is broken into two parts. By pushing the two nanotubes head to head close enough, a new one is formed. To test the strength of this nanotube, it was broken again. The fact that the nanotube breaks at a different site suggests that the tensile strength of the connected nanotubes is not weaker than that of the original nanotube itself. We have determined that no type of connection based on van der Waals interactions can provide such a strong connection strength [18.112]. Also, we have shown that, from the measured tensile strength (1.3 TPa), chemical bonds must have been formed when the junction formed, and that these are most likely to be covalent bonds (sp^2-hybrid type, as in a nanotube).

3-D nanorobotic manipulation has opened a new route for structuring and assembly nanotubes into nanodevices. However, at present nanomanipulation is still performed in a serial manner with master–slave control, which is not a large-scale production-oriented technique. Nevertheless, with advances in the exploration of mesoscopic physics, better control of the synthesis of nanotubes, more accurate actuators, and effective tools for manipulation, high-speed and automatic nanoassembly will be possible. Another approach might be parallel assembly by positioning building blocks with an array of probes [18.122] and joining them together simultaneously, e.g., with the parallel EBID [18.103] approach we presented. Further steps might progress towards exponential assembly [18.123], and in the far future to self-replicating assembly [18.4].

18.4.3 Nanocoils

The construction of nanocoil-based NEMS involves the assembly of as-grown or as-fabricated nanocoils, which is a significant challenge from a fabrication standpoint. Focusing on the unique aspects of manipulating nanocoils due to their helical geometry, high elasticity, single end fixation, and strong adhesion to the substrate from wet etching, a series of new processes is presented using a manipulator (MM3A, Kleindiek) installed in an SEM (Zeiss DSM962). As-fabricated SiGe/Si bilayer nanocoils are shown in Fig. 18.13. Special tools have been fabricated including a nanohook prepared by controlled *tip-crashing* of a commercially available tungsten sharp probe (Picoprobe T-4-10-1 mm and T-4-10) onto a substrate, and a *sticky* probe prepared by tip dipping into a double-sided SEM silver conductive tape (Ted Pella, Inc.). As shown in Fig. 18.14, experiments demonstrate that nanocoils can be released from a chip by

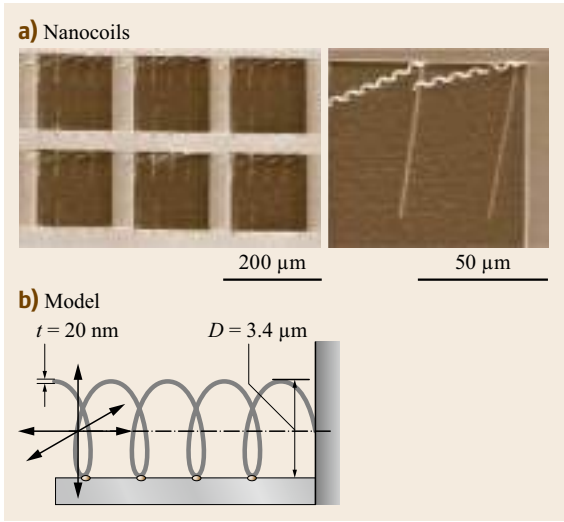


Fig. 18.13a,b As-fabricated nanocoils with a thickness of $t = 20$ nm (without Cr layer) or 41 nm (with Cr layer). Diameter: $D = 3.4\,\mu$m

lateral pushing, picked up with a nanohook or a *sticky* probe, and placed between the probe/hook and another probe or an AFM cantilever (Nanoprobe, NP-S). Axial pulling/pushing, radial compressing/releasing, and bending/buckling have also been demonstrated. These processes have shown the effectiveness of manipulation for the characterization of coil-shaped nanostructures and their assembly for NEMS, which have been otherwise unavailable.

Configurations of nanodevices based on individual nanocoils are shown in Fig. 18.15. Cantilevered nanocoils as shown in Fig. 18.15a can serve as nanosprings. Nanoelectromagnets, chemical sensors and nanoinductors involve nanocoils bridged between two electrodes, as shown in Fig. 18.15b. Electromechanical sensors can use a similar configuration but with one end connected to a moveable electrode, as shown in Fig. 18.15c. Mechanical stiffness and electric conductivity are fundamental properties for these devices that must be further investigated.

As shown in Fig. 18.14h, axial pulling is used to measure the stiffness of a nanocoil. A series of SEM images are analyzed to extract the AFM tip displacement and the nanospring deformation, i. e., the relative displacement of the probe from the AFM tip. From this displacement data and the known stiffness of the AFM cantilever, the tensile force acting on the nanospring versus the nanospring deformation was plotted. The deformation of the nanospring was measured relative to the first measurement point. This was necessary because proper attachment of the nanospring to the AFM cantilever must be verified. Afterwards, it was not pos-

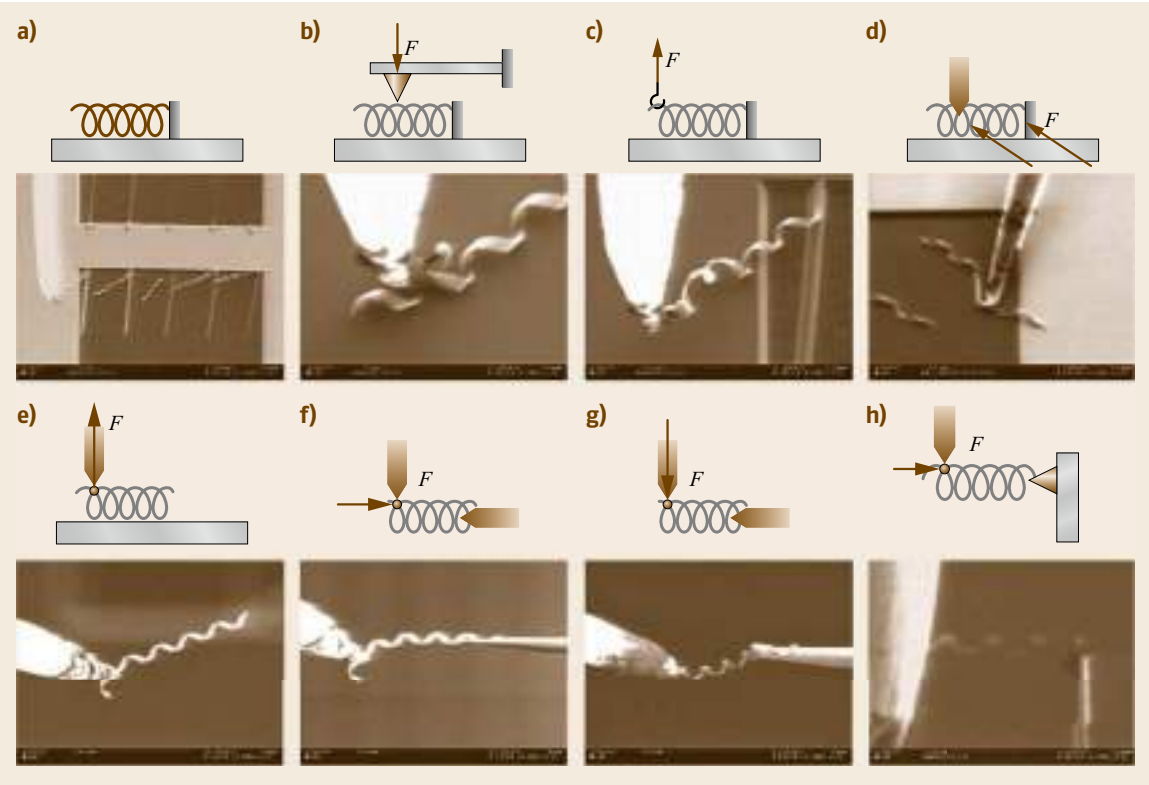


Fig. 18.14a–h Nanorobotic manipulation of nanocoils: **(a)** original state, **(b)** compressing/releasing, **(c)** hooking, **(d)** lateral pushing/breaking, **(e)** picking, **(f)** placing/inserting, **(g)** bending, and **(h)** pushing and pulling

sible to return to the point of zero deformation. Instead, the experimental data, as presented in Fig. 18.15d, has been shifted such that, with the calculated linear elastic spring stiffness, the line begins at zero force and zero deformation. From Fig. 18.15d, the stiffness of the spring was estimated to be 0.0233 N/m. The linear elastic region of the nanospring extends to a deformation of 4.5 μm. An exponential approximation was fitted to the nonlinear region. When the applied force reached 0.176 μN, the attachment between the nanospring and the AFM cantilever broke. Finite element simulation (ANSYS 9.0) was used to validate the experimental data [18.84]. Since the exact region of attachment cannot be identified according to the SEM images, simulations were conducted for 4, 4.5, and 5 turns to get an estimate of the possible range, given that the apparent number of turns of the nanospring is between 4 and 5. The nanosprings in the simulations were fixed on one end and had an axial load of 0.106 μN applied to the other end. The simulation results for the spring with 4 turns yield a stiffness of 0.0302 N/m; for the nanospring with 5 turns it is 0.0191 N/m. The measured stiffness falls into this range at 22.0% above the minimum value and 22.8% below the maximum value, and very close to the stiffness of a 4.5 turn nanospring, which has a stiffness of 0.0230 N/m according to the simulation.

Figure 18.15e shows the results from electrical characterization experiments on a nanospring with 11 turns using the configuration as shown in Fig. 18.14g. The *I–V* curve is nonlinear, which may be caused by the resistance change of the semiconductive bilayer due to ohmic heating. Another possible reason is the decrease in contact resistance caused by thermal stress. The maximum current was found to be 0.159 mA under an 8.8 V bias. Higher voltage causes the nanospring to *blow off*. From the fast scanning screen of the SEM, the extension of the nanospring on the probes was observed around the peak current so that the current does not drop abruptly. At 9.4 V, the extended nanospring is broken down, causing an abrupt drop in the *I–V* curve.

From fabrication and characterization results, the helical nanostructures appear to be suitable to function as inductors. They would allow further miniaturization compared to state-of-the-art microinductors. For this purpose, higher doping of the bilayer and an additional metal layer would result in the required conductance. Conductance, inductance, and quality factor can be

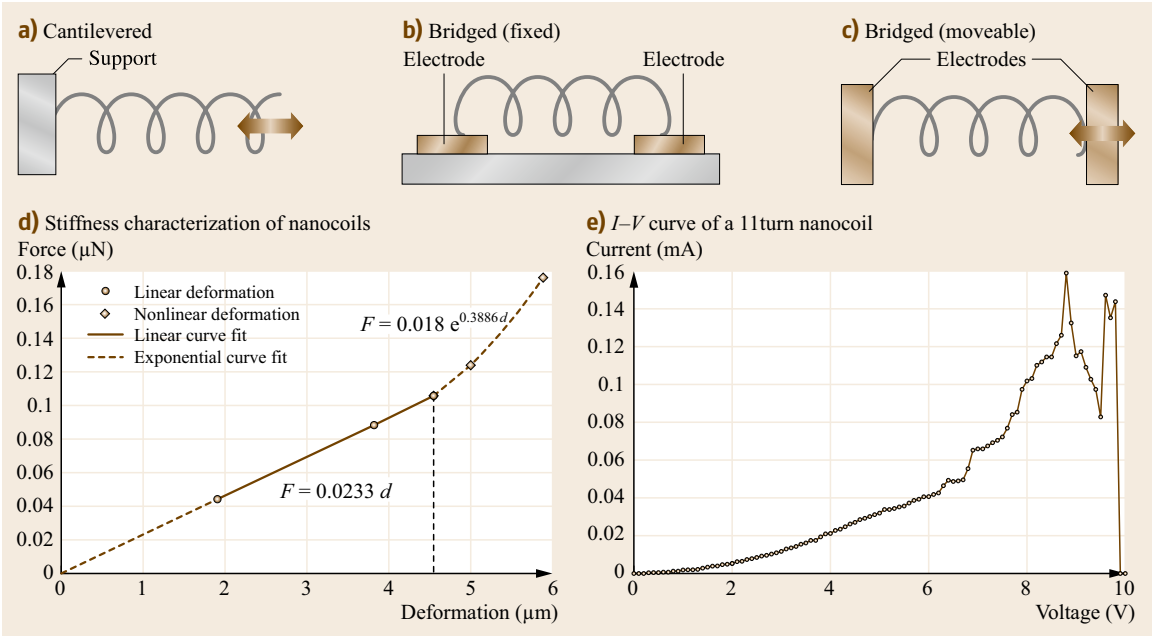


Fig. 18.15a–e Nanocoil-based devices. Cantilevered nanocoils **(a)** can serve as nanosprings. Nanoelectromagnets, chemical sensors, and nanoinductors involve nanocoils bridged between two electrodes **(b)**. Electromechanical sensors can use a similar configuration but with one end connected to a moveable electrode **(c)**. Mechanical stiffness **(d)** and electric conductivity **(e)** are basic properties of interest for these devices

further improved if, after curling up, additional metal is electroplated onto the helical structures. Moreover, a semiconductive helical structure, when functionalized with binding molecules, can be used for chemical sensing using the same principle as demonstrated with other types of nanostructures [18.124]. With bilayers in the range of a few monolayers, the resulting structures would exhibit a very high surface-to-volume ratio with the whole surface exposed to an incoming analyst.

18.5 Applications

Material science, biotechnology, electronics, and mechanical sensing and actuation will benefit from advances in nanorobotics. Research topics in bionanorobotics include the autonomous manipulation of single cells or molecules, the characterization of biomembrane mechanical properties using nanorobotic systems with integrated vision and force-sensing modules, and more. The objective is to obtain a fundamental understanding of single-cell biological systems and provide characterized mechanical models of biomembranes for deformable cell tracking during biomanipulation and cell injury studies. Robotic manipulation at nanometer scales is a promising technology for structuring, characterizing and assembling nanobuilding blocks into NEMS. Combined with recently developed nanofabrication processes, a hybrid approach is realized to build NEMS and other nanorobotic devices from individual carbon nanotubes and SiGe/Si nanocoils.

18.5.1 Robotic Biomanipulation

Biomanipulation – Autonomous Robotic Pronuclei DNA Injection

To improve the low success rate of manual operation, and to eliminate contamination, an autonomous robotic system (shown in Fig. 18.16) has been developed to deposit DNA into one of the two nuclei of a mouse embryo without inducing cell lysis [18.125, 126]. The laboratory's experimental results show that the success rate for the autonomous embryo pronuclei DNA injection is dramatically improved over manual conventional injection methods. The autonomous robotic system features a hybrid controller that combines visual servoing

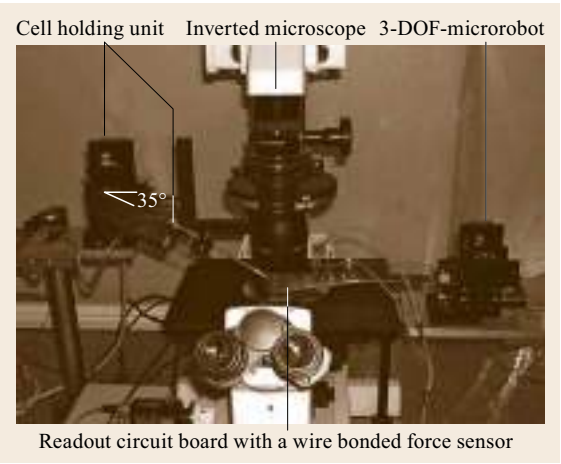


Fig. 18.16 Robotic biomanipulation system with vision and force feedback

and precision position control, pattern recognition for detecting nuclei, and a precise autofocusing scheme. Figure 18.17 illustrates the injection process.

To realize large-scale injection operations, a MEMS cell holder was fabricated using anodic wafer-bonding techniques. Arrays of holes are aligned on the cell holder, which are used to contain and fix individual cells for injection. When well-calibrated, the system with the cell holder makes it possible to inject large numbers of cells using position control. The cell-injection operation can be conducted in a move–inject–move manner.

Successful injection is determined greatly by injection speed and trajectory, and the forces applied to cells (Figure 18.17). To further improve the robotic system's performance, a multi-axial MEMS-based capacitive cellular force sensor is being designed and fabricated to provide real-time force feedback to the robotic system. The MEMS cellular force sensor also aids our research in biomembrane mechanical property characterization.

MEMS-Based Multi-Axis Capacitive Cellular Force Sensor

The MEMS-based two-axis cellular force sensor [18.127] shown in Fig. 18.18 is capable of resolving normal forces applied to a cell as well as tangential forces generated by improperly aligned cell probes. A high-yield microfabrication process was developed to form the 3-D high-aspect-ratio structure using deep reactive ion etching (DRIE) on silicon-on-insulator (SOI) wafers. The constrained outer frame and the inner movable structure are connected by four curved springs. A load applied to the probe causes the inner structure to move, changing the gap between each pair of interdigitated comb capacitors. Consequently, the total capacitance change resolves the applied force. The interdigitated capacitors are orthogonally configured to

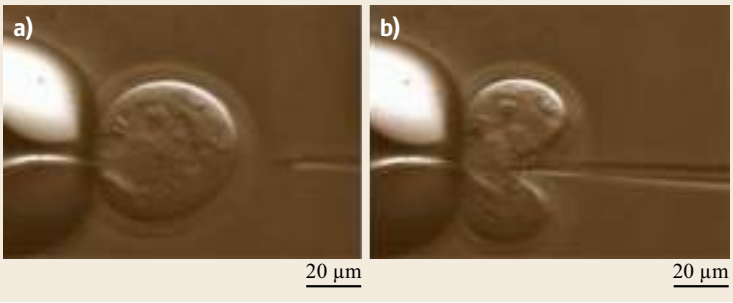


Fig. 18.17a,b Cell injection process. **(a,b)** Cell injection of a mouse oocyte

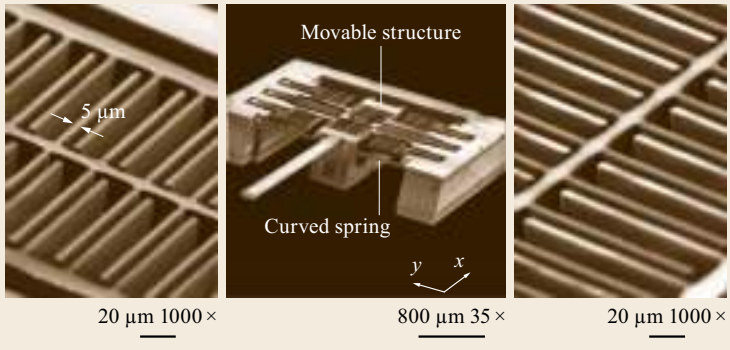


Fig. 18.18 A cellular force sensor with orthogonal comb drives detailed

make the force sensor capable of resolving forces in both the x- and y-directions. The cellular force sensors used in the experiments are capable of resolving forces up to 25 μN with a resolution of 0.01 μN.

Tip geometry affects the quantitative force measurement results. A standard injection pipette (Cook K-MPIP-1000-5) tip section with a tip diameter of 5 μm is attached to the probe of the cellular force sensors.

The robotic system and high-sensitivity cellular force sensor are also applied to biomembrane mechanical property studies [18.128]. The goal is to obtain a general parameterized model describing cell membrane deformation behavior when an external load is applied. This parameterized model serves two chief purposes. First, in robotic biomanipulation, it allows online parameter recognition so that cell membrane deformation behavior can be predicted. Second, for a thermodynamic model of membrane damage in cell injury and recovery studies, it is important to appreciate the mechanical behavior of the membranes. This allows the interpretation of such reported phenomena as mechanical resistance to cellular volume reduction during dehydration, and its relationship to injury. The establishment of such a biomembrane model will greatly facilitate cell injury studies.

Experiments demonstrate that robotics and MEMS technology can play important roles in biological studies such as automating biomanipulation tasks. Aided by robotics, the integration of vision and force-sensing modules, and MEMS design and fabrication techniques, investigations are being conducted in biomembrane mechanical property modeling, deformable cell tracking, and single-cell and biomolecule manipulation.

18.5.2 Nanorobotic Devices

Nanorobotic devices involve tools, sensors, and actuators at the nanometer scale. Shrinking device size makes it possible to manipulate nanosized objects with nanosized tools, measure mass in femtogram ranges, sense force at piconewton scales, and induce GHz motion, among other amazing advancements.

Top-down and bottom-up strategies for manufacturing such nanodevices have been independently investigated by a variety of researchers. Top-down strategies are based on nanofabrication and include technologies such as nanolithography, nanoimprinting, and chemical etching. Presently, these are 2-D fabrication processes with relatively low resolution. Bottom-up strategies are assembly-based techniques. At present, these strategies include such techniques as self-assembly, dip-pen lithography, and directed self-assembly. These techniques can generate regular nanopatterns at large scales. With the ability to position and orient nanometer-scale objects, nanorobotic manipulation is an enabling technology for structuring, characterizing and assembling many types of nanosystems [18.101]. By combining bottom-up and top-down processes, a hybrid nanorobotic approach (as shown in Fig. 18.19) based on nanorobotic manipulation provides a third way to fabricate NEMS by structuring as-grown nanomaterials or nanostructures. This new nanomanufacturing technique can be used to create complex 3-D nanodevices with such building blocks. Nanomaterial science, bionanotechnology, and nanoelectronics will also benefit from advances in nanorobotic assembly.

The configurations of nanotools, sensors, and actuators based on individual nanotubes that have been experimentally demonstrated are summarized as shown in Fig. 18.20.

For detecting deep and narrow features on a surface, cantilevered nanotubes (Fig. 18.20a, [18.103]) have been demonstrated as probe tips for an AFM [18.129], an STM and other types of SPM. Nanotubes provide ultrasmall diameters, ultralarge aspect ratios, and excellent mechanical properties. Manual assembly, direct growth [18.130] and nanoassembly [18.131] have proven effective for their construction. Cantilevered nanotubes have also been demonstrated as probes for the measurement of ultrasmall physical quantities, such as femtogram masses [18.67], mass flow sensors [18.132], and piconewton-order force sensors [18.133] on the basis of their static deflections or change of resonant frequencies detected within an electron microscope. Deflections cannot be measured from micrographs in real time, which limits the application of this kind of sensor. Interelectrode distance changes cause emission current variation of a nanotube emitter and may serve as a candidate to replace microscope images.

Bridged individual nanotubes (Fig. 18.20b, [18.134]) have been the basis for electric characterization. A nanotube-based gas sensor adopted this configuration [18.135].

Opened nanotubes (Fig. 18.20c, [18.136]) can serve as an atomic or molecular container. A thermometer based on this structure has been shown by monitoring the height of the gallium inside the nanotube using TEM [18.137].

Bulk nanotubes can be used to fabricate actuators based on charge-injection-induced bond-length change [18.138], and, theoretically, individual nanotubes also work on the same principle. Electrostatic deflection of a nanotube has been used to construct a relay [18.139]. A new family of nanotube actuators can be constructed by taking advantage of the ultralow inter-layer friction of a multiwalled nanotube. Linear bearings based on telescoping nanotubes have been demonstrated [18.110]. Recently, a micro actu-

ator with a nanotube as a rotation bearing has been demonstrated [18.140]. A preliminary experiment on a promising nanotube linear motor with field emission current serving as position feedback has been shown with nanorobotic manipulation (Fig. 18.20d, [18.136]).

Cantilevered dual nanotubes have been demonstrated as nanotweezers [18.141] and nanoscissors (Fig. 18.20e, [18.91]) by manual and nanorobotic assembly, respectively.

Based on electric resistance change under different temperatures, nanotube thermal probes (Fig. 18.20f) have been demonstrated for measuring the temperature at precise locations. These thermal probes are more advantageous than nanotube-based thermometers because the thermometers require TEM imaging. The probes also have better reproducibility than devices based on dielectrophoretically assembled bulk nanotubes [18.142]. Gas sensors and hot-wire-based mass-flow sensors can also be constructed in this configuration rather than a bridged one.

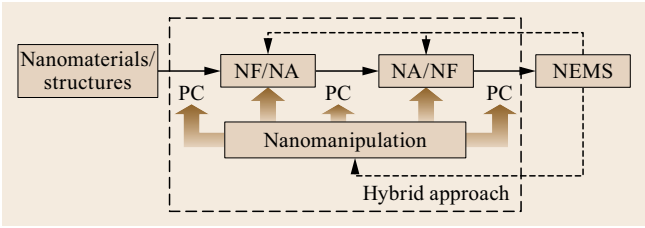


Fig. 18.19 Hybrid approach to NEMS (PC: property characterization, NF: nanofabrication, NA: nanoassembly)

The integration of these devices can be realized using the configurations shown in Figs. 18.20g [18.143] and 18.20h [18.90]. Arrays of individual nanotubes can also be used to fabricate nanosensors, such as position encoders [18.144].

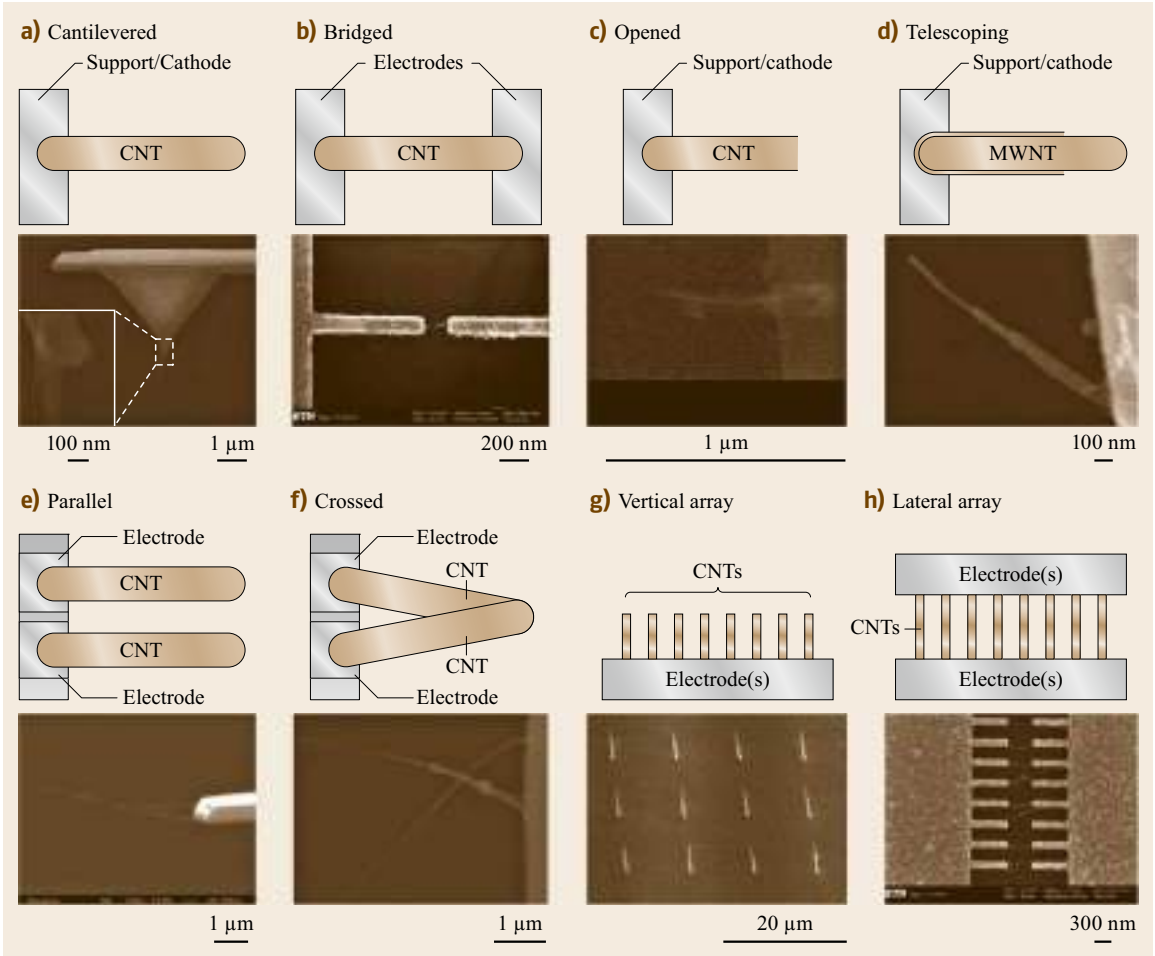


Fig. 18.20a–h Configurations of individual nanotube-based NEMS. *Scale bars*: **(a)** 1 µm (*inset*: 100 nm), **(b)** 200 nm, **(c)** 1 µm, **(d)** 100 nm, **(e)** and **(f)** 1 µm, **(g)** 20 µm, and **(h)** 300 nm. All examples are from the authors' work

Nanotube-based NEMS remains a rich research field with a large number of open problems. New materials and effects at the nanoscale will enable a new family of sensors and actuators for the detection and actuation of ultrasmall quantities or objects with ultrahigh precision and frequencies. Through random spreading, direct growth, and nanorobotic manipulation, prototypes have been demonstrated. However, for integration into NEMS, self-assembly processes will become increasingly important. Among them, we believe that dielectrophoretic nanoassembly will play a significant role for large-scale production of 2-D regular structures.

References

18.1 R.P. Feynman: There's plenty of room at the bottom, Caltech Eng. Sci. **23**, 22–36 (1960)

18.2 M.C. Roco, R.S. Williams, P. Alivisatos: Nanotechnology research directions. In: *Interag. Work. Group Nanosci. Eng. Technol. (IWGN) Workshop Rep.* (Kluwer, Dordrecht 2000)

18.3 Committee on Technology, M.L. Downey, D.T. Moore, G.R. Bachula, D.M. Etter, E.F. Carey, L.A. Perine: *National Nanotechnology Initiative: Leading to the Next Industrial Revolution, A Report by the Interagency Working Group on Nanoscience, Engineering and Technology* (National Science and Technology Council, Washington 2000)

18.4 K. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley Interscience, New York 1992)

18.5 G. Binnig, H. Rohrer, C. Gerber, E. Weibel: Surface studies by scanning tunneling microscopy, Phys. Rev. Lett. **49**, 57–61 (1982)

18.6 W.F. Degrado: Design of peptides and proteins, Adv. Protein Chem. **39**, 51–124 (1998)

18.7 J.-M. Lehn: *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim 1995)

18.8 G.M. Whitesides, B. Grzybowski: Self-assembly at all scales, Science **295**, 2418–2421 (2002)

18.9 M. Fujita, N. Fujita, K. Ogura, K. Yamaguchi: Spontaneous assembling of ten small components into a three-dimensionally interlocked compound consisting of the same two cage frameworks, Nature **400**, 52–55 (1999)

18.10 T. Ebefors, G. Stemme: Microrobotics. In: *The MEMS Handbook*, ed. by M. Gad-el-Hak (CRC, Boca Raton 2002)

18.11 C.-J. Kim, A.P. Pisano, R.S. Muller: Silicon-processed overhanging microgripper, IEEE/ASME J. MEMS **1**, 31–36 (1992)

18.12 C. Liu, T. Tsao, Y.-C. Tai, C.-M. Ho: Surface micromachined magnetic actuators. In: *Proc. 7th IEEE Int. Conf. Micro Electro Mech. Syst., Oiso* (IEEE, Piscataway 1994) pp. 57–62

18.13 J. Judy, D.L. Polla, W.P. Robbins: A linear piezoelectric stepper motor with submicron displacement and centimeter travel, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 428–437 (1990)

18.14 K. Nakamura, H. Ogura, S. Maeda, U. Sangawa, S. Aoki, T. Sato: Evaluation of the micro wobbler motor fabricated by concentric buildup process. In: *Proc. 8th IEEE Int. Conf. Micro Electro Mech. Syst., Amsterdam* (IEEE, Piscataway 1995) pp. 374–379

18.15 A. Teshigahara, M. Watanabe, N. Kawahara, I. Ohtsuka, T. Hattori: Performance of a 7-mm microfabricated car, IEEE/ASME J. MEMS **4**, 76–80 (1995)

18.16 K.R. Udayakumar, S.F. Bart, A.M. Flynn, J. Chen, L.S. Tavrow, L.E. Cross, R.A. Brooks, D.J. Ehrlich: Ferroelectric thin-film ultrasonic micromotors. In: *Proc. 4th IEEE Int. Conf. Micro Electro Mech. Syst., Nara* (IEEE, Piscataway 1991) pp. 109–113

18.17 W. Trimmer, R. Jebens: Actuators for micro robots. In: *Proc. 1989 IEEE Int. Conf. Robot. Autom., Scottsdale* (IEEE, Piscataway 1989) pp. 1547–1552

18.18 P. Dario, R. Valleggi, M.C. Carrozza, M.C. Montesi, M. Cocco: Review – Microactuators for microrobots: A critical survey, J. Micromech. Microeng. **2**, 141–157 (1992)

18.19 I. Shimoyama: Scaling in microrobots. In: *Proc. IEEE/RSJ Intell. Robot. Syst., Pittsburgh* (IEEE, Piscataway 1995) pp. 208–211

18.20 R.S. Fearing: Powering 3-dimensional microrobots: power density limitations. In: *Tutorial on Micro Mechatronics and Micro Robotics, Proc. IEEE Int. Conf. Robot. Autom, Leuven*, ed. by G. Girait, P. Dario (IEEE, Piscataway 1998)

18.21 R.G. Gilbertson, J.D. Busch: A survey of microactuator technologies for future spacecraft missions, J. Br. Interplanet. Soc. **49**, 129–138 (1996)

18.22 M. Mehregany, P. Nagarkar, S.D. Senturia, J.H. Lang: Operation of microfabricated harmonic and ordinary side-drive motors. In: *Proc. 3th IEEE Int. Conf. Micro Electro Mech. Syst., Napa Valley* (IEEE, Piscataway 1990) pp. 1–8

18.23 Y.C. Tai, L.S. Fan, R.S. Muller: IC-processed micromotors: design, technology, and testing. In: *Proc. 2nd IEEE Int. Conf. Micro Electro Mech. Syst., Salt Lake City* (IEEE, Piscataway 1989) pp. 1–6

18.24 T. Ohnstein, T. Fukiura, J. Ridley, U. Bonne: Micromachined silicon microvalve. In: *Proc. 3rd IEEE Int. Conf. Micro Electro Mech. Syst., Napa Valley* (IEEE, Piscataway 1990) pp. 95–99

18.25 L.Y. Chen, S.L. Zhang, J.J. Yao, D.C. Thomas, N.C. MacDonald: Selective chemical vapor deposition of tungsten for microdynamic structures. In: *Proc. 2nd IEEE Int. Conf. Micro Electro Mech. Syst., Salt Lake City* (IEEE, Piscataway 1989) pp. 82–87

18.26 K. Yanagisawa, H. Kuwano, A. Tago: An electromagnetically driven microvalve. In: *Proc. 7th Int. Conf. Solid-State Sens. Actuators, Yokohama* (IEEE, Piscataway 1993) pp. 102–105

18.27 S. Brand: New applications of piezo-electric actuators. In: *Proc. 3rd Int. Conf. New Actuators* (Messe

Bremen GmbH, Bremen 1992) p. 59
18.28 M. Esashi, S. Shoji, A. Nakano: Normally close microvalve and micropump fabricated on a silicon wafer. In: *Proc. 4th IEEE Int. Conf. Micro Electro Mech. Syst., Salt Lake City* (IEEE, Piscataway 1989) pp. 29–34
18.29 R. Petrucci, K. Simmons: An introduction to piezoelectric crystals, Sens. Mag. **May**, 26–31 (1994)
18.30 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56**, 93–96 (1986)
18.31 A. Ashkin, J.M. Dziedzic: Optical trapping and manipulation of viruses and bacteria, Science **235**, 1517–1520 (1987)
18.32 F.H.C. Crick, A.F.W. Hughes: The physical properties of cytoplasm: A study by means of the magnetic particle method, Part I: Experimental, Exp. Cell Res. **1**, 37–80 (1950)
18.33 M.F. Yu, M.J. Dyer, G.D. Skidmore, H.W. Rohrs, X.K. Lu, K.D. Ausman, J.R. Von Ehr, R.S. Ruoff: Three-dimensional manipulation of carbon nanotubes under a scanning electron microscope, Nanotechnology **10**, 244–252 (1999)
18.34 L.X. Dong, F. Arai, T. Fukuda: 3-D nanorobotic manipulation of nano-order objects inside SEM. In: *Proc. 2000 Int. Symp. Micromechatron. Hum. Sci., Nagoya* (IEEE, Piscataway 2000) pp. 151–156
18.35 D.M. Eigler, E.K. Schweizer: Positioning single atoms with a scanning tunneling microscope, Nature **344**, 524–526 (1990)
18.36 P. Avouris: Manipulation of matter at the atomic and molecular levels, Acc. Chem. Res. **28**, 95–102 (1995)
18.37 M.F. Crommie, C.P. Lutz, D.M. Eigler: Confinement of electrons to quantum corrals on a metal surface, Science **262**, 218–220 (1993)
18.38 L.J. Whitman, J.A. Stroscio, R.A. Dragoset, R.J. Cellota: Manipulation of adsorbed atoms and creation of new structures on room-temperature surfaces with a scanning tunneling microscope, Science **251**, 1206–1210 (1991)
18.39 I.-W. Lyo, P. Avouris: Field-induced nanometer-scale to atomic-scale manipulation of silicon surfaces with the STM, Science **253**, 173–176 (1991)
18.40 G. Dujardin, R.E. Walkup, P. Avouris: Dissociation of individual molecules with electrons from the tip of a scanning tunneling microscope, Science **255**, 1232–1235 (1992)
18.41 T.-C. Shen, C. Wang, G.C. Abeln, J.R. Tucker, J.W. Lyding, P. Avouris, R.E. Walkup: Atomic-scale desorption through electronic and vibrational-excitation mechanisms, Science **268**, 1590–1592 (1995)
18.42 M.T. Cuberes, R.R. Schittler, J.K. Gimzewski: Room-temperature repositioning of individual C_60 molecules at Cu steps: Operation of a molecular counting device, Appl. Phys. Lett. **69**, 3016–3018 (1996)
18.43 H.J. Lee, W. Ho: Single-bond formation and characterization with a scanning tunneling microscope, Science **286**, 1719–1722 (1999)
18.44 T. Yamamoto, O. Kurosawa, H. Kabata, N. Shimamoto, M. Washizu: Molecular surgery of DNA based on electrostatic micromanipulation, IEEE Trans. IA **36**, 1010–1017 (2000)
18.45 C. Haber, D. Wirtz: Magnetic tweezers for DNA micromanipulation, Rev. Sci. Instrum. **71**, 4561–4570 (2000)
18.46 D.M. Schäfer, R. Reifenberger, A. Patil, R.P. Andres: Fabrication of two-dimensional arrays of nanometer-size clusters with the atomic force microscope, Appl. Phys. Lett. **66**, 1012–1014 (1995)
18.47 T. Junno, K. Deppert, L. Montelius, L. Samuelson: Controlled manipulation of nanoparticles with an atomic force microscope, Appl. Phys. Lett. **66**, 3627–3629 (1995)
18.48 P.E. Sheehan, C.M. Lieber: Nanomachining, manipulation and fabrication by force microscopy, Nanotechnology **7**, 236–240 (1996)
18.49 C. Baur, B.C. Gazen, B. Koel, T.R. Ramachandran, A.A.G. Requicha, L. Zini: Robotic nanomanipulation with a scanning probe microscope in a networked computing environment, J. Vac. Sci. Technol. B **15**, 1577–1580 (1997)
18.50 R. Resch, C. Baur, A. Bugacov, B.E. Koel, A. Madhukar, A.A.G. Requicha, P. Will: Building and manipulating 3-D and linked 2-D structures of nanoparticles using scanning force microscopy, Langmuir **14**, 6613–6616 (1998)
18.51 J. Hu, Z.-H. Zhang, Z.-Q. Ouyang, S.-F. Chen, M.-Q. Li, F.-J. Yang: Stretch and align virus in nanometer scale on an atomically flat surface, J. Vac. Sci. Technol. B **16**, 2841–2843 (1998)
18.52 M. Sitti, S. Horiguchi, H. Hashimoto: Controlled pushing of nanoparticles: Modeling and experiments, IEEE/ASME Trans. Mechatron. **5**, 199–211 (2000)
18.53 M. Guthold, M.R. Falvo, W.G. Matthews, S. Paulson, S. Washburn, D.A. Erie, R. Superfine, F.P. Brooks Jr., R.M. Taylor II: Controlled manipulation of molecular samples with the nanoManipulator, IEEE/ASME Trans. Mechatron. **5**, 189–198 (2000)
18.54 F. Arai, D. Andou, T. Fukuda: Micro manipulation based on micro physics – strategy based on attractive force reduction and stress measurement. In: *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., Pittsburgh* (IEEE, Piscataway 1995) pp. 236–241
18.55 L.X. Dong, F. Arai, T. Fukuda: 3-D nanorobotic manipulations of nanometer scale objects, J. Robot. Mechatron. **13**, 146–153 (2001)
18.56 H.W.P. Koops, J. Kretz, M. Rudolph, M. Weber, G. Dahm, K.L. Lee: Characterization and application of materials grown by electron-beam-induced deposition, Jpn. J. Appl. Phys. **33**, 7099–7107 (1994)
18.57 S. Iijima: Helical microtubules of graphitic carbon, Nature **354**, 56–58 (1991)
18.58 S.J. Tans, A.R.M. Verchueren, C. Dekker: Room-temperature transistor based on a single carbon nanotube, Nature **393**, 49–52 (1998)
18.59 Y. Huang, X.F. Duan, Y. Cui, L.J. Lauhon, K.-H. Kim, C.M. Lieber: Logic gates and computation from as-

sembled nanowire building blocks, Science **294**, 1313–1317 (2001)
18.60 A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker: Logic circuits with carbon nanotube transistors, Science **294**, 1317–1320 (2001)
18.61 R.H. Baughman, A.A. Zakhidov, W.A. de Heer: Carbon nanotubes – the route toward applications, Science **297**, 787–792 (2002)
18.62 N. Wang, Z.K. Tang, G.D. Li, J.S. Chen: Single-walled 4 Å carbon nanotube arrays, Nature **408**, 50–51 (2000)
18.63 Z.W. Pan, S.S. Xie, B.H. Chang, C.Y. Wang, L. Lu, W. Liu, W.Y. Zhou, W.Z. Li, L.X. Qian: Very long carbon nanotubes, Nature **394**, 631–632 (1998)
18.64 H.W. Zhu, C.L. Xu, D.H. Wu, B.Q. Wei, R. Vajtai, P.M. Ajayan: Direct synthesis of long single-walled carbon nanotube strands, Science **296**, 884–886 (2002)
18.65 M.J. Treacy, T.W. Ebbesen, J.M. Gibson: Exceptionally high Young's modulus observed for individual carbon nanotubes, Nature **381**, 678–680 (1996)
18.66 E.W. Wong, P.E. Sheehan, C.M. Lieber: Nanobeam mechanics: Elasticity, strength, and toughness of nanorods and nanotubes, Science **277**, 1971–1975 (1997)
18.67 P. Poncharal, Z.L. Wang, D. Ugarte, W.A. de Heer: Electrostatic deflections and electromechanical resonances of carbon nanotubes, Science **283**, 1513–1516 (1999)
18.68 M.F. Yu, O. Lourie, M.J. Dyer, K. Moloni, T.F. Kelley, R.S. Ruoff: Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load, Science **287**, 637–640 (2000)
18.69 A. Krishnan, E. Dujardin, T.W. Ebbesen, P.N. Yianilos, M.M.J. Treacy: Young's modulus of single-walled nanotubes, Phys. Rev. B **58**, 14–013–14–019 (1998)
18.70 J.P. Salvetat, G.A.D. Briggs, J.-M. Bonard, R.R. Bacsa, A.J. Kulik, T. Stockli, N.A. Burnham, L. Forro: Elastic and shear moduli of single-walled carbon nanotube ropes, Phys. Rev. Lett. **82**, 944–947 (1999)
18.71 M.F. Yu, B.S. Files, S. Arepalli, R.S. Ruoff: Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties, Phys. Rev. Lett. **84**, 5552–5555 (2000)
18.72 D.A. Walters, L.M. Ericson, M.J. Casavant, J. Liu, D.T. Colbert, K.A. Smith, R.E. Smalley: Elastic strain of freely suspended single-wall carbon nanotube ropes, Appl. Phys. Lett. **74**, 3803–3805 (1999)
18.73 R. Saito, M. Fujita, G. Dresselhaus, M.S. Dresselhaus: Electronic structure of graphene tubules based on C_{60}, Phys. Rev. B **46**, 1804–1811 (1992)
18.74 T.W. Ebbesen, H.J. Lezec, H. Hiura, J.W. Bennett, H.F. Ghaemi, T. Thio: Electrical conductivity of individual carbon nanotubes, Nature **382**, 54–56 (1996)
18.75 H.J. Dai, E.W. Wong, C.M. Lieber: Probing electrical transport in nanomaterials: conductivity of individual carbon nanotubes, Science **272**, 523–526 (1996)
18.76 P. Kim, L. Shi, A. Majumdar, P.L. McEuen: Thermal transport measurements of individual multiwalled nanotubes, Phys. Rev. Lett. **87**, 215502 (2001)
18.77 W.J. Liang, M. Bockrath, D. Bozovic, J.H. Hafner, M. Tinkham, H. Park: Fabry–Perot interference in a nanotube electron waveguide, Nature **411**, 665–669 (2001)
18.78 S. Frank, P. Poncharal, Z.L. Wang, W.A. de Heer: Carbon nanotube quantum resistors, Science **280**, 1744–1746 (1998)
18.79 X.B. Zhang, D. Bernaerts, G. Van Tendeloo, S. Amelincks, J. Van Landuyt, V. Ivanov, J.B. Nagy, P. Lambin, A.A. Lucas: The texture of catalytically grown coil-shaped carbon nanotubules, Europhys. Lett. **27**, 141–146 (1994)
18.80 X.Y. Kong, Z.L. Wang: Spontaneous polarization-induced nanohelixes, nanosprings, and nanorings of piezoelectric nanobelts, Nano Lett. **3**, 1625–1631 (2003)
18.81 S.V. Golod, V.Y. Prinz, V.I. Mashanov, A.K. Gutakovsky: Fabrication of conducting GeSi/Si micro- and nanotubes and helical microcoils, Semicond. Sci. Technol. **16**, 181–185 (2001)
18.82 L. Zhang, E. Deckhardt, A. Weber, C. Schönenberger, D. Grützmacher: Controllable fabrication of SiGe/Si and SiGe/Si/Cr helical nanobelts, Nanotechnology **16**, 655–663 (2005)
18.83 D.J. Bell, L.X. Dong, Y. Sun, L. Zhang, B.J. Nelson, D. Grützmacher: Manipulation of nanocoils for nanoelectromagnets. In: *Proc. 5th IEEE Conf. Nanotechnol., Nagoya* (IEEE, Piscataway 2005) pp. 149–152
18.84 D.J. Bell, Y. Sun, L. Zhang, L.X. Dong, B.J. Nelson, D. Grützmacher: Three-dimensional nanosprings for electromechanical sensors. In: *Proc. 13th Int. Conf. Solid-State Sens. Actuators Microsyst., Seoul* (IEEE, Piscataway 2005) pp. 15–18
18.85 S. Iijima, T. Ichihashi: Single-shell carbon nanotubes of 1-nm diameter, Nature **363**, 603–605 (1993)
18.86 D.S. Bethune, C.H. Kiang, M.S. de Vries, G. Gorman, R. Savoy, J. Vazquez, R. Beyers: Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls, Nature **363**, 605–607 (1993)
18.87 R. Martel, T. Schmidt, H.R. Shea, T. Hertel, P. Avouris: Single- and multi-wall carbon nanotube field-effect transistors, Appl. Phys. Lett. **73**, 2447–2449 (1998)
18.88 N.R. Franklin, Y.M. Li, R.J. Chen, A. Javey, H.J. Dai: Patterned growth of single-walled carbon nanotubes on full 4-inch wafers, Appl. Phys. Lett. **79**, 4571–4573 (2001)
18.89 T. Rueckes, K. Kim, E. Joselevich, G.Y. Tseng, C.-L. Cheung, C.M. Lieber: Carbon nanotube-based non-volatile random access memory for molecular computing, Science **289**, 94–97 (2000)
18.90 A. Subramanian, B. Vikramaditya, L.X. Dong, D. Bell, B.J. Nelson: Micro and nanorobotic assembly using dielectrophoresis. In: *Robotics:*

Science and Systems I, ed. by S. Thrun, G.S. Sukhatme, S. Schaal, O. Brock (MIT Press, Cambridge 2005) pp. 327–334

18.91 T. Fukuda, F. Arai, L.X. Dong: Assembly of nanodevices with carbon nanotubes through nanorobotic manipulations, Proc. IEEE **91**, 1803–1818 (2003)

18.92 M.R. Falvo, G.J. Clary, R.M. Taylor, V. Chi, F.P. Brooks, S. Washburn, R. Superfine: Bending and buckling of carbon nanotubes under large strain, Nature **389**, 582–584 (1997)

18.93 H.W.C. Postma, A. Sellmeijer, C. Dekker: Manipulation and imaging of individual single-walled carbon nanotubes with an atomic force microscope, Adv. Mater. **12**, 1299–1302 (2000)

18.94 H.W.C. Postma, M. de Jonge, Z. Yao, C. Dekker: Electrical transport through carbon nanotube junctions created by mechanical manipulation, Phys. Rev. B **62**, R10653–R10656 (2000)

18.95 T. Hertel, R. Martel, P. Avouris: Manipulation of individual carbon nanotubes and their interaction with surfaces, J. Phys. Chem. B **102**, 910–915 (1998)

18.96 P. Avouris, T. Hertel, R. Martel, T. Schmidt, H.R. Shea, R.E. Walkup: Carbon nanotubes: nanomechanics, manipulation, and electronic devices, Appl. Surf. Sci. **141**, 201–209 (1999)

18.97 L. Roschier, J. Penttila, M. Martin, P. Hakonen, M. Paalanen, U. Tapper, E.I. Kauppinen, C. Journet, P. Bernier: Single-electron transistor made of multiwalled carbon nanotube using scanning probe manipulation, Appl. Phys. Lett. **75**, 728–730 (1999)

18.98 M. AhIskog, R. Tarkiainen, L. Roschier, P. Hakonen: Single-electron transistor made of two crossing multiwalled carbon nanotubes and its noise properties, Appl. Phys. Lett. **77**, 4037–4039 (2000)

18.99 M.R. Falvo, R.M. Taylor II, A. Helser, V. Chi, F.P. Brooks Jr, S. Washburn, R. Superfine: Nanometre-scale rolling and sliding of carbon nanotubes, Nature **397**, 236–238 (1999)

18.100 M.R. Falvo, J. Steele, R.M. Taylor II, R. Superfine: Gearlike rolling motion mediated by commensurate contact: Carbon nanotubes on HOPG, Phys. Rev. B **62**, R10665–R10667 (2000)

18.101 L.X. Dong: *Nanorobotic Manipulations of Carbon Nanotubes*, Ph.D. Thesis (Nagoya University, Nagoya 2003)

18.102 J.H. Hafner, C.-L. Cheung, T.H. Oosterkamp, C.M. Lieber: High-yield assembly of individual single-walled carbon nanotube tips for scanning probe microscopies, J. Phys. Chem. B **105**, 743–746 (2001)

18.103 L.X. Dong, F. Arai, T. Fukuda: Electron-beam-induced deposition with carbon nanotube emitters, Appl. Phys. Lett. **81**, 1919–1921 (2002)

18.104 L.X. Dong, F. Arai, T. Fukuda: 3-D nanorobotic manipulations of multi-walled carbon nanotubes. In: *Proc. 2001 IEEE Int. Conf. Robot. Autom. (ICRA2001), Seoul* (IEEE, Piscataway 2001) pp. 632–637

18.105 L.X. Dong, F. Arai, T. Fukuda: Three-dimensional nanorobotic manipulations of carbon nanotubes, J. Robot. Mechatron. JSME **14**, 245–252 (2002)

18.106 L.X. Dong, F. Arai, T. Fukuda: Inter-process measurement of MWNT rigidity and fabrication of MWNT junctions through nanorobotic manipulations. In: *Conf. Proc. Nanonetw. Mater. Fuller. Nanotub. Relat. Mater.*, Vol. 590, (AIP, Melville 2001) pp. 71–74

18.107 L.X. Dong, F. Arai, T. Fukuda: Shape modification of carbon nanotubes and its applications in nanotube scissors. In: *Proc. IEEE Int. Conf. Nanotechnol. (IEEE-NANO2002), Washington* (IEEE, Piscataway 2002) pp. 443–446

18.108 L.X. Dong, F. Arai, T. Fukuda: Destructive constructions of nanostructures with carbon nanotubes through nanorobotic manipulations, IEEE/ASME Trans. Mechatron. **9**, 350–357 (2004)

18.109 J. Cumings, P.G. Collins, A. Zettl: Peeling and sharpening multiwall nanotubes, Nature **406**, 58 (2000)

18.110 J. Cumings, A. Zettl: Low-friction nanoscale linear bearing realized from multiwall carbon nanotubes, Science **289**, 602–604 (2000)

18.111 L.X. Dong, F. Arai, T. Fukuda: 3-D nanoassembly of carbon nanotubes through nanorobotic manipulations. In: *Proc. 2002 IEEE Int. Conf. Robot. Autom. (ICRA2002), Washington* (IEEE, Piscataway 2002) pp. 1477–1482

18.112 L.X. Dong, F. Arai, T. Fukuda: Mechanochemical nanorobotic manipulations of carbon nanotubes, Jpn. J. Appl. Phys. **42**, 295–298 (2003)

18.113 R. Saito, G. Dresselhaus, M.S. Dresselhaus: Tunneling conductance of connected carbon nanotubes, Phys. Rev. B **53**, 2044–2050 (1996)

18.114 L. Chico, V.H. Crespi, L.X. Benedict, S.G. Louie, M.L. Cohen: Pure carbon nanoscale devices: nanotube heterojunctions, Phys. Rev. Lett. **76**, 971–974 (1996)

18.115 M. Menon, D. Srivastava: Carbon nanotube *T junctions*: nanoscale metal–semiconductor–metal contact devices, Phys. Rev. Lett. **79**, 4453–4456 (1997)

18.116 Z. Yao, H.W.C. Postma, L. Balents, C. Dekker: Carbon nanotube intramolecular junctions, Nature **402**, 273–276 (1999)

18.117 H.W.C. Postma, T. Teepen, Z. Yao, M. Grifoni, C. Dekker: Carbon nanotube single-electron transistors at room temperature, Science **293**, 76–79 (2001)

18.118 M.S. Fuhrer, L. Shih, M. Forero, Y.-G. Yoon, M.S.C. Mazzoni, H.J. Choi, J. Ihm, S.G. Louie, A. Zettl, P.L. McEuen: Crossed nanotube junctions, Science **288**, 494–497 (2000)

18.119 T. Rueckes, K. Kim, E. Joselevich, G.Y. Treng, C.L. Cheung, C.M. Lieber: Carbon nanotube-based nonvolatile random access memory for molecular computing science, Science **289**, 94–97 (2000)

18.120 A.G. Rinzler, J.H. Hafner, P. Nikolaev, L. Lou, S.G. Kim, D. Tománek, P. Nordlander, D.T. Colbert, R.E. Smalley: Unraveling nanotubes: field emission from an atomic wire, Science **269**, 1550–1553

(1995)
18.121 Y. Saito, K. Hamaguchi, K. Hata, K. Uchida, Y. Tasaka, F. Ikazaki, M. Yumura, A. Kasaya, Y. Nishina: Conical beams from open nanotubes, Nature **389**, 554 (1997)
18.122 S.C. Minne, G. Yaralioglu, S.R. Manalis, J.D. Adams, J. Zesch, A. Atalar, C.F. Quate: Automated parallel high-speed atomic force microscopy, Appl. Phys. Lett. **72**, 2340–2342 (1998)
18.123 G.D. Skidmore, E. Parker, M. Ellis, N. Sarkar, R. Merkle: Exponential assembly, Nanotechnology **11**, 316–321 (2001)
18.124 Y. Cui, Q.Q. Wei, H.K. Park, C.M. Lieber: Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species, Science **293**, 1289–1292 (2001)
18.125 Y. Sun, B.J. Nelson: Microrobotic cell injection. In: *Proc. 2001 IEEE Int. Conf. Robot. Autom. (ICRA2001), Seoul* (IEEE, Piscataway 2001) pp. 620–625
18.126 Y. Sun, B.J. Nelson: Autonomous injection of biological cells using visual servoing. In: *Int. Symp. Exp. Robot. (ISER)* (Lecture Notes Control Inform. Sci., Hawaii 2000) pp. 175–184
18.127 Y. Sun, B.J. Nelson, D.P. Potasek, E. Enikov: A bulk microfabricated multi-axis capacitive cellular force sensor using transverse comb drives, J. Micromech. Microeng. **12**, 832–840 (2002)
18.128 Y. Sun, K. Wan, K.P. Roberts, J.C. Bischof, B.J. Nelson: Mechanical property characterization of mouse zona pellucida, IEEE Trans. Nanobiosci. **2**, 279–286 (2003)
18.129 H.J. Dai, J.H. Hafner, A.G. Rinzler, D.T. Colbert, R.E. Smalley: Nanotubes as nanoprobes in scanning probe microscopy, Nature **384**, 147–150 (1996)
18.130 J.H. Hafner, C.L. Cheung, C.M. Lieber: Growth of nanotubes for probe microscopy tips, Nature **398**, 761–762 (1999)
18.131 H. Nishijima, S. Kamo, S. Akita, Y. Nakayama, K.I. Hohmura, S.H. Yoshimura, K. Takeyasu: Carbon-nanotube tips for scanning probe microscopy: preparation by a controlled process and observation of deoxyribonucleic acid, Appl. Phys. Lett. **74**, 4061–4063 (1999)
18.132 L.X. Dong, F. Arai, M. Nakajima, P. Liu, T. Fukuda: Nanotube devices fabricated in a nano laboratory. In: *Proc. 2003 IEEE Int. Conf. Robot. Autom. (ICRA2003), Taipei* (IEEE, Piscataway 2003) pp. 3624–3629
18.133 M. Nakajima, F. Arai, L.X. Dong, T. Fukuda: Pico-Newton order force measurement using a calibrated carbon nanotube probe inside a scanning electron microscope, J. Robot. Mechatron. JSME **16**(2), 155–162 (2004)
18.134 A. Subramanian, B.J. Nelson, L.X. Dong, D. Bell: Dielectrophoretic nanoassembly of nanotube-based NEMS with nanoelectrodes. In: *6th IEEE Int. Symp. Assem. Task Plan., Montréal* (IEEE, Piscataway 2005) pp. 200–205
18.135 J. Kong, N.R. Franklin, C.W. Zhou, M.G. Chapline, S. Peng, K.J. Cho, H.J. Dai: Nanotube molecular wires as chemical sensors, Science **287**, 622–625 (2000)
18.136 L.X. Dong, B.J. Nelson, T. Fukuda, F. Arai: Towards linear nano servomotors with integrated position sensing. In: *Proc. 2005 IEEE Int. Conf. Robot. Autom., Barcelona* (IEEE, Piscataway 2005) pp. 867–872
18.137 Y.H. Gao, Y. Bando: Carbon nanothermometer containing gallium, Nature **415**, 599 (2002)
18.138 R.H. Baughman, C.X. Cui, A.A. Zakhidov, Z. Iqbal, J.N. Barisci, G.M. Spinks, G.G. Wallace, A. Mazzoldi, D. De Rossi, A.G. Rinzler, O. Jaschinski, S. Roth, M. Kertesz: Carbon nanotube actuators, Science **284**, 1340–1344 (1999)
18.139 S.W. Lee, D.S. Lee, R.E. Morjan, S.H. Jhang, M. Sveningsson, O.A. Nerushev, Y.W. Park, E.E.B. Campbell: A three-terminal carbon nanorelay, Nano Lett. **4**, 2027–2030 (2004)
18.140 A.M. Fennimore, T.D. Yuzvinsky, W.-Q. Han, M.S. Fuhrer, J. Cumings, A. Zettl: Rotational actuators based on carbon nanotubes, Nature **424**, 408–410 (2003)
18.141 P. Kim, C.M. Lieber: Nanotube nanotweezers, Science **286**, 2148–2150 (1999)
18.142 C.K.M. Fung, V.T.S. Wong, R.H.M. Chan, W.J. Li: Dielectrophoretic batch fabrication of bundled carbon nanotube thermal sensors, IEEE Trans. Nanotech. **3**, 395–403 (2004)
18.143 A. Subramanian, L.X. Dong, B.J. Nelson: Selective eradication of individual nanotubes from vertically aligned arrays. In: *Proc. 2005 IEEE/ASME Int. Conf. Adv. Intell. Mechatron., Monterey* (IEEE, Piscataway 2005) pp. 105–110
18.144 L.X. Dong, A. Subramanian, B.J. Nelson: Nano encoders based on arrays of single nanotube emitter. In: *Proc. 5th IEEE Conf. Nanotechnol., Nagoya* (IEEE, Piscataway 2005) pp. 211–214

BioMEMS Part D

Part D BioMEMS/NEMS

19 Applications of MEMS to Cell Biology
Georg E. Fantner, Lausanne, Switzerland
Pascal D. Odermatt, Lausanne, Switzerland
Haig Alexander Eskandarian, Lausanne, Switzerland

20 Contact-free Mechanical Manipulation of Biological Materials
Jörg Schnauß, Leipzig, Germany
Josef A. Käs, Leipzig, Germany
David M. Smith, Leipzig, Germany

21 Nano-Particles for Biomedical Applications
Paolo Decuzzi, Genova, Italy
Alessandro Coclite, Genova, Italy
Aeju Lee, Kumamoto, Japan
Anna Lisa Palange, Genova, Italy
Daniele Di Mascolo, Genova, Italy
Ciro Chiappini, London, UK
Hélder A. Santos, Helsinki, Finland
Maria Laura Coluccio, Catanzaro, Italy
Gerardo Perozziello, Catanzaro, Italy
Patrizio Candeloro, Catanzaro, Italy
Enzo Di Fabrizio, Thuwal, Saudi Arabia
Francesco Gentile, Naples, Italy

22 Biological Molecules in Therapeutic Nanodevices
Stephen C. Lee, Columbus, USA
Bharat Bhushan, Columbus, USA

19. Applications of MEMS to Cell Biology

Georg E. Fantner, Pascal D. Odermatt, Haig Alexander Eskandarian

Relatively simple microscale devices such as cantilevers already have found distinct applications in the study of cells and subcellular structures. The ability to measure small forces, small features, and small masses has led to unique and elegant measurement solutions. In this chapter, we give a brief background of some relevant fundamental biology and provide examples of research in cell biology, ranging from mechanobiology, physiology, microbiology to cancer biology, where MEMS- and NEMS-based methods have had a significant and, most likely, lasting impact.

19.1 **Biological Background** 588
19.2 **High Resolution Microscopy Methods for Live Cell Imaging** 592
19.3 **Measuring Mechanical Properties on Living Cells** 593
19.3.1 Using MEMS Cantilevers to Measure Cell Mechanical Properties 593
19.3.2 Cantilever Calibration 594
19.3.3 Instrument Calibration 594
19.3.4 Extracting Materials Properties from Cell-Indentation Measurements 595
19.4 **Cancer Detection Using Nanomechanical Sensors** 596
19.5 **Measurement of Cell Adhesion Using Cantilever Based Force Sensor** 599
19.5.1 Tip Functionalization Methods 599
19.5.2 Bacterial Adhesion 600
19.5.3 Cell–Substrate Interaction (Biomaterials/Tissue Engineering) 601
19.5.4 Cell–Cell Interaction 601
19.6 **Manipulation of Cells Using MEMS Cantilevers** 602
19.7 **Mass Measurements of Single Cells Using Resonant Cantilevers** 603
19.7.1 Resonant Mechanical Sensors 603
19.7.2 Cellular Growth 604
19.8 **Characterization of Antibiotic Action Using MEMS Devices** 606
19.8.1 Characterizing the Impact of Antibiotics 606
19.8.2 Detecting Fluctuations in the Buoyant Mass of Single Bacteria 606
19.8.3 Detecting the Dynamic Motion of Bacteria in Fluid 606
19.8.4 Antibiotic Effect on the Bacterial Cell Surface 607
19.9 **Conclusions and Outlook** 609
References 609

In size, NEMS and MEMS are on a similar scale as many of the important building blocks of life (Fig. 19.1). This makes them amenable tools for studying these biological entities in a variety of new ways. A nanowire can be used to measure the mass of individual molecules on the order of tens to hundreds of angstroms in size. An atomic force microscope (AFM)-cantilever tip has a radius of around 1–10 nm, similar to molecules like DNA (Deoxyribonucleic acid). The height of the tip is around 10 μm, which makes it a suitable tool to measure bacterial and yeast cells, which are on the order of 0.5–3 μm in height. An AFM-cantilever body has dimensions of around 70 μm in length and 10 μm in width, on which bacterial micro-colonies and single cells, as well as individual eukaryotic cells, may be placed in order to measure the cell mass. Yet thicker cantilevers that have been made hollow permit cells to pass through the cantilever body, allowing for the measurement of buoyant mass. Taken together, the diversity of micro- and nano-electro-mechanical systems being developed is directly contributing to the broadening of human understanding of the animate sphere and its morphogenesis.

In the next section, we describe briefly some fundamental aspects of biology that should aid the reader to understand the use of MEMS and NEMS in the cell-biology applications discussed in the later sections.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_19

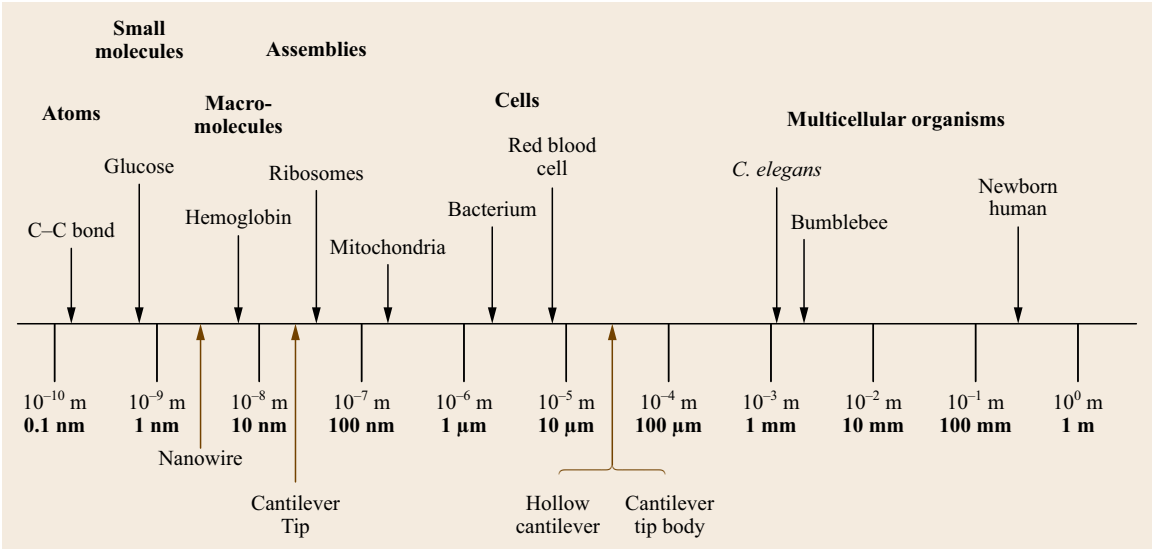


Fig. 19.1 Components of life range in size from sub-nanometers to meters. Above the marginal line are components of life and below are the sizes the specific MEMS devices. These MEMS devices are mainly utilized to measure living components smaller in size than the dimensions of the device, itself

19.1 Biological Background

The most basic unit of animate life is the cell. While cells vary in their form, composition, and morphogenesis, certain fundamental components are conserved [19.1]. Lipid-bilayer membranes enclose all cells. The internal compartment is a hydrophilic space known as the cytosol. It is filled with components necessary for sustaining and propagating life. These components are wide ranging and can be categorized into raw materials, as well as synthesized biochemical compounds. The reactions that produce these chemicals are the result of catalysis by sequences of enzymatically active amino acids, known as peptides or proteins, or ribonucleic-acid sequences, RNAs. These enzymatic components are constructed and encoded by a core of biochemical components common to all known life forms. The concept that describes these core components from which life is comprised is called the *Central Dogma*. Protein synthesis (known as translation) is catalyzed by a molecule known as the ribosome. Ribosomes are structurally composed of proteins and function enzymatically through the activity of RNA molecules. The amino-acid sequence of proteins is encoded by other RNA sequences. However, unprotected RNA molecules in the cytosol have a low half-life, which means that the sequence specificity of unique proteins and RNA must be encoded by another chemical component retained within the cell. This molecule happens to be deoxyribonucleic acid, DNA. Specific proteins known as polymerases read the DNA code and catalyze the synthesis of RNA molecules. The process of RNA synthesis from DNA is known as transcription. DNA encodes the traits that are inherited for a specific species from one generation to the next. In order to equally partition the genetic code among cellular progeny, a new copy of DNA is synthesized in a process known as replication. Building upon the central dogma are multiple regulatory systems that control the expression of RNA sequences and the ensuing enzymatic processes. Some basic regulatory attributes are conserved among the various kingdoms of life, although some attributes are also the main conduit to cell differentiation and specialization. The ability for life to exist and thrive in a multitude of diverse environmental conditions underlies the multitude of ways in which these regulatory mechanisms manipulate the basic building blocks of the central dogma.

The physical and chemical constraints of the environment can have a high selective pressure on cell survival. Species of the various kingdoms of life have developed unique attributes to counter these specific physical attributes to which they are constrained. One physical property that constrains cell structure and function is osmotic pressure. In order for cells to carry out various biochemical functions, particles and ions

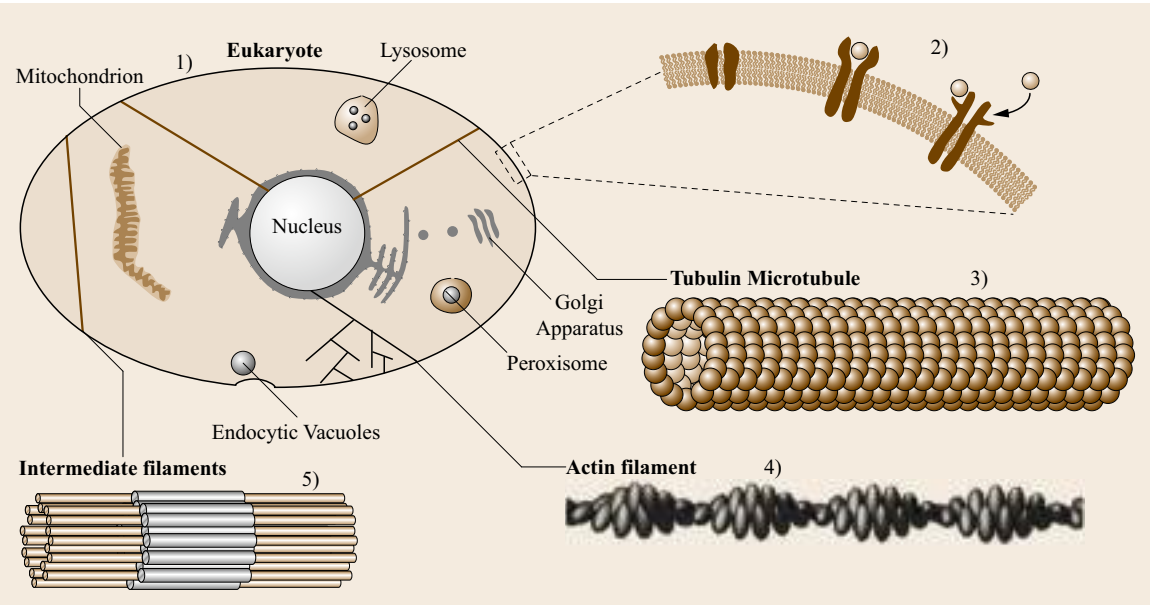


Fig. 19.2 Components and structure of eukaryotic cells, including basic organelles and the parts of the cytoskeleton. Pictured clockwise from the top left are 1) a schematic representation of a eukaryotic cells with some of its most fundamental components; followed by 2) the cell membrane, which is composed of a lipid bilayer and protein pores and signaling receptors (*red*) that are activated upon binding of their specific agonist (*yellow*); 3) a microtubule, composed of tubulin monomers; 4) actin filaments; and finally, 5) intermediate filaments

are concentrated in the cytoplasm. Without the proper structural and biophysical constraints, once an ionic flux has been created, water molecules would rush in to equilibrate the colligative properties of the cell and, in the process, burst cells open. How therefore do cells maintain a high internal osmotic pressure? In order to address this question central to the existence of life, we must describe the core structural and biochemical components of cell.

Two major groups of life forms (called kingdoms) are conceptually of particular interest: prokaryotes and eukaryotes. Prokaryotes are largely unicellular organisms that range in size from several hundred nanometers (nm) to about 10 micrometers (μm) (Fig. 19.1). Bacteria are one of the most prominent representatives of this group. In contrast, eukaryotes are comprised of both unicellular (yeast) and multicellular organisms (animals, plants). Eukaryotic cells are on the order of tens to hundreds of micrometers in size (Fig. 19.1) and possess internal compartments within the cytoplasm, called organelles (Fig. 19.2). A eukaryote's organelles can be classified into two groups. The first are believed to be derived from prokaryotes and form a symbiotic relationship with the eukaryotic host. These organelles are called the mitochondrion among members of the animal kingdom and the chloroplasts in plants. The second group of eukaryotic organelles is derived from the cell membrane. These organelles enclose essential components destined for core-cell biochemical functions and more refined intercellular communication. These organelles include the nuclear membrane, endoplasmic reticulum, Golgi apparatus, lysosomes, peroxisomes, vacuoles (both endocytic and exocytic), and various types of inclusion bodies (Fig. 19.2). In order to maintain these organelles' functioning, the cell must provide basic nutrients, ions, and biochemical gradients. Eukaryotes use passive and active diffusion of particles through pores which are gated by the presence of specific ions and the hydrolysis of energy-storing molecules. Cells maintain these selective pores at the cell surface in order to regulate the level of permeability of the cell membrane. However, as ionic concentration increases inside cells, so does the turgor pressure, which puts stress on the cell membrane, compartmentalizing the internal cellular components. To maintain a cell's morphological integrity, eukaryotes use three types of mechano-proteins to form mechanically rigid filaments of actin or tubulin and intermediate filaments (Fig. 19.2). All three groups of proteins require the hydrolysis of energy-storing molecules to catalyze the formation of filaments. In many cells, actin filaments form a semi-rigid network known as the cell cortex, offering some coherent morphology near the

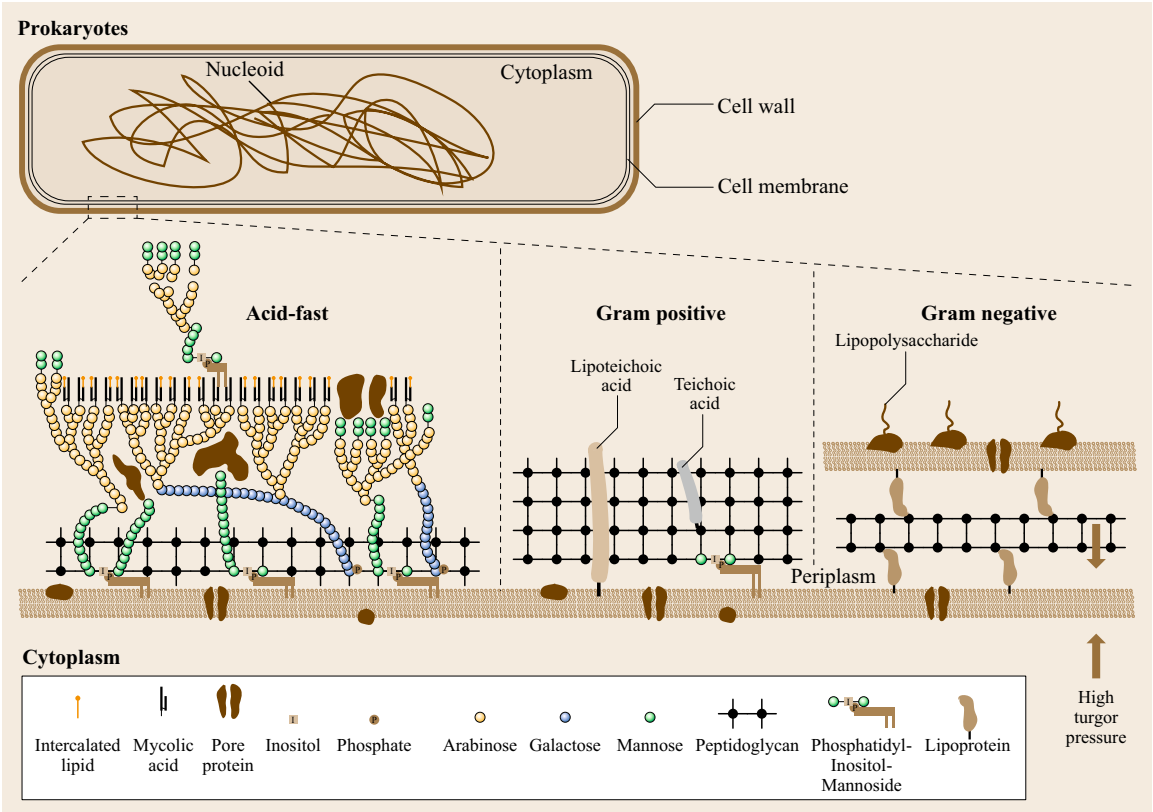


Fig. 19.3 Bacterial cell structure and components of the cell wall. Bacteria can be classified into several groups based on the composition of their cell wall. Depicted are the cell walls of three major types of bacteria based on their pattern of staining with specific dyes. Acid-fast bacteria include actinomycetes like mycobacteria and are comprised of a mycolic acid layer at its surface, conferring a high level of hydrophobicity. Gram-positive bacteria are composed of a thick peptidoglycan layer, giving the cell structural rigidity. Gram-negative bacteria have a second lipid-bilayer surrounding the peptidoglycan providing extra selectivity for molecules that may gain access to the cell cytoplasm

cell surface. The second group of mechano-proteins, the tubulins, is described as fibrous molecules enabling cells to perform active transport of cargo molecules within the cytoplasm. The third group of mechano-proteins, intermediate filaments, is believed to support tension though it cannot create it [19.2]. Together, these mechano-proteins comprise what is considered as the cytoskeleton.

Prokaryotes are not considered to possess cytoskeletal elements functioning to maintain cell morphology. In contrast, most bacteria are able to maintain a high-turgor pressure by mechanically constraining the impermeable cytosolic membrane with the formation of chemical components cross-linked into a rigid meshwork on the outer side of the cell membrane. This outer shell is known as the cell wall and is essential for maintaining bacterial cell morphology and impermeability to ions and water. The cell wall is composed of functional units of carbohydrate and peptide chains which are connected to one another by bonding adjoining sugar molecules forming a stress-resistant sacculus. These sugar bonds provide the bulk of the mechanical rigidity of the bacterial cell wall. The sugar molecules are portions of individual component units of the cell wall, for which the peptidoglycan layer is a central component of most bacterial cell walls. The peptidoglycan layer contains a specific sugar, called muramic acid, which helps to incorporate peptides into the cell wall (Fig. 19.3). The outer layers of the cell wall differ among bacteria. Some, known as the Gram-negative bacteria have a second impermeable cell membrane surrounding the peptidoglycan layer. Yet other bacteria lack a second lipid bilayer and are traditionally classified by the various dyes that they retain while staining, like Gram-staining or the acid-fast stain, Ziehl–Neelsen. One class of bacteria, the actinomycetes, possesses yet other thick layers above the peptidoglycan layer which provide the cell

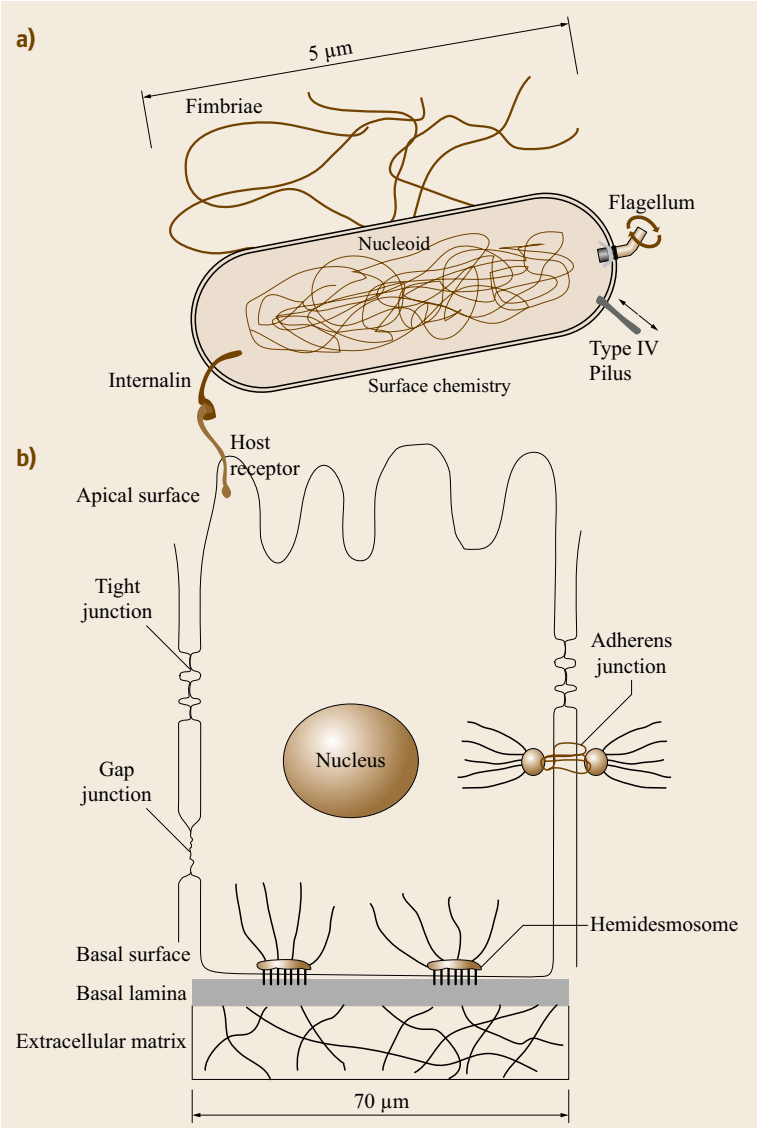


Fig. 19.4a,b Components of prokaryotic and eukaryotic cell adherence. **(a)** Components of bacterial cells conferring adherence on surfaces are diverse in morphology and chemistry. The cell surface provides nonspecific adherence, like the fibrous fimbriae. Fimbriae are passive pili, which is in contrast to type IV pili that can extend and contract in order to provide motion and reorientation to bacterial cells. Flagella provide motion for bacteria, but are also identifiable by receptors of the host immune system. Bacteria also exhibit molecules involved in inducing internalization, like listerial internalins. Internalins bind to host signaling receptors like cadherins or Met. **(b)** Adherence in higher eukaryotes is made by membrane–protein interactions and protein–protein interactions. Tight junctions provide close connections and cell impermeability. Gap junctions also are tight connections between cells, which provide pores at their protein center through which small molecules from either cell cytoplasm can be exchanged. Hemi-desmosomes and desmosomes provide connections with basal laminar layers and with neighboring cells, respectively. Adherens junctions are just one sort of cell-adherence molecules (CAMs) connecting neighboring cells to one another

with increased selectivity in the permeability of small molecules and water. Taken together, the cell wall in prokaryotes and the cytoskeleton in eukaryotes act to counter the effects of turgor pressure, a major physical force affecting cells.

Another biophysical process that cells must undertake is adherence to a cell surface. Both prokaryotes and eukaryotes have developed a variety of mechanisms to confer specificity in surface adherence. Adherence provides bacteria with numerous functions, including: 1) host-cell invasion as the first step of pathogenesis; and 2) the transfer of genetic material between bacterial cells in a process known as conjugation (Fig. 19.4a). Among eukaryotic cells, which possess only an amoebic lipid bilayer to separate the cytoplasm from the extracellular space, surface adhesion is dependent on adherence molecules recognizing specific ligands. A ligand is a molecule that can form a complex with another biomolecule on the host cell, called the receptor. Adhesion molecules provide eukaryotic cells the ability to bind to surfaces, cells, or particles (Fig. 19.4), which play important functions in accumulating nutrients, cell growth, division, and differentiation (otherwise known as morphogenesis).

Prokaryotes are found in various environments, both inanimate and those teeming with life. In instances when bacteria encounter chemical or physical

stresses, they can utilize adherence as a tool to initiate a process of host-cell invasion wherein they access a new environment void of the stresses they previously encountered. Often, this process is associated with disease in host cells, even though the aim of the pathogen is to access an environment supportive of growth. Bacteria exhibit both nonspecific and selective methods by which to confer adherence. Among nonspecific adherence is the chemical characteristic of the cell wall. For example, mycobacteria exhibit mycolic acids on their outer surfaces, which are hydrophobic in nature. Another method of adherence employs fimbriae, thin strands of protein associated together in a filament that reaches several micrometers in length. Shorter versions of fimbriae are known as pili and provide cells selective attachment to bacteria and host cells, alike. The mechanics of pili interactions include both: 1) rigid bodies that provide bacteria tunnel-like structures through which to exchange DNA and small particles; and 2) contractile filaments, which are involved in bacterial–host cell interactions and ratcheting bacteria into close proximity and in an ideal orientation for host-cell invasion. Bacteria also exhibit receptors on their surfaces that bind as ligands to host-cell receptors and induce the process of host internalization. Many of these bacterial factors involved in host-cell interactions and internalization are considered virulence factors, as they induce a pathogenic process resulting in disease in the host organism.

Eukaryotes express a variety of adherence factors necessary for cell–cell, cell–matrix, cell–bacterial, and cell–antigen interactions (Fig. 19.4b). Tight-junctions provide cells the ability to hold together in tissues. Tight-junctions are sometimes called occluding junctions since they serve the purpose of sealing gaps between cells in order to make impermeable cell barriers. Structurally, tight-junctions are composed of proteins known as cell-adhesion molecules (CAMs). CAMs and focal adhesions function to bind cells together and cells to extracellular matrix (ECM). The ECM is a structurally rigid material composed of a mesh-like network of proteins (such as collagens) and sugars (polysaccharides) that are secreted by cells. Adherence between cells can also serve as a mechanism of inter-cytosolic communication through gap junctions. Another group of CAMs, adherens junctions and desmosomes, function to produce cell–cell specific interactions. Binding of these molecules induces receptors signaling a cascade of chemical reactions downstream leading to changes in gene regulation. This is a general concept by which cells can sense interactions and modify their behaviors. Together, these interactions provide cells the uncanny ability to make many types of connections with substances in their surrounding environment, ranging in size from small antigens on the order of several nanometers to eukaryotic cells of several hundred micrometers.

In this chapter we explore approaches taken to uncover some fundamental principles governing cell structure, cell growth, and cell behavior under conditions of stress and disease. We will highlight the use of micro- and nano-electro-mechanical systems (MEMS and NEMS) for making these fundamental discoveries in biology possible.

19.2 High Resolution Microscopy Methods for Live Cell Imaging

Due to the small size scale of most cells, high-magnification microscopy methods are required to study the subcellular structures. Additionally, the frequent need to observe the cells in their living state limits the number of applicable microscopy techniques. Atomic force microscopy (AFM) and optical microscopy techniques are both amenable to image living cells and often the sample preparation is similar for both methods. A sample prepared for optical microscopy imaging can often be used for AFM measurements as well. While the AFM enables one to characterize the surface, optical microscopy can map labeled structures of a transparent sample.

The combination of these techniques therefore enables access to the surface to measure the structure and mechanical properties of the sample, while optical microscopy can identify the biochemical nature of the sample, supplying complementary information. This combination has been implemented for some time and instruments combining an inverted optical microscope and an AFM are commercially available. Using optical microscopy in conjunction with AFM is the basis for many of the measurements of cell mechanics, cell adhesion, and the cell diagnostic and manipulation experiments described later in this chapter.

In combined systems, AFM normally gains access to the sample from the top, while the optical technique accesses the sample deposited on a transparent coverslip from the bottom, using high-numerical aperture objectives on an inverted optical microscope (Fig. 19.5). The optical microscope helps to screen the sample and to choose a region of interest before starting AFM experiments. Thereby, regions of the sample containing cells or biomolecules of interest can be located quickly.

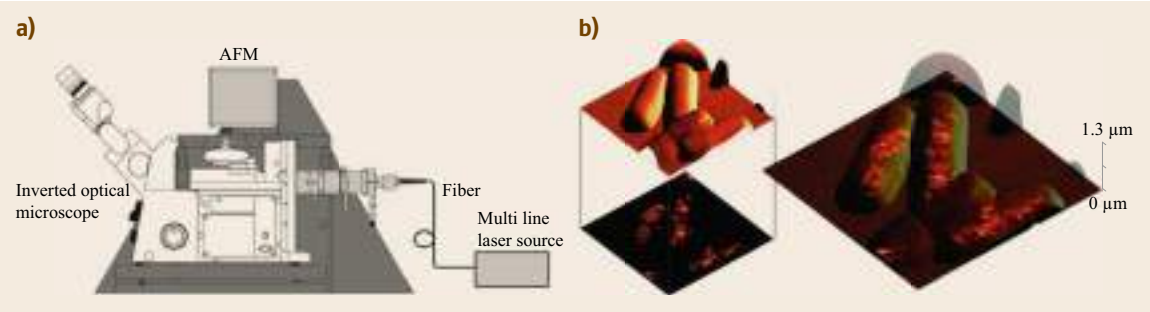


Fig. 19.5 **(a)** Schematic of a combined atomic force and optical microscopy setup. The inverted optical fluorescence microscope is set below the atomic force microscope. A large metal base stabilizes the AFM and sample above the inverted microscope so as to reduce mechanical noise. **(b)** Correlated AFM-PALM imaging of *E. coli* labelled with RNAP-mEos2 (after [19.3])

It also enables one to optically observe mechanical manipulations done by the AFM.

While AFM imaging on living cells can reach a resolution of $\approx 10\,\text{nm}$, conventional fluorescent microscopy suffers from the fundamental resolution limit described in [19.4]. The fundamental diffraction limited or optical resolution limit is dependent on the wavelength and the numerical aperture of the optical system, and is typically in the order of 150–250 nm, which is more than ≈ 10 times larger than the resolution achieved by AFM on living cells. Using advanced microscopy techniques, such as single molecule localization microscopy (SMLM) (e.g., photo-activated localization microscopy (PALM) [19.5, 6], (direct) stochastic optical reconstruction microscopy (dSTORM) [19.7], or stimulated emission depletion microscopy (STED) [19.8]), make it possible to decrease the resolution limit by a factor of ≈ 10. These super-resolution optical techniques thereby approach the resolution of the AFM. Thus, biological structures can be studied using both methods (AFM and super-resolution microscopy) at a similar resolution, providing complementary information about two different aspects of the same biological sample. The combined technique was demonstrated on various static biological samples including biomolecules, bacteria, and mammalian cells, using various super-resolution techniques, e.g., STED, (d-)STORM, and PALM [19.3, 9, 10]. Recently, a high-speed AFM combined with time-resolved PALM imaging even enabled the study of transient processes, such as focal adhesion dynamics in a correlated manner on living cells [19.3]. Such advances in correlated AFM/optical microscopy will greatly aid in areas such as mechanobiology.

19.3 Measuring Mechanical Properties on Living Cells

Quantitatively measuring mechanical properties of living cells has become of great interest, as recent results have shown a correlation between the mechanical properties of a cell and its function. Altered cell functions due to a disease can severely influence mechanical properties of the cell, or vice versa, and altered mechanical properties can lead to a change in the function of the cell or affect its behavior. Cancer, malaria, kidney disease, and many other pathological conditions have been attributed to adverse mechanical properties of cells [19.11–13]. Thus, quantitative information about the mechanical properties of cells could help to establish accurate diagnosis and early detection of a disease or to evaluate prognosis.

19.3.1 Using MEMS Cantilevers to Measure Cell Mechanical Properties

Probing mechanical properties requires applying a physical force to the cell and detecting its physical response. If the amount of force applied is known and the physical response can be measured precisely, the mechanical properties can be quantified. The AFM enables one to apply a specified force on the sample using a calibrated cantilever while constantly recording its position and deflection [19.14].

Most measurements of cell mechanical properties using MEMS cantilevers are done with AFM equipment, based on the recording of a force–distance curve (Fig. 19.6). For a desired position of the sample, the

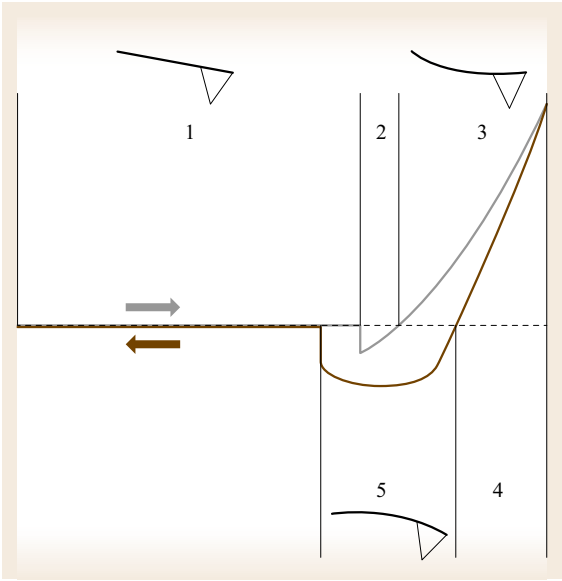


Fig. 19.6 Schematic representation of a force curve. 1) The cantilever begins at an extended above the sample surface. 2) The cantilever tip is subject to weak chemical forces and snaps onto the sample surface. 3) The cantilever deflects away from the cell surface as it is approached nearer the cell surface. 4) Once the force set-point is reached (at the apex of the y-axis), the cantilever is retracted away from the surface. 5) Once more, the cantilever is subject to adhesion and weak chemical forces before losing contact with the sample surface. The approach curve is labeled in *gray* and the retraction curve is labeled in *brown*

cantilever is brought to the sample, often using piezoelectric positioners, and the deflection of the cantilever is recorded as a function of the piezo-extension. The force-distance curve can be divided into five regions:

1. Region without contact (baseline)
2. Snap in with point of contact
3. Compressive loading of the material (indentation)
4. Relaxation
5. Adhesion.

Region 1 is used as the baseline, and its intersection with region 3 is often defined as the contact point. For perfectly elastic materials, regions 3 and 4 overlap, however with real materials the deviation of region 4 from region 3 gives information about the plastic or viscoelastic properties of the material. In region 5, the cantilever is bent in the other direction due to adhesion with the sample. Regions 1, 2 and 3 constitute the approach – regions 4, and 5 are the retract curve. The area between the approach and retract curve corresponds to the energy dissipated during one loading cycle. Different biologically relevant aspects of cell behavior can be extracted from the various parts of the force curve. In order to obtain quantitative measures of these aspects, both the MEMS cantilevers, as well as the AFM instrument, need to be properly calibrated.

19.3.2 Cantilever Calibration

The cantilever of an AFM can be characterized as a spring with an intrinsic spring constant, a resonance frequency, and a quality factor. According to the simplified Hooke's law, the spring constant is

$$k = \frac{F}{\Delta x} ,$$

which is the force required per unit of displacement (N/m). Two frequently used methods, the thermal noise method [19.15] and the Sader method [19.16], are generally employed to precisely calibrate the spring constant of an AFM cantilever.

In the thermal-noise method, the cantilever is treated as a simple harmonic oscillator (SHO). The spring constant k is calculated by relating the thermal (i. e., Brownian) motion of the cantilever's fundamental mode to its thermal energy [19.15]

$$k = \frac{k_B T}{\langle z_c \rangle^2} ,$$

where k_B is the Boltzmann constant, T the temperature, and $\langle z_c \rangle^2$ the squared mean displacement of the cantilever. Various corrections for the methods have been proposed [19.17–19].

The Sader method uses the in-plane dimensions of the cantilever, the density (ρ), and the viscosity (η) of the medium in which they are measured, and the cantilever resonance frequency f_R and quality factor Q to calculate the spring constant [19.16]. The thermal spectrum of a rectangular cantilever can be fitted with the SHO model to extract f_R and Q. The spring constant is then given by [19.16]

$$k = 7.5246 \rho b^2 L \Gamma_i(Re) f_R^2 Q \quad \text{with}$$

$$Re = \frac{\pi \rho b^2 f_R}{2\eta} ,$$

where b is the width of the cantilever, L the length of the cantilever, and Γ_i the imaginary part of the hydrodynamic function [19.18, 20].

19.3.3 Instrument Calibration

The deflection of the cantilever in most AFMs is measured based on the change of the angle of a laser beam

reflected off the backside of the cantilever. The change of the position of the reflected laser beam will change the voltage on a position-sensitive detector (PSD), which is determining its position. The signal from the PSD needs to be correlated to the actual cantilever-tip displacement by recording a force–distance curve on a hard surface. Because on a hard surface there is theoretically no indentation of the tip into the substrate, a *deflection sensitivity* can be derived from the linear portion of region 3 of the force–distance curve. This value can then be used to correlate a specific value read from the PSD with a specific cantilever deflection, and, in turn (using the spring constant of a cantilever), to a specific force.

Many factors influence the accurate measurement of cell's mechanical properties, leading to a large spread in the values reported in the literature. Therefore, in order for quantitative cell-mechanical properties measured by AFM to be comparable across different labs a standardized procedure for the calibration of the cantilever and the read-out system is required. In this regard, researchers across various labs in Europe have developed a procedure, called the Dubrovnik procedure [19.21].

19.3.4 Extracting Materials Properties from Cell–Indentation Measurements

From the properly calibrated and transformed force–distance curves, mechanical information about the sample can be extracted, such as the stiffness and the indentation depth. The Young's modulus is a measurement of a substances' resistance to being elastically strained when a stress is applied to it. It can be derived from the slope of the stress–strain curve in the elastic deformation region. However, the geometry of the indentor (AFM tip) has to be known for proper calculation of the Young's modulus. One way to describe the interaction between a spherical tip of known radius and a cell is the Hertz contact model of two elastic bodies [19.22–24]. The Hertz model is valid under the assumption of elastic deformation of two homogenous smooth surfaces in contact [19.25].

Assuming the Hertz contact model and that the tip used for probing is infinitely stiff compared to the cell, the relationship between the indentation δ, and the loading force F is given by

$$F = \frac{4ER^{1/2}\delta^{3/2}}{3(1-\nu^2)} ,$$

$$F = \frac{4}{3}E * R^{1/2}\delta^{3/2} ,$$

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} ,$$

where E is the elastic modulus of the cell, ν the Poisson ratio of the cell and R_1 and R_2 the radii of the cell and the cantilever tip, respectively.

Under the assumption that the cell can be described as a homogeneous and isotropic (similar in all dimensions) material, valid for small-indentation depth only, the cell can be approximately described by two numbers, for example, by the Young's modulus and the Poisson ratio. The Poisson ratio for soft materials typically ranges 0.3–0.5. Since the maximum error due to the unknown Poisson ratio is small, it is typically sufficient to characterize soft materials with just the Young's modulus [19.26, 27]. Thus knowing the force, the radii, and the indentation depth are enough to calculate the Young's modulus using the equations above [19.28].

Other models applied to mechanical measurements done by AFM include the Johnson–Kendall–Roberts (JKR) and Derjaguin–Muller–Toporov (DMT) models [19.29, 30]. The DMT model [19.30] is applicable to stiff contacts with small tip radii and low adhesion forces. The JKR model is best suited for situations with relatively low stiffness samples, large tip radii, and high adhesion forces [19.31]. It should be noted, however, that all the above-mentioned models have been derived for uniform materials. As described earlier, cells are very complex structures consisting of many different materials in a variety of geometries (cell membrane, cytoskeleton, etc.). Numbers derived from these models should therefore be understood as *equivalent Young's modulus*, and we use the term Young's modulus as such in this chapter.

As the measurements of Young's modulus is dependent on the loading rate due to the viscous contribution, measurements of mechanical properties have been done at rather slow speeds, minimizing the viscous effect, with typical rates of 1–2 Hz. By performing multiple force curves in a raster pattern, the topography and mechanical properties can be recorded simultaneously (force–volume mode [19.32]). To speed up the measurement acquisition, several *fast-force curve* modes have been developed where hundreds or thousands of ramps are performed per second. To deal with the increasing speed, deviations from the linear trajectory have been introduced, and the triggering on the peak force has been replaced by a constant feedback loop on the measured peak force (Quantitative Nanomechanical Mapping (PF-QNM), Bruker, HybriD mode, NT-MDT) [19.33, 34].

For a small scan area within bacterial cells it was found that elastic models remain relevant for indentation frequencies below 20 Hz, while above this frequency incorrect apparent Young's moduli were calculated [19.28]. Additionally, an effect of penetration depth on the modulus values was observed as well.

A too-low penetration depth (< 20 nm) leads to an increase of measured modulus, as the hard cell wall is tested. On the other hand, too high penetration depths also dramatically increased the elastic moduli due to the hard-substrate effect and incompressibility of the cellular membrane. This introduces artifacts for very thin cells adhered to substrates, which is for example the case for fibroblasts (connective tissue cells) [19.35]. However, even thin materials can still be reliably measured with conically shaped tips by applying a correction factor, namely, the bottom-effect cone correction (BECC) [19.36].

Further instrument development, standardization of experimental procedures, and improvements in data analysis will make it possible to measure quantitatively mechanical properties of biological samples, which can be compared across different laboratories and instruments. This is particularly important to be able to use mechanical-property measurements of biological samples for diagnostic or prognostic purposes.

19.4 Cancer Detection Using Nanomechanical Sensors

> *Cancer is the uncontrolled growth of cells, which can invade and spread to distant sites of the body. Cancer can have severe health consequences, and is a leading cause of death.* [...] *Early detection, accurate diagnosis, and effective treatment* [...] *help increase cancer survival rates and reduce suffering* – World Health Organization [19.37]

The World Health Organization states that the number of new cancer cases is expected to rise by about 70% over the next two decades, while already now 8.2 million people die each year from cancer. The major cause of death from cancer is metastases. In order for metastases to occur, the cells have to complete a series of steps:

1. Normal benign cells have to undergo transformation acquiring a cancerous state that leads to uncontrolled growth.
2. The primary tumor initially nourished by diffusion requires the development of new blood vessels in order to grow beyond 1–2 mm in diameter [19.38–41].
3. Tumor cells then invade the connective tissue surrounding the organ [19.42], detach from their primary site and penetrate the blood vessels (intravasation) [19.43].
4. The cells are then disseminated via the blood vessels and become trapped in the capillary beds of distant organs by adhering to the capillary walls [19.44].
5. Cells then penetrate the capillary walls and invade the tissue at the distant site by crossing the extracellular matrix barrier.
6. Finally, cell growth and division within the distant organ completes the metastatic process (Fig. 19.7).

To continue growing, the micrometastasis must develop a vascular network and evade destruction by the host's defenses. The cells can then invade blood vessels, enter the circulation, and produce additional metastases [19.38, 45–49]. Thus, in order to successfully metastasize, a cell faces many challenges on its journey. During entry into, and exit from, the vascular system, tumor cells undergo dramatic shape changes driven by cytoskeletal remodeling [19.50]. Such remodeling of the cytoskeleton affects the mechanical properties of cells [19.51, 52]. A reduction in elastic modulus of cells therefore could facilitate substantial, rapid, and reversible migration of tumor cells through small pores that are often present in tumors infiltrating blood vessels [19.11, 53, 54]. Measurements of the mechanical properties of cells could be used as indicators of their biological states and could offer valuable insights into the pathogenic basis of disease [19.11, 53, 55–57]. Therefore, effort has been put into understanding how changes in the biomechanical properties of cells influence, and are influenced by, the onset and progression of human disease. Several techniques have been developed in order to assess the mechanical properties of cells or tissues. These include micropipette aspiration [19.58–61], magnetic-bead twisting [19.62], microfluidics [19.63], and optical tweezers [19.64] or atomic force microscopy (AFM) [19.65, 66]. The main advantage of these techniques lies in their capability to measure living cells on a single-cell level and in liquid conditions, imitating their natural environment [19.67]. It has been found that cancerous-breast epithelial cells are more deformable than nontumorigenic epithelial cells, and that the metastatic variant is even more deformable [19.68]. Various cell-model systems have been applied to correlate the mechanical properties of a cell with their metastatic potential. For some cells, it is possible to regulate the metastatic potential by modulating the expression of specific genes, such as breast cancer metastasis suppressor 1 (BRMS1) or Syk, a protein involved in many different cell-signaling cascades in breast cancer cells [19.69, 70]. Cells expressing BRMS1 or Syk generally exhibited lower metastatic

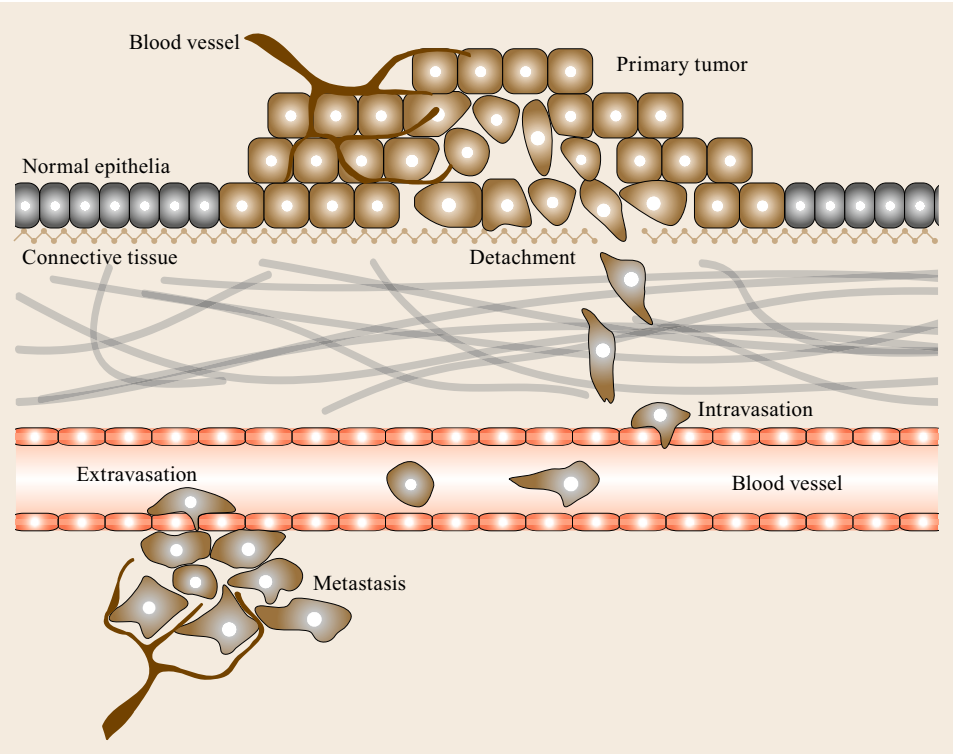


Fig. 19.7 Graphical representation of cancer cells, their spread, tissue invasion, and tumor formation

potential, while expression of BMP2 [19.71] promotes migration and invasion. Specifically, these cells exhibited increased cell adhesion, cellular spring constant, Young's modulus, and viscosity [19.72]. Additionally, an altered cellular topography, including reduction in cell height and a more substantial and stable microtubule network, was observed [19.51]. These results show that a number of relevant mechanical properties change as the cell undergoes transformation from a benign to malignant state. Similarly, using colloidal AFM probes, it was found for a variety of ovarian cell lines that stiffer cells exhibited a lower migratory and invasive potential than the softer cells. A decrease in cell stiffness was found to correlate with a low degree of coalignment of cytoskeletal microfilaments [19.73]. This suggests that the structure of the cytoskeleton affects the mechanical properties, which in turn influence the migratory and invasive potential of cells, and, in consequence, the metastatic potential [19.74].

In accordance with these results, a decrease of the elastic modulus was found upon drug-induced destabilization of actin filaments of noncancer cells [19.75, 76]. Transformed fibroblasts, besides being softer than the nontransformed cells, showed disorganization of the actin cytoskeleton, lack of actin stress fibers, and decline in the number and size of focal adhesions [19.77]. Similarly, in a study comparing three epithelial breast-cancer cell lines with various degrees of malignancy, it was found that local stiffness is at a maximum where filamentous actin is organized as well-aligned stress fibers. A difference was also found in the spatial distribution of actin stress fibers. In tumorigenic cells, actin stress fibers appeared only at basal regions (facing the substrate), where they cannot contribute to the stiffness probed by AFM. In contrast, in healthy cells, actin stress fibers were present at apical regions (facing away from the substrate), where they contribute to the stiffness measured by AFM [19.78].

Using patient samples, the group of Gimzewski measured by AFM the elastic moduli of metastatic cells obtained from body (pleural) fluids from patients with suspected lung, breast, and pancreas cancer. Compared to benign cells, metastatic cells were 70% softer. The study showed that a shift in biologically-driven biomechanical properties towards a decrease in cell stiffness correlates with an increase in metastatic potential [19.12]. Demonstrating that nanomechanical characterization could be applied to patient samples to reliably provide information about the progression of the disease makes this biomarker an additional tool with diagnostic potential.

However, whether the elastic modulus is a general biomechanical marker for cancer progression is not entirely clear. A comparison of bladder-cancer cells

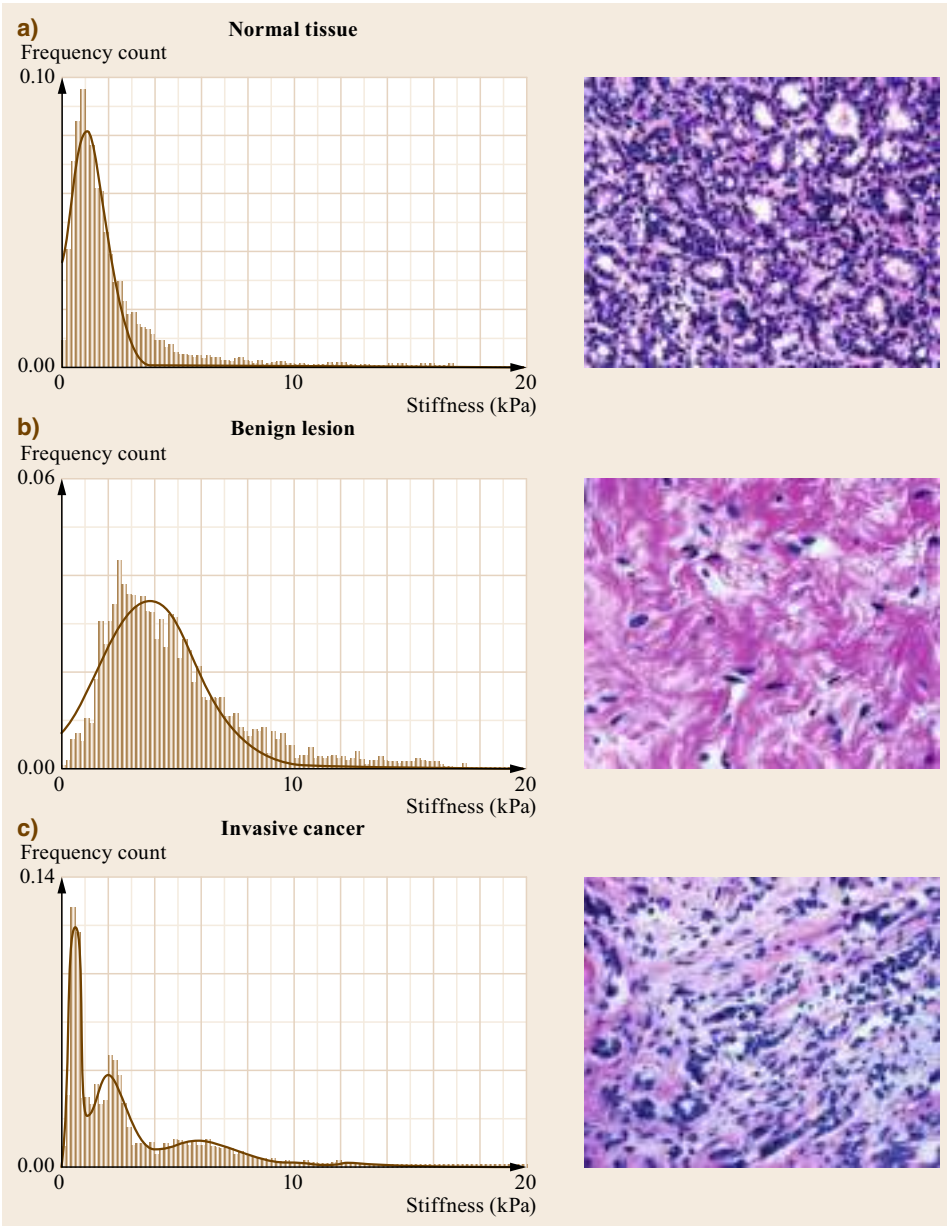


Fig. 19.8a–c Graphical representation of the mechanical properties measured on the surface of breast-cancer tissue sections. On the *left* are graphs of the distribution in the mechanical properties of **(a)** normal tissue, **(b)** benign lesions, and **(c)** invasive cancers. On the *right* are tissue slices with dyes of cell nuclei (*purple*) and proteins (*pink*) (after [19.79])

from various histological stages of cancer progression showed that cancerous cells are more deformable than nonmalignant cells. But, this was regardless of the state of cancer progression. Soft cells thus indicated a malignant phenotype. But, the histologically-determined tumor grade does not seem to be directly reflected in the elastic properties of the cells. Cancer cells from advanced stages of tumor progression were not softer than cancer cells from earlier stages [19.80]. This suggests that just determining elastic properties of individual cells might reflect a qualitative state of the cell in relation to cancer, but might not enable one to link it to the severity of the disease. Some researchers therefore suggested that stiffness is a dynamic property, and proposed that, rather than a general decrease in cell stiffness, the cells' ability to vary their degree of stiffness towards both very low and very high values as a marker of malignancy [19.81].

Stiffness at the tissue level might as well be different from the stiffness measurements presented above, which almost exclusively focused on measurements of single isolated cells. Paradoxically, and counterintu-

itive to the results obtained on single cells, cancerous tissues are often found to appear stiffer than normal tissue, for example, in breast-cancer tissue [19.82, 83]. A comprehensive study of breast-cancer biopsies, correlating histopathological information with nanomechanical measurements on a tissue-scale using AFM, has been reported by *Plodinec* et al. [19.79]. Nanomechanical stiffness measurements were done on breast-cancer biopsies of ex vivo human and mice samples under physiological buffer conditions. It was found that healthy tissue exhibited a uniform stiffness profile, characterized by a unimodal distribution with a single distinct peak (Fig. 19.8). Similarly, while the stiffness of tissue samples containing benign lesions remained uniform, the values were higher than in the normal tissue. The biopsies of tissues with cancerous lesions, on the other hand, exhibited a distribution of stiffness values that was much broader and contained a long tail at higher values. A dominant peak was measured at a value at least a factor of two softer than healthy mammary-gland epithelium, which corresponds to regions densely populated with cancer cells. This suggests that individual cancerous cells in the tissue are softer, confirming the results obtained on single cells. However, the nanomechanical signature of cancer tissue included a second peak at higher values, which corresponds to fibrotic regions (regions with extensive extracellular matrix material).

Taken together, all these nanomechanical cues suggest that abnormal mechanical properties are general traits of cancerous cells. Efforts are under way to utilize this in clinical research and diagnostics. Additionally, there is still much room for fundamental research and technology development in this area.

19.5 Measurement of Cell Adhesion Using Cantilever Based Force Sensor

Specific molecular recognition as well as nonspecific adhesion, are important events regulating and influencing a vast range of biochemical processes and biological behaviors in living organisms [19.84]. Cell adhesion molecules are involved in regulating fundamental processes in multicellular organisms, such as embryonic development, binding of white blood cells to the walls of blood vessels, and neuromigration. These diverse processes are enabled by a multitude of various adhesion molecules present on cell surfaces [19.85].

Cantilever-based force spectroscopy can be used to measure these adhesions. It is quantitative and has a very-wide force range from 10 pN to 1 μN [19.86]. This excellent force resolution, coupled with the positional precision, enable the use of this technique for measurements at the single-molecule level. Small adhesion forces are exhibited between biomolecules, such as receptor-ligand association, DNA entanglement [19.87], antigen-antibody complexes [19.88], and cell-integrin binding [19.89]. Besides single-molecule measurements, the binding strength between individual cells or between a cell and a substrate can be measured under physiological conditions [19.85, 90].

19.5.1 Tip Functionalization Methods

A requirement for such an experiment is the functionalization of the cantilever. One part, e.g., the ligand, is chemically tethered at the tip of an AFM cantilever. This cantilever can then be brought into contact with the other part, e.g., the receptor, which is adsorbed on the substrate or present on the cell membrane. The AFM cantilever motion can be precisely controlled, allowing one to modulate parameters such as contact time, contact force or speed of approach/retraction.

A force–distance curve can be obtained by recording the deflection of the cantilever during such a contact/release cycle and applying Hooke's law to the cantilever deflection (Fig. 19.6). The cantilever starts deflecting when it touches the surface, and e.g., the ligand and receptor start to interact consequently. The position can then be held constant during a certain length of time before the cantilever is retracted again from the surface and the deflection reverses, until the force applied on the surface is again zero. If there is an interaction between, e.g., the ligand on the cantilever and the receptor on the surface, the deflection of the cantilever increases in the opposite direction because it is held down by the molecular interaction. At some point, the pulling force is higher than the molecular interaction, and the connection will break. The profile of the adhesion and the strength of the interaction can thus be quantified by analyzing the deflection during retraction of the cantilever [19.91]. Parameters that are conventionally extracted from such data are the maximum force applied during the detachment process, the work of detachment, the distance of detachment, and the contour length [19.92].

The required cantilever functionalization can be achieved by either coating or grafting the material on the pinpoint of the tip [19.91] or attaching a functionalized bead on a tip-less cantilever [19.94]. Force-spectroscopy experiments using a functionalized tip have successfully been performed to study many spe-

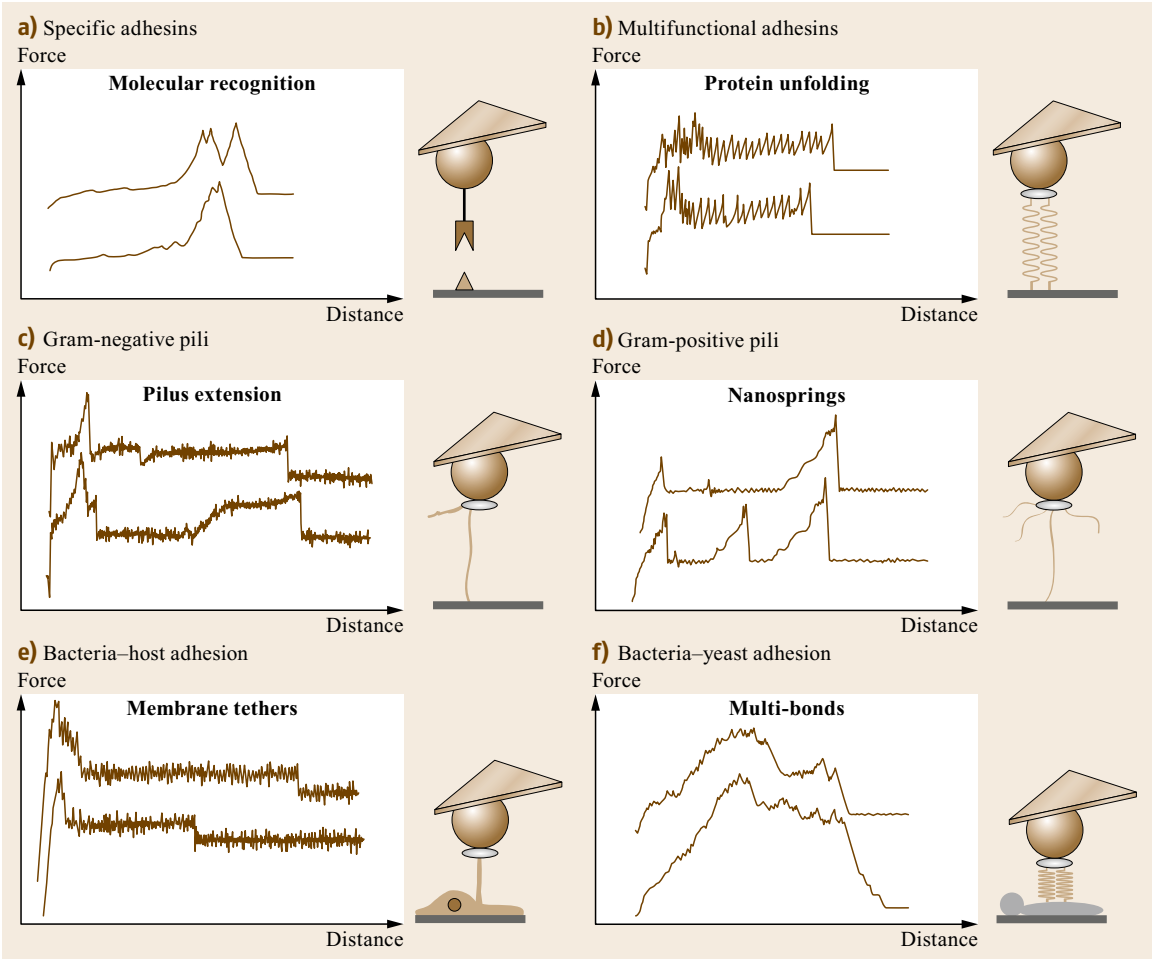


Fig. 19.9a–f Schematic representation of a variety of cell-adhesion experiments and their representative force spectrum (after [19.93])

cific interaction, ranging from the interaction between avidin and biotin and antibody and antigen [19.95] to the interaction between the individual-cell adhesion proteoglycan (AP) chains of the marine sponge *Microciona prolifera* as a model [19.84] (Fig. 19.9).

To measure cell–cell or cell–surface interactions, cells can be directly immobilized on the bottom side of a cantilever. To pick up single mammalian cells, cantilevers are previously functionalized with biomolecules that bind factors present on the cell surface, such as Concanavalin A (ConA) [19.96]. The living cell on the cantilever can then be used as a bioprobe allowing a wide range of experiments, where variables such as the surface decoration, the contact time, and the proteins expressed at the surface can be modulated [19.92, 97–100].

Various tip functionalizations have been reported to attach bacterial cells on cantilevers, including the use of poly-L-lysine (PLL), polyethyleneimine (PEI), lectin, silane, ConA, and a bioinspired-polydopamine wet adhesive (Cell-Tak) [19.85, 86, 101–106]. Bioinspired polydopamine was used to measure adhesion between single bacteria or fungi and metal surfaces [19.107], or using functionalized colloidal probes to measure the adhesion forces between single bacteria and biotic (lectin monolayer) or abiotic (hydrophobic monolayer) surfaces [19.108, 109].

19.5.2 Bacterial Adhesion

Bacterial adhesion on biomaterial surfaces is the initial step in establishing infections and leads to the formation of biofilms [19.110]. *Boks* et al. reported that there are likely differences between adhesion to hydrophilic or hydrophobic surfaces [19.103]. The authors studied the time-dependent adhesion of four strains of *Staphylo-*

coccus epidermidis to hydrophobic and hydrophilic surfaces. They found that, over time, strengthening of the adhesion forces was virtually absent on hydrophobic-coated glass, while on hydrophilic glass strengthening of the adhesion occurred within 5–35 s. This suggests that the hydrophobic effect causes instantaneous adhesion, while strengthening of the bonds on hydrophilic surfaces is dominated by noninstantaneous hydrogen-bond formation.

Various experiments have been reported demonstrating how the contribution of various bacterial biomolecules or structures to the adhesion can be assessed. Type IV pili (T4P), for example, are very thin protein filaments produced by a large number of bacteria. They play key roles in a wide variety of processes, including attachment to surfaces, bacteriophage attachment, twitching motility on surfaces and virulence [19.111–114]. Results suggest that adhesion via T4P is mediated by a low-affinity, high-avidity type of attachment to surfaces where nonspecific adhesion occurs along the length of the pili. Consistent with this model, force plateaus were observed that were attributed to the forced desorption or peeling of extended lengths of T4P from both hydrophilic and hydrophobic substrates [19.115, 116].

In *E. coli*, mutational studies indicate that type I pili play a critical role in biofilm formation [19.117]. Adhesion between type I pili expressing bacteria and various surfaces (fluorinated, aminated and PEG-like monolayers, silicon wafer, and mica) revealed a distinct adhesion profile for each case in terms of both force and distance, highlighting the cells' ability to adapt their adhesive properties to disparate surfaces [19.118].

19.5.3 Cell–Substrate Interaction (Biomaterials/Tissue Engineering)

Adhesion of cells to a substrate is a critical step in the successful integration of biomaterials for medical applications [19.119]. Bioactive glasses, an oxide-based biocompatible ceramic, is routinely used for orthopaedic and dental purposes [19.120]. Osteoblasts (bone-synthesizing cells) deposit new bone on the surface of the bioactive ceramics [19.121], which gradually replaces damaged or diseased bones in the body [19.122]. Gluing various bioactive glass microspheres on cantilevers and placing them in contact with the osteoblasts grown on a culture dish can measure adhesion between osteoblasts and bioactive glasses. Such colloidal probe experiments enable the study of the adhesion of cells to biomaterials of varying composition and the dynamics of the adhesion over time. It was found that osteoblasts reacted most rapidly and strongly with calcium borate glasses [19.122–127]. In a similar experiment with chitosan microspheres, osteoblasts exhibited particularly-high adhesion forces compared to other cell types. Chitosan is a widely-used biomaterial with reported osteo-conductive and osteo-inductive properties [19.91, 128, 129].

Besides overall cell adhesion to biomaterials the effect of specific adhesion-mediating proteins expressed on the cell surface can be studied in detail using a similar approach. By attaching single cells expressing certain integrin receptors that are specific for, e.g., fibrillar collagen type I on the AFM cantilever, the adhesion dynamics can be assessed. Integrins are a family of receptors involved in cell–cell adhesion and cell attachment to the extracellular matrix [19.130, 131]. After a contact time of > 60 s, overall adhesion strongly increased with the smallest, discrete rupture events, increasing above single-integrin levels. This suggests the establishment of cooperative integrin-receptor binding after an initial weak single-integrin-mediated binding event [19.100].

Conversely, by detaching cells from their substrate, the adhesion of living cells grown on various substrates can be studied. From the cantilever deflection during the detachment process [19.86, 132, 133], parameters, such as percentage of cell detachment, maximum force of detachment, distance of detachment, and work of detachment, can be analyzed. This method enables the study of long-term adhesion and the effects of surface topography [19.92].

19.5.4 Cell–Cell Interaction

Specific recognition of cell-surface molecules between single cells is essential for multicellular development. Combining single-molecule force spectroscopy with genetic manipulation, Benoit and co-workers quantified de-adhesion forces at the resolution of individual cell-adhesion molecules using eukaryotic cells of *Dictyostelium discoideum* [19.85]. These cells expressed only one particular type of adhesion molecules (contact site A (csA)). This way, the expression of a gene was linked to the function of its product in cell adhesion on a single biomolecule interaction level using whole cells [19.134].

Cell–cell interactions also play a major role during the immune response. One of the first steps, in recruiting white blood cells to a site of infection, is the specific interaction between factors expressed on the blood-vessel wall and receptors present on the membrane of circulating white-blood cells [19.135]. This specific interaction is dependent on various sets of cell-adhesion molecules [19.136–138]. AFM shows that the adhesion between white blood cells and endothelial cells is a complex process involving several step-like rupture

events and the viscoelastic deformation of cells on the scale of several microns [19.139].

Cancer research is another area where adhesion forces are of particular interest. Cancerous cells need to detach from their neighboring cells in order to metastasize [19.140]. An altered expression profile of adhesion molecules has been reported in breast cancer cells [19.141, 142]. Epithelial cadherin (E-cadherin), a cell adhesion molecule responsible for providing firm adhesion between two epithelial cells, is downregulated, while N-cadherins are upregulated. The loss of E-cadherin facilitates cell disconnection and invasion of malignant cells [19.143, 144]. Single-cell force spectroscopy revealed that the unbinding force between two E-cadherin molecules is much higher than that of two N-cadherin molecules [19.145, 146]. Thus, the low expression of E-cadherin and its replacement by N-cadherin generate lower cell–cell adhesiveness and assist in the detachment of cancerous cells from their primary tumor. Force-spectroscopy measurements showed that cells with increased invasion potential exhibited decreasing adhesion force, demonstrating the role of adhesion in metastasis [19.147].

Measurements done using the detachment approach of cells from their substrate, with subsequent analysis of the deflection during detachment, suffer from some major limitations [19.148]. Each cantilever has to be functionalized in order for cells to firmly attach. Strong attachment often requires one to stay in contact with the cell for tens of minutes, which is a time-consuming step. What's more, this procedure might even require changing the cantilever after a successful detachment of a single cell. To get statistically relevant data, higher throughput methods are required. Recent advances in MEMS fabrication offer routes towards such methods.

19.6 Manipulation of Cells Using MEMS Cantilevers

Among the various MEMS techniques available for imaging and manipulating cells, fluid-force microscopy (FluidFM) is a recent development. FluidFM relies on the conventional capabilities of an atomic force microscope by which a cantilever is resonated above a surface in order to measure the height. However, where FluidFM goes beyond conventional AFM is in the use of a cantilever that has a hollow fluidic channel connecting a reservoir of fluid to the cantilever tip, where fluid can be dispensed, making it a micropipette. The ability to dispense fluid through this hollow cantilever makes possible precise targeting of single cells and sample regions to administer various chemical treatments in an otherwise controlled environment. The functional use of FluidFM can be categorized into four distinct applications:

1. Measuring surface adhesion
2. Displacing cells and cell sorting
3. Fluid delivery
4. Access to the cell interior through injection/extraction (Fig. 19.10).

For cell-adhesion measurements, conventional AFM relies on the specific interactions between a chemically functionalized cantilever and the receptor of interest. Such functionalization is tedious and limits the throughput to one cell per cantilever. Using hollow cantilevers, however, controlled pressure in the channel can be used to pick up cells mechanically. The cells can then be tested against the target surface and subsequently released from the cantilever to make room for the next cell, enabling high-throughput experiments [19.145]. Such studies have been conducted on eukaryotic samples including the yeast, *Candida albicans*, and human cell lines [19.149, 150].

The ability to pick up cells by suction control also means that cells can be manipulated spatially, wherein cells in disparate locations can be made to interact with one another at close range. This technique has been demonstrated with bacteria, yeast, and mammalian cells [19.151, 152]. These manipulations provide an interesting perspective for isolating specific cells of interest for further analysis. To that end, bacteria exhibiting specific phenotypic traits could be isolated and genotyped by 16S rRNA gene amplification and sequencing at a later time [19.153]. Recently, one report demonstrated the possibility that eukaryotic cells could be probed for analysis of specific biomolecules produced in the cell cytoplasm and nucleus [19.154]. The possibility to probe cells at specific sites makes possible the determination of the composition and behavior of cells in ways that are otherwise not feasible or destructive to cell physiology.

The delivery of molecules to cells by a micropipette has been used to control the dosage and location at which a cell was treated with a bioactive molecule [19.152]. To date, it has been demonstrated that a membrane-impermeable dye could be injected into the host cytosol and nucleus [19.155, 156], whereby intracellular material and space can be marked with markers normally not present inside

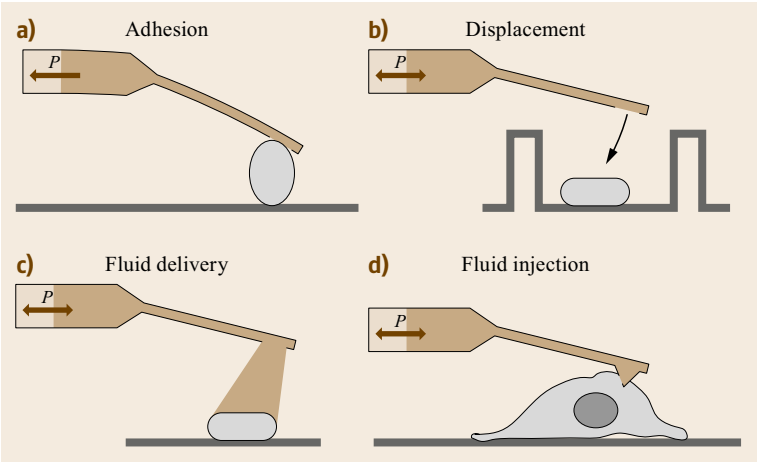


Fig. 19.10a–d Schematic representation of the versatile uses of FluidFM. **(a)** FluidFM can be used to adhere cells by aspirating the cell with the hollow cantilever tip by the use of the pressure controller. **(b)** Cell can be displaced space by aspiration, displacement of the cantilever and/or the stage, and by out-pressure to break the adherence between the cell and the tip. Finally, fluid can be delivered to the outside of the cell **(c)** or injected into the cytoplasm or nucleus of the cell **(d)**

the cell. Future study may aim to inject controlled amounts of nucleic acids or other exogenous materials to study cell processes with single-cell resolution that have only otherwise been studied using bulk measurement techniques on hundreds and thousands of cells.

Another field that may benefit from the controlled delivery of biomolecules to living cells is host–pathogen interactions, namely, in the controlled release of a virus above human cells [19.157]. Many of the physical principles underlying the process of viral infection remain poorly understood.

19.7 Mass Measurements of Single Cells Using Resonant Cantilevers

Determining the characteristics of whole-cell physiology at the single-cell level requires techniques sensitive to subtle and dynamic changes. While many behavioral and physiological phenotypes can be quantified at the whole-cell level, one such physiological attribute of cells that is of particular interest in biology is cell size. Fundamental behavioral processes, like growth and division, regulate cell size. How *cell size* is quantified is determined by the tools available to measure it. Optical microscopy is used to measure cell length and width. Together, these measures can provide an estimate of cell volume [19.158]. However, cell size is also a representation of the molecular contents which together are organized into a cellular compartment. One way to measure the full contents of a cell is to quantify its mass. Here, resonant mechanical sensors provide a unique opportunity.

19.7.1 Resonant Mechanical Sensors

Beam-based mass sensors comprise cantilevers (singly clamped), bridges (doubly clamped), and more complex geometries such as membranes. Cantilevers can have various geometries, including simple rectangular beams ranging from hundreds of micrometers in length down to single micrometers [19.159]. The minimum detectable mass is related to the ratio of mass and resonance frequency of the beam [19.160]. Beam size, geometry, fluidic environment, and the quality factor (Q-factor) of the beam material also play crucial roles. Manipulating each of these factors impacts the sensitivity of the beam for measuring the mass of a loaded sample. Cantilevers can be used in various ways to measure cell mass. They use the principle that cells attached on the cantilever affect its resonance frequency [19.161] (Fig. 19.11). The shift in resonance frequency is related to the mass of the sample, as well as where cells are located on, or in, the cantilever [19.162]. Functionalizing the cantilever surface allows for the adherence and immobilization of cells in specific sites along the beam [19.163]. Cantilevers can be functionalized with antibodies raised against specific bacterial receptors [19.161] or with agarose, which acts as an indiscriminate immobilization platform [19.163]. An alternative method to use cantilevers for cell-mass measurement is to flow cells through a hollow microchannel resonator within the beam. This has additional advantages, as the mass sensitivity is dependent on the resonance frequency and the Q-factor. Having the liquid medium on the inside of the cantilever and surrounding it with a vacuum make possible much higher resonance frequencies and Q-factors than achievable for cantilever

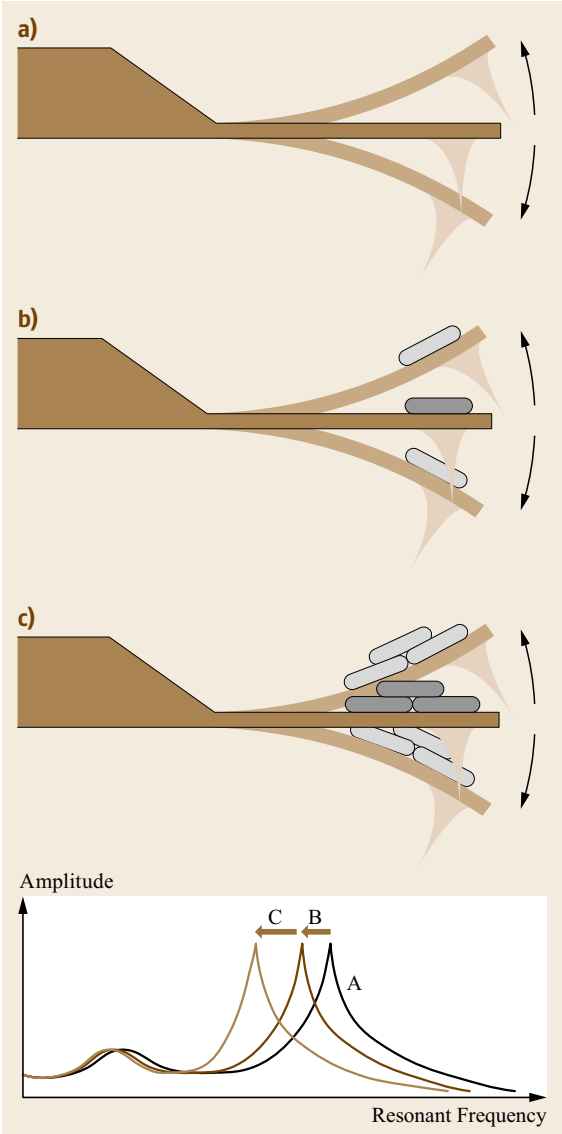


Fig. 19.11a–c Mass measurement of particles/cells adhered to the cantilever-tip surface. The resonant frequency of a cantilever is shifted depending on the mass loaded on the cantilever-tip end, while the amplitude remains constant. **(a–c)** depict schematic representations of how the cantilever-tip body can be loaded with various quantities of mass. The graph *below* depicts how the addition of increased mass shifts the resonance frequency peak downward

resonators immersed in cell medium [19.164]. These suspended micro-channel resonators (SMRs) have the benefit that the medium, in which single cells can flow back and forth through the micro-cantilever channel, can be exchanged rapidly to detect very minute changes in cell physiology (Fig. 19.12). Together, these methods for mass measurement provide unique opportunities to characterize cell growth in terms of added cell mass. Hereafter, we describe the ways in which mass measurement is used to uncover fundamental principles of cell physiology and behavior.

19.7.2 Cellular Growth

A critical factor regulating cell size is growth. Cell growth is dependent on the addition of new cell-wall material, as well as the accumulation of cytosolic components. The rate at which each of these processes function in relation to one another has a direct impact on cell morphology. Specific protein systems are involved in synthesizing components of which the various wall layers are comprised. Peptidoglycan, lipid, and sugar molecules together contribute to maintaining cell impermeability and morphology. The space in the bacterial-cell cytoplasm is comprised primarily of water, ions, DNA, and proteins. The accumulation of material within the cytoplasm is dependent on the influx of water, ions, and small molecules, and their conversion into the synthesized biological factors necessary for life. The rate of synthesis within the cytosol directly impacts the rate of mass accumulation. While volume can change reversibly under various switches in physiological environment [19.165], mass accumulates as a function of the retention and synthesis of materials necessary for sustaining and propagating life.

One method by which to measure the *instantaneous* growth rate is to quantify the cell mass, which represents the amount of material from which a cell is constituted. Researchers have used agarose-coated resonant cantilevers to measure growth dynamics at the micro-colony level, where they observed the same growth dynamics as do spectrophotometry measurements in bulk cultures of bacteria [19.163]. The growth of yeast cells is one example of a more complex organism that has been studied [19.166]. Together with previous bacterial growth studies, it was reported that, as in bacteria, a micro-colony of yeast grows in multiple phases, including a lag phase, an exponential growth phase, and a plateau [19.166, 167]. Another example that MEMS resonators can be adapted to measuring cell mass of higher organisms is the mass measurement of human cell lines [19.168]. This last study employed arrays of MEMS resonator sensors, by which high-throughput, single-cell studies can be achieved on a population of cells. Whereas in this last study they used arrays of cantilever beams, they later reported the use of four beam-spring sensor arrays to record human cell growth for over 50 hours, a relevant time scale for following 2–3 generations of cell growth

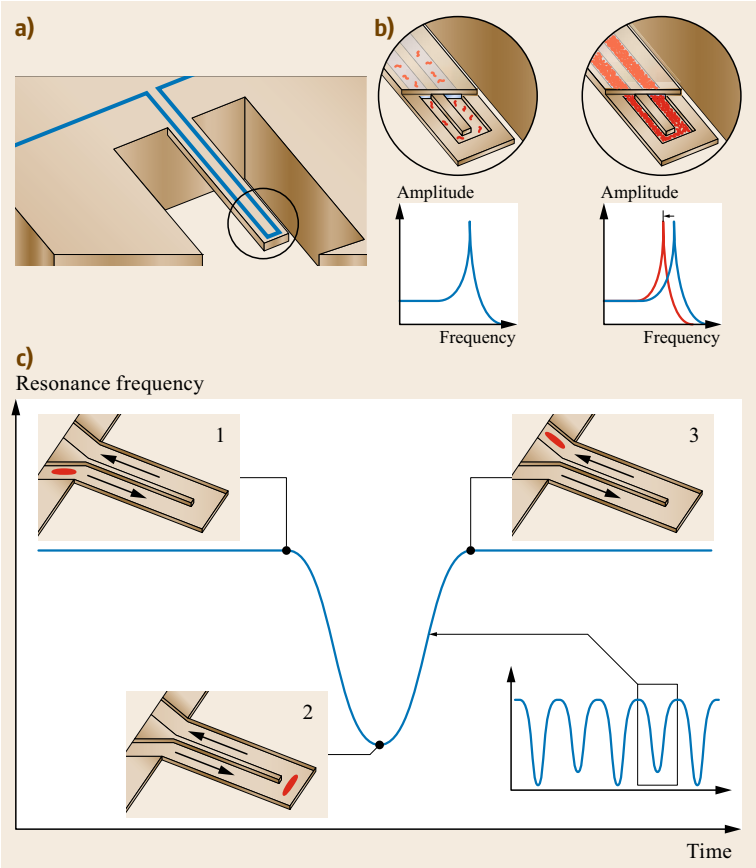


Fig. 19.12a–c Buoyant mass detected by variations in the cantilever resonance frequency. On **(b)** is depicted a schematic representation of a SMR with the cantilever body in *grey* and the hollow tunnel filled with a medium in *blue*. On **(a)** are depicted two conditions in which few or many particles/cells are flowed through the SMR. An increase in buoyant mass is concurrent with a down-shift in the resonance frequency of the cantilever (after [19.164])

and division [19.169]. Taken together, these micromechanical resonators can be employed to determine cell mass by measuring the change in resonance frequency upon immobilization and growth of cells on the resonator.

Another method used to measure cell mass and *instantaneous* growth is to transiently pass cells through the channel of a SMR [19.170]. Compared to the previous methods that measure the total mass of the cell (including the water inside), SMRs measure the buoyant mass. Buoyant mass is described as cell volume multiplied by the difference in cell density and fluid density

$$m_{\text{buoyant}} = V(\rho_{\text{cell}} - \rho_{\text{fluid}}) \, .$$

For the most part, the density of the cell components is greater than water, with the exception of some solubilized small molecules. By measuring the same cell in fluids of various densities, both cell volume and the density of the cell can be calculated [19.171]. The accumulation of buoyant mass was reported to increase exponentially for bacterial cells and eukaryotes alike [19.172]. For yeast cells treated with cell cycle inhibitors, it was observed that the change in the density of the cell occurred predominantly before bud-formation, whereas cell volume and buoyant mass increase throughout the cell cycle [19.173].

Buoyant mass has also been used as a tool to distinguish various species of bacteria [19.174, 175]. In addition, it has been used to distinguish various stages in bacterial cellular differentiation, like the process of sporulation in *Bacillus subtilis* [19.176]. More subtle changes in buoyant mass are also detectable, like the distinction between *E. coli* mini-cells that lack DNA and are the product of highly asymmetric divisions associated with chromosome missegregation [19.177]. Taken together, micromechanical resonators provide powerful tools to study cell growth by measuring cell mass, volume, and cell density. These studies, if combined with optical microscopy studies of cell elongation can provide a powerful and more accurate characterization of the principles underlying cell growth.

19.8 Characterization of Antibiotic Action Using MEMS Devices

Infectious disease imposes a major burden on human society. Treating infectious disease relies heavily on drug therapies whereby antibiotics clear the pathogen in order for the body to repair or replace diseased tissue. Antibiotics are a diverse group of substances aimed at eliminating host diseases or infections. Disease clearance can be measured by pathogen count (mainly measured in colony-forming units (CFU)). The impact of antibiotics on bacterial pathogens can be classified in one of three ways:

1. Bacteriocidal (bacterial killing)
2. Bacteriostatic (bacterial count remains constant)
3. Resistance.

Characterizing the impact of antibiotics on pathogens is a critical field of interest for evaluating drug efficacy. Concurrently, identifying drug targets is important for understanding the mode of action. Estimates suggest that nearly 10% of drugs have no known assigned target [19.178]. Concurrently, $\approx 20\%$ of drugs lack a known mechanism-of-action [19.179]. What remains is simply the knowledge that these drugs possess a positive effect for treating disease [19.180]. Determining how such drugs provide a protective or therapeutic function against infection is an expensive and uncertain endeavor [19.181]. Emerging MEMS-based assays hold much promise in this area.

19.8.1 Characterizing the Impact of Antibiotics

How antibiotics directly affect pathogens is of keen interest in the quest for more potent and effective disease treatments. In microbes, antibiotics target a limited number of cell processes central to life. One major target is cell-wall biosynthesis. A second target is protein biosynthesis by affecting ribosome translation [19.182]. A third is DNA transcription [19.183] or replication [19.184]. These direct effects have historically been identified and characterized in detail using biochemical and molecular biology. Often, antibiotic resistance arises due to mutations in these direct targets. While some mechanisms of bacterial antibiotic resistance have been described recently, these account for only a small portion of the vast numbers of cell–antibiotic interactions. Identifying the responsible factors that lead to antibiotic resistance is a major challenge. Here, the creative use of pre-existent technologies and the development of new MEMS-based techniques could result in substantial advances in our ability to study antibiotic action.

19.8.2 Detecting Fluctuations in the Buoyant Mass of Single Bacteria

As described above, suspended micro-channel resonators (SMR) can detect subtle fluctuations in physiology at the single-cell level [19.164]. One use of such a device is measuring the buoyant mass of bacteria in the presence of an antibiotic. If bacteria are to flow through a SMR containing antibiotics, one can determine physiological changes in the buoyant mass of a cell. The minimum reported detectable limit of transiting mass is 10^{-15} grams (femtograms), which is sensitive enough to detect changes in cell density at the molecular level. Using this method, *Knudsen* et al. were able to distinguish between the pathogenic bacterium *Citrobacter rodentium* known to exhibit resistance to ampicillin (Amp^R) and its ampicillin-sensitive (Amp^S) mutant strain [19.185]. These observations suggest that changes in the buoyant mass can be used as a proxy for determining the antibiotic sensitivity of clinical-bacterial strains or for the diagnosis of antibiotic resistance.

19.8.3 Detecting the Dynamic Motion of Bacteria in Fluid

Another measure used to assess antibiotic efficacy is the dynamic motion of bacterial cells in a fluid. The detection of bacterial viability makes the use of a cantilever as a biosensor of potential interest to the biomedical fields [19.186]. The temporal fluctuations of a cantilever have been monitored over the course of an experiment, wherein bacteria are adhered to the top surface of the cantilever and are subjected to antibiotic stress [19.187]. The fluctuation amplitude increases only when living cells are adhered to the surface of the cantilever (Fig. 19.13). The increase in cantilever fluctuation is attributed to a *nanomotion* of cells [19.189]. What biophysical attribute of living cells is being measured remains unclear, but is currently a topic of keen interest. Experiments done with *Escherichia coli*, treated with ampicillin, revealed a minimum inhibitory concentration (MIC) of about $1.2\pm0.6\,\mu\text{g/ml}$ and a minimum bactericidal concentration of $7.6\pm0.6\,\mu\text{g/ml}$ [19.187]. These values are within the same order of magnitude as values described by conventional biological techniques that entail counting CFUs of bacteria grown on nutrient agar plates containing varying concentrations of an antibiotic. The benefit of the cantilever-based method is that one could potentially detect bacterial resistance of a clinical strain

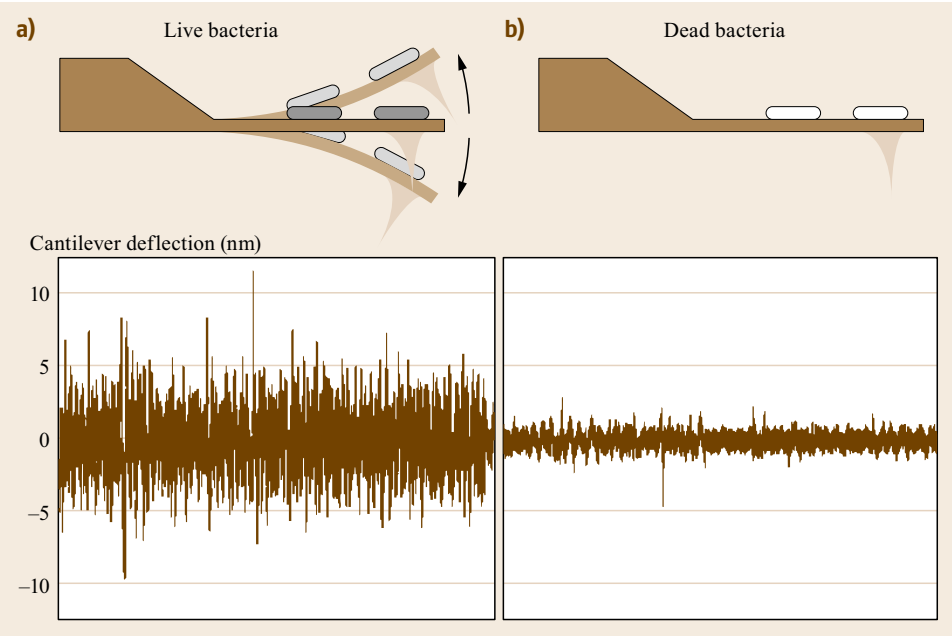


Fig. 19.13a,b Bacterial cell permeability and viability detected by variations in cantilever deflection. Above is depicted a schematic representation of the large variations in cantilever deflection exhibited when living cells **(a)** and dead cells **(b)** are adherent on the surface of the cantilever tip. *Below* is a graphical representation of the range in deflection exhibited by the cantilever in living **(a)** and dead **(b)** cells. A cantilever with adherent live cells has a larger range in cantilever deflection than with dead cells (after [19.188])

more quickly and with a lower titer than needed to perform more conventional techniques currently used in the biomedical field.

19.8.4 Antibiotic Effect on the Bacterial Cell Surface

The biosynthesis of the cell surface is a process directly targeted by multiple first-line antibiotics used for infectious diseases. Antibiotics, like vancomycin, target and inhibit the crosslinking of peptidoglycan, while others block the formation of precursor molecules from which wall components are made. Their impact on bacteria therefore perturbs the physical stability of bacterial cells, making them more fragile, and exhibiting membrane permeability or death. One direct way to assess cell-surface morphology and mechanical properties is to use atomic force microscopy [19.190]. AFM remains to date the only tool to visualize the bacterial-cell surface in time-lapse [19.191]. It is also the only method to directly map changes in mechanical properties along the cell surface, under conditions of growth and stress.

How bacteria are affected morphologically at the cell surface is a central point of interest. AFM can provide unique insights into how cell-surface perturbation leads to cell death. *E. coli* were imaged at high speed and treated with the pore-forming antimicrobial peptide CM15 [19.192]. The cell-surface morphology of *E. coli* appeared to rapidly change from smooth to corrugated over the course of minutes (Fig. 19.14). Curiously, the cells imaged behaved differently in their response to CM15, becoming corrugated on their surfaces at different times in relation to one another. This observation suggests an inherent heterogeneity of single cells in a genetically identical population of bacteria in their ability to tolerate antimicrobial peptide treatment. While the bulging of the *E. coli* cell-surface morphology was visible upon treatment with the peptidoglycan-crosslinking inhibitor, vancomycin, (observable by optical microscopy [19.193]), the emergence of surface corrugations visualized by tapping-mode AFM is undetectable by optical microscopy. Further study into modifications in surface morphology by AFM has been conducted on myocbacteria [19.194], among which *Mycobacterium tuberculosis* is the causative agent of tuberculosis. By following subtle changes in cell morphology, AFM studies have revealed the various ways in which antibiotics and antimicrobial peptides affect cells and offered hypotheses as to how cell death ensues.

It has been suggested that, if the cell-surface morphology is modified by antibiotic treatment, then the

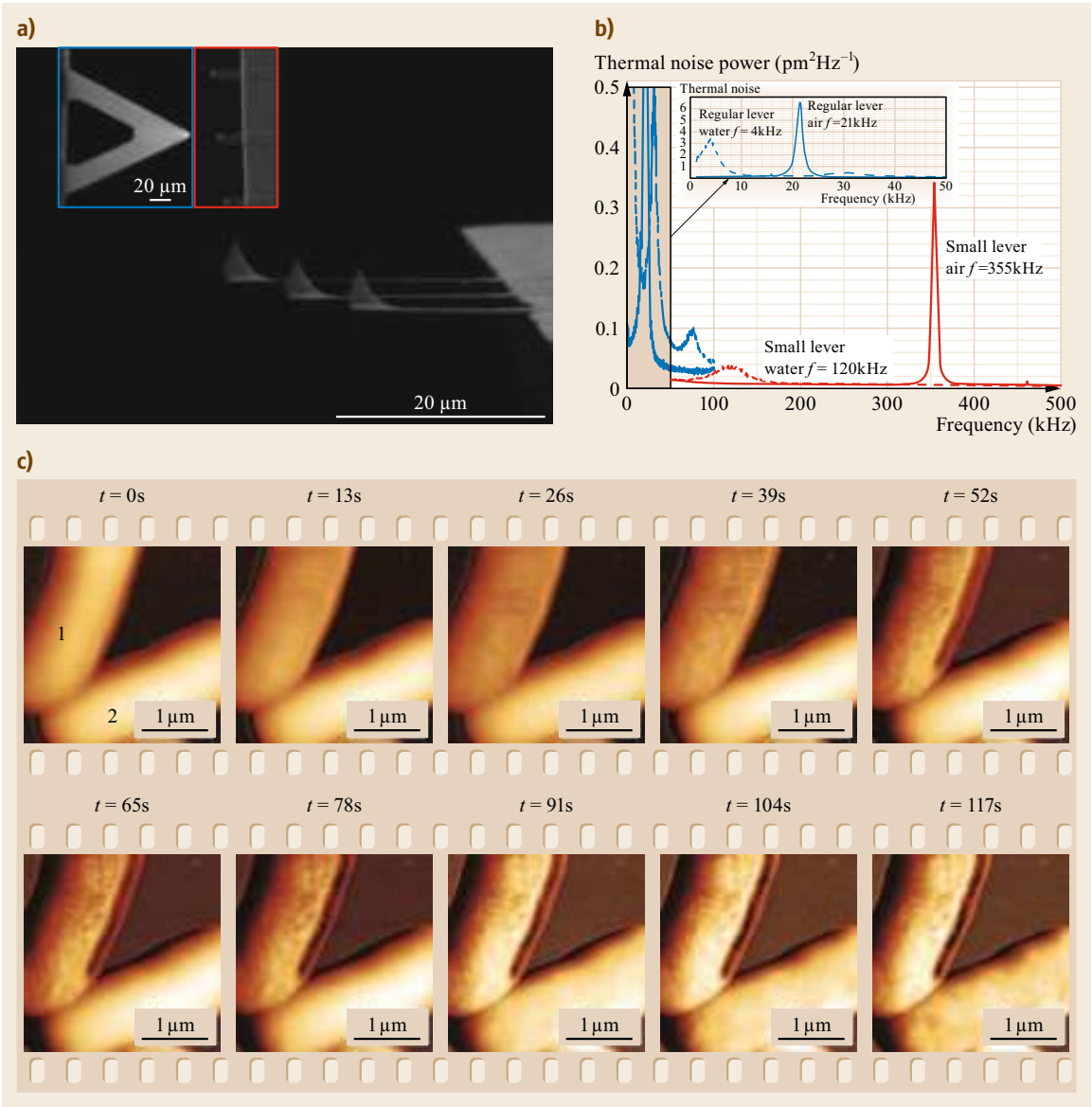


Fig. 19.14a–c Surface-topology fluctuations detected by high-speed AFM using tapping mode. On **(a)** are depicted scanning-electron-micrograph images of small cantilevers. On **(b)** are depicted graphical representations of the resonance frequencies of large (*blue*) and small (*red*) cantilevers. Large cantilevers resonate at ≈ 21 kHz at the first eigenmode whereas small cantilevers resonate at ≈ 355 kHz. **(c)** Graphical film representation of a short time-lapse experiment, wherein cells treated with an antimicrobial peptide exhibit the appearance of cell surface corrugations (after [19.192])

mechanical properties are as well. This could link the surface roughening to the disruption of cell-wall layers composed of arabinogalactan or peptidoglycan crosslinks. These layers are considered to bear the brunt of the architectural load maintaining gross cell morphology [19.195]. Therefore, a second way in which AFM can generally be used to assess the effect of antibiotics on bacterial cells is to quantitatively map surface-mechanical properties [19.186]. By measuring force curves, one can estimate the mechanical properties of the bacteria. Upon treatment of *E. coli* with the β-lactam antibiotic ampicillin regions of greater initial stiffness along the length of the cell undergo the greatest change in mechanical properties following cell death [19.196]. Some studies, however, report a smooth bacterial surface for *E. coli* associated with

homogeneous mechanical properties along that same cell length [19.197, 198].

The ongoing study of how mechanical properties change on the surface of cells treated with antibiotic will reveal the physical principles that are fundamental to our understanding of how cells are formed and how their morphology is maintained. Studies into the repair of the cell surface upon stress will give insights into how cells respond to antibiotics and might lead to ways to prevent the occurrence of antibiotic resistance.

19.9 Conclusions and Outlook

The size scale of MEMS and NEMS make them well-suited for studying biological systems at the single-cell and subcellular level. A myriad of proof-of-concept experiments have demonstrated their potential impact in cell biology. For these techniques to be broadly exploited in the cell biology community, however, some improvements are necessary. To enable wide adoption of these methods requires the development of robust, reproducible, and cost-effective protocols. In addition, to further develop MEMS devices, this will require extensive effort in instrumentation development, sample handling, and ease-of-use improvements.

A second requirement is to correlate the new MEMS-based methods with existing standard cell-biology techniques. Over the most recent decades, a tremendous amount of protocols have been developed and knowledge has been gathered using traditional cell-biology methods. Vast databases have been created, and standardized tests have been established for diagnostic purposes. For new MEMS-based methods to be accepted in the cell-biology community, they must tie in with this existing knowledge base. Firstly, results of the new methods must be validated on reference samples by traditional techniques. Secondly, they need to provide significant advantages in everyday lab use over established technologies. These advantages can be in terms of sensitivity, measurement speed, throughput, or cost. However, great care has to be taken that these advantages don't come at the expense of factors such as standardization, reliability, versatility, or ease-of-use. The successful integration of MEMS technology into cell biology will require an intimate collaboration between engineers and cell biologists.

References

19.1 H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, A. Ammon, K. Martin: *Molecular Cell Biology*, 8th edn. (W. H. Freeman, New York 2016)

19.2 A. Koch: *Bacterial Growth and Form* (Springer, Dordrecht 2001)

19.3 P.D. Odermatt, A. Shivanandan, H. Deschout, R. Jankele, A.P. Nievergelt, L. Feletti, M.W. Davidson, A. Radenovic, G.E. Fantner: High-resolution correlative microscopy: Bridging the gap between single molecule localization microscopy and atomic force microscopy, Nano Lett **15**(8), 4896–4904 (2015)

19.4 C.J.R. Sheppard: Resolution and superresolution, Microsc. Res. Tech. **80**(6), 590–598 (2017)

19.5 E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, H.F. Hess: Imaging intracellular fluorescent proteins at nanometer resolution, Science **313**(5793), 1642–1645 (2006)

19.6 S.T. Hess, T.P. Girirajan, M.D. Mason: Ultra-high resolution imaging by fluorescence photoactivation localization microscopy, Biophys J **91**(11), 4258–4272 (2006)

19.7 M.J. Rust, M. Bates, X. Zhuang: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), Nat. Methods **3**(10), 793–795 (2006)

19.8 S.W. Hell, J. Wichmann: Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy, Opt. Lett. **19**(11), 780–782 (1994)

19.9 B. Harke, J. Chacko, H. Haschke, C. Canale, A. Diaspro: A novel nanoscopic tool by combining AFM with STED microscopy, Opt. Nanoscopy **1**(1), 3 (2012)

19.10 J.V. Chacko, F.C. Zanacchi, A. Diaspro: Probing cytoskeletal structures by coupling optical super-resolution and AFM techniques for a correlative approach, Cytoskeleton (Hoboken) **70**(11), 729–740 (2013)

19.11 S. Suresh, J. Spatz, J.P. Mills, A. Micoulet, M. Dao, C.T. Lim, M. Beil, T. Seufferlein: Connections between single-cell biomechanics and human disease states: Gastrointestinal cancer and malaria, Acta Biomaterialia **1**(1), 15–30 (2005)

19.12 S.E. Cross, Y.-S. Jin, J. Rao, J.K. Gimzewski: Nanomechanical analysis of cells from cancer patients, Nat. Nanotechnol. **2**(12), 780–783 (2007)

19.13 Y.I. Rabinovich, S. Daosukho, K.J. Byer, H.E. El-Shall, S.R. Khan: Direct AFM measurements of adhesion forces between calcium oxalate monohydrate and kidney epithelial cells in the presence

of Ca2+ and Mg2+ ions, J. Colloid Interface Sci. **325**(2), 594–601 (2008)
19.14 B. Cappella, G. Dietler: Force-distance curves by atomic force microscopy, Surf. Sci. Rep. **34**(1), 1–104 (1999)
19.15 J.L. Hutter, J. Bechhoefer: Calibration of atomic-force microscope tips, Rev. Sci. Instrum. **64**(7), 1868–1873 (1993)
19.16 J.E. Sader, J.W.M. Chon, P. Mulvaney: Calibration of rectangular atomic force microscope cantilevers, Rev. Sci. Instrum. **70**(10), 3967–3969 (1999)
19.17 J.P. Cleveland, S. Manne, D. Bocek, P.K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64**(2), 403–405 (1993)
19.18 J. te Riet, A.J. Katan, C. Rankl, S.W. Stahl, A.M. van Buul, I.Y. Phang, A. Gomez-Casado, P. Schön, J.W. Gerritsen, A. Cambi, A.E. Rowan, G.J. Vancso, P. Jonkheijm, J. Huskens, T.H. Oosterkamp, H. Gaub, P. Hinterdorfer, C.G. Figdor, S. Speller: Interlaboratory round robin on cantilever calibration for AFM force spectroscopy, Ultramicroscopy **111**(12), 1659–1669 (2011)
19.19 H.J. Butt, M. Jaschke: Calculation of thermal noise in atomic force microscopy, Nanotechnology **6**(1), 1 (1995)
19.20 J.E. Sader: Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope, J. Appl. Phys. **84**(1), 64–76 (1998)
19.21 European Cooperation in Science and Technology: European network on applications of Atomic Force Microscopy to NanoMedicine and Life Sciences (AFM4NanoMed&Bio), http://www.cost.eu/COST_Actions/bmbs/TD1002 (2011)
19.22 J.L. Alonso, W.H. Goldmann: Feeling the forces: atomic force microscopy in cell biology, Life Sci **72**(23), 2553–2560 (2003)
19.23 H. Hertz: Ueber die Beruehrung fester elastischer Koerper, J. reine angew. Math. **92**, 156–171 (1881)
19.24 A. Touhami, B. Nysten, Y.F. Dufrêne: Nanoscale mapping of the elasticity of microbial cells by atomic force microscopy, Langmuir **19**(11), 4539–4543 (2003)
19.25 I. Sokolov, M.E. Dokukin, N.V. Guz: Method for quantitative measurements of the elastic modulus of biological cells in AFM indentation experiments, Methods **60**(2), 202–213 (2013)
19.26 I. Diaconu, D. Dorohoi: Properties of polyurethane thin films, Optoelectron. Adv. Mater **7**(2), 921–924 (2005)
19.27 M.E. Dokukin, I. Sokolov: On the measurements of rigidity modulus of soft materials in nano-indentation experiments at small depth, Macromolecules **45**(10), 4277–4288 (2012)
19.28 G. Smolyakov, C. Formosa-Dague, C. Severac, R.E. Duval, E. Dague: High speed indentation measures by FV, QI and QNM introduce a new understanding of bionanomechanical experiments, Micron **85**, 8–14 (2016)
19.29 R. García, R. Pérez: Dynamic atomic force microscopy methods, Surf. Sci. Rep. **47**(6–8), 197–301 (2002)
19.30 B.V. Derjaguin, V.M. Muller, Y.P. Toporov: Effect of contact deformations on the adhesion of particles, J. Colloid Interface Sci. **53**(2), 314–326 (1975)
19.31 R. Garcia, R. Perez: Dynamic atomic force microscopy methods, Surface Sci. Rep. **47**(6–8), 197–301 (2002)
19.32 M. Radmacher, J.P. Cleveland, M. Fritz, H.G. Hansma, P.K. Hansma: Mapping interaction forces with the atomic force microscope, Biophys. J. **66**(6), 2159–2165 (1994)
19.33 D. Alsteens, V. Dupres, S. Yunus, J.-P. Latgé, J.J. Heinisch, Y.F. Dufrêne: High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy, Langmuir **28**(49), 16738–16744 (2012)
19.34 Y.F. Dufrene, D. Martinez-Martin, I. Medalsy, D. Alsteens, D.J. Muller: Multiparametric imaging of biological systems by force-distance curve-based AFM, Nat. Methods **10**(9), 847–854 (2013)
19.35 E.K. Dimitriadis, F. Horkay, J. Maresca, B. Kachar, R.S. Chadwick: Determination of elastic moduli of thin layers of soft material using the atomic force microscope, Biophys. J. **82**(5), 2798–2810 (2002)
19.36 N. Gavara, R.S. Chadwick: Determination of the elastic moduli of thin samples and adherent cells using conical atomic force microscope tips, Nat. Nanotechnol. **7**(11), 733–736 (2012)
19.37 World Health Organization: Cancer, http://www.who.int/cancer/en/ (2016)
19.38 J. Folkman: How is blood-vessel growth regulated in normal and neoplastic tissue – GHA clowes memorial award lecture, Cancer Res. **46**(2), 467–473 (1986)
19.39 J. Folkman: The role of angiogenesis in tumor growth, Seminars Cancer Biol **3**(2), 65–71 (1992)
19.40 E.C. Woodhouse, R.F. Chuaqui, L.A. Liotta: General mechanisms of metastasis, Cancer **80**(S8), 1529–1537 (1997)
19.41 I.J. Fidler: Critical determinants of cancer metastasis: Rationale for therapy, Cancer Chemother. Pharmacol. **43**(1), S3–S10 (1999)
19.42 L.A. Liotta: Tumor invasion and metastases – Role of the extracellular matrix: Rhoads Memorial Award lecture, Cancer Res. **46**(1), 1–7 (1986)
19.43 J.B. Wyckoff, J.G. Jones, J.S. Condeelis, J.E. Segall: A critical step in metastasis: In vivo analysis of intravasation at the primary tumor, Cancer Res. **60**(9), 2504–2511 (2000)
19.44 G.L. Nicolson: Cancer metastasis: tumor cell and host organ properties important in metastasis to specific secondary sites, Biochim. Biophys. Acta Rev. Cancer **948**(2), 175–224 (1988)
19.45 G. Poste, I.J. Fidler: The pathogenesis of cancer metastasis, Nature **283**(5743), 139–146 (1980)
19.46 I.J. Fidler: The organ microenvironment and cancer metastasis, Differentiation **70**(9/10), 498–505 (2002)
19.47 I.J. Fidler: The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited, Nat. Rev.

Cancer **3**(6), 453–458 (2003)

19.48 A.F. Chambers, A.C. Groom, I.C. MacDonald: Metastasis: Dissemination and growth of cancer cells in metastatic sites, Nat. Rev. Cancer **2**(8), 563–572 (2002)

19.49 M. Yilmaz, G. Christofori: EMT, the cytoskeleton, and cancer cell invasion, Cancer Metastasis Rev **28**(1), 15–33 (2009)

19.50 D. Wirtz, K. Konstantopoulos, P.C. Searson: The physics of cancer: the role of physical interactions and mechanical forces in metastasis, Nat. Rev. Cancer **11**(7), 512–522 (2011)

19.51 M.O. Krisenko, A. Cartagena, A. Raman, R.L. Geahlen: Nanomechanical property maps of breast cancer cells as determined by multiharmonic atomic force microscopy reveal syk-dependent changes in microtubule stability mediated by MAP1B, Biochemistry **54**(1), 60–68 (2015)

19.52 E.K.F. Yim, E.M. Darling, K. Kulangara, F. Guilak, K.W. Leong: Nanotopography-induced changes in focal adhesions, cytoskeletal organization, and mechanical properties of human mesenchymal stem cells, Biomaterials **31**(6), 1299–1306 (2010)

19.53 S. Suresh: Biomechanics and biophysics of cancer cells, Acta Biomaterialia **3**(4), 413–438 (2007)

19.54 K. Greish: Enhanced permeability and retention (EPR) effect for anticancer nanomedicine drug targeting. In: *Cancer Nanotechnology: Methods and Protocols*, ed. by R.S. Grobmyer, M.B. Moudgil (Humana Press, Totowa 2010) pp. 25–37

19.55 A.M. Handorf, Y. Zhou, M.A. Halanski, W.-J. Li: Tissue stiffness dictates development, homeostasis, and disease progression, Organogenesis **11**(1), 1–15 (2015)

19.56 W.L.K. Chen, C.A. Simmons: Lessons from (patho)physiological tissue stiffness and their implications for drug screening, drug delivery and regenerative medicine, Adv. Drug Deliv. Rev. **63**(4/5), 269–276 (2011)

19.57 K.D. Costa: Single-cell elastography: Probing for disease with the atomic force microscope, Dis. Mark. **19**(2/3), 139–154 (2003)

19.58 E.H. Zhou, S.T. Quek, C.T. Lim: Power-law rheology analysis of cells undergoing micropipette aspiration, Biomech. Model. Mechanobiol. **9**(5), 563–572 (2010)

19.59 Y.-S. Chu, W.A. Thomas, O. Eder, F. Pincet, E. Perez, J.P. Thiery, S. Dufour: Force measurements in E-cadherin-mediated cell doublets reveal rapid adhesion strengthened by actin cytoskeleton remodeling through Rac and Cdc42, J. Cell Biol. **167**(6), 1183–1194 (2004)

19.60 M. Daoudi, E. Lavergne, A. Garin, N. Tarantino, P. Debré, F. Pincet, C. Combadière, P. Deterre: Enhanced adhesive capacities of the naturally occurring Ile249–Met280 variant of the chemokine receptor CX3CR1, J. Biol. Chem. **279**(19), 19649–19657 (2004)

19.61 K. Sung, L.A. Sung, M. Crimmins, S.J. Burakoff, S. Chien: Determination of junction avidity of cytolytic T cell and target cell, Science **234**(4782), 1405–1408 (1986)

19.62 R. Fodil, V. Laurent, E. Planus, D. Isabey: Characterization of cytoskeleton mechanical properties and 3D-actin structure in twisted adherent epithelial cells, Biorheology **40**(1–3), 241–245 (2003)

19.63 D.R. Gossett, H.T. Tse, S.A. Lee, Y. Ying, A.G. Lindgren, O.O. Yang, J. Rao, A.T. Clark, D. Di Carlo: Hydrodynamic stretching of single cells for large population mechanical phenotyping, Proc. Natl. Acad. Sci. USA **109**(20), 7630–7635 (2012)

19.64 Y. Li, C. Wen, H. Xie, A. Ye, Y. Yin: Mechanical property analysis of stored red blood cell using optical tweezers, Colloids Surf. B **70**(2), 169–173 (2009)

19.65 D. Kirmizis, S. Logothetidis: Atomic force microscopy probing in the measurement of cell mechanics, Int. J. Nanomed. **5**(137), e45 (2010)

19.66 K.C. Neuman, A. Nagy: Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, Nat. Methods **5**(6), 491–505 (2008)

19.67 M. Lekka, K. Pogoda, J. Gostek, O. Klymenko, S. Prauzner-Bechcicki, J. Wiltowska-Zuber, J. Jaczewska, J. Lekki, Z. Stachura: Cancer cell recognition – Mechanical phenotype, Micron **43**(12), 1259–1266 (2012)

19.68 J. Guck, S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H.M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Käs, S. Ulvick, C. Bilby: Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence, Biophys. J. **88**(5), 3689–3698 (2005)

19.69 Y. Wu, G.D. McEwen, S. Harihar, S.M. Baker, D.B. DeWald, A. Zhou: BRMS1 expression alters the ultrastructural, biomechanical and biochemical properties of MDA-MB-435 human breast carcinoma cells: An AFM and Raman microspectroscopy study, Cancer Lett **293**(1), 82–91 (2010)

19.70 Y.M. Sung, X. Xu, J. Sun, D. Mueller, K. Sentissi, P. Johnson, E. Urbach, F. Seillier-Moiseiwitsch, M.D. Johnson, S.C. Mueller: Tumor suppressor function of syk in human MCF10A in vitro and normal mouse mammary epithelium in vivo, PLoS ONE **4**(10), e7445 (2009)

19.71 H. Jin, J. Pi, X. Huang, F. Huang, W. Shao, S. Li, Y. Chen, J. Cai: BMP2 promotes migration and invasion of breast cancer cells via cytoskeletal reorganization and adhesion decrease: An AFM investigation, Appl. Microbiol. Biotechnol. **93**(4), 1715–1723 (2012)

19.72 X. Zhang, U. Shrikhande, B.M. Alicie, Q. Zhou, R.L. Geahlen: Role of the protein tyrosine kinase Syk in regulating cell-cell adhesion and motility in breast cancer cells, Mol. Cancer Res. **7**(5), 634–644 (2009)

19.73 W. Xu, R. Mezencev, B. Kim, L. Wang, J. McDonald, T. Sulchek: Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells, PLoS ONE **7**(10), e46609 (2012)

19.74 A. Moustakas, C. Stournaras: Regulation of actin organisation by TGF-beta in H-ras-transformed fibroblasts, J. Cell Sci. **112**(8), 1169–1179 (1999)

19.75 M.E. Grady, R.J. Composto, D.M. Eckmann: Cell elasticity with altered cytoskeletal architectures across multiple cell types, J. Mech. Behav. Biomed. Mater. **61**, 197–207 (2016)

19.76 T. Wakatsuki, B. Schwab, N.C. Thompson, E.L. Elson: Effects of cytochalasin D and latrunculin B on mechanical properties of cells, J. Cell Sci. **114**(5), 1025–1036 (2001)

19.77 Y.M. Efremov, M.E. Lomakina, D.V. Bagrov, P.I. Makhnovskiy, A.Y. Alexandrova, M.P. Kirpichnikov, K.V. Shaitan: Mechanical properties of fibroblasts depend on level of cancer transformation, Biochimica et Biophysica Acta (BBA) – Mol. Cell Res. **1843**(5), 1013–1019 (2014)

19.78 A. Calzado-Martín, M. Encinar, J. Tamayo, M. Calleja, A. San Paulo: Effect of actin organization on the stiffness of living breast cancer cells revealed by peak-force modulation atomic force microscopy, ACS Nano **10**(3), 3365–3374 (2016)

19.79 M. Plodinec, M. Loparic, C.A. Monnier, E.C. Obermann, R. Zanetti-Dallenbach, P. Oertle, J.T. Hyotyla, U. Aebi, M. Bentires-Alj, Y.H. LimRoderick, C.-A. Schoenenberger: The nanomechanical signature of breast cancer, Nat. Nanotechnol. **7**(11), 757–765 (2012)

19.80 J.R. Ramos, J. Pabijan, R. Garcia, M. Lekka: The softening of human bladder cancer cells happens at an early stage of the malignancy process, Beilstein J. Nanotechnol. **5**, 447–457 (2014)

19.81 G. Weder, M.C. Hendriks-Balk, R. Smajda, D. Rimoldi, M. Liley, H. Heinzelmann, A. Meister, A. Mariotti: Increased plasticity of the stiffness of melanoma cells correlates with their acquisition of metastatic properties, Nanomed. Nanotechnol. Biol. Med. **10**(1), 141–148 (2014)

19.82 A. Goddi, M. Bonardi, S. Alessi: Breast elastography: A literature review, J. Ultrasound **15**(3), 192–198 (2012)

19.83 S. Abbas, Z. Judit, P. Donald: Elastic moduli of normal and pathological human breast tissues: an inversion-technique-based investigation of 169 samples, Phys. Med. Biol. **52**(6), 1565 (2007)

19.84 U. Dammer, O. Popescu, P. Wagner, D. Anselmetti, H.-J. Guntherodt, G.N. Misevic: Binding strength between cell adhesion proteoglycans measured by atomic force microscopy, Science **267**(5201), 1173–1175 (1995)

19.85 M. Benoit, D. Gabriel, G. Gerisch, H.E. Gaub: Discrete interactions in cell adhesion measured by single-molecule force spectroscopy, Nat. Cell Biol. **2**(6), 313–317 (2000)

19.86 J. Helenius, C.-P. Heisenberg, H.E. Gaub, D.J. Muller: Single-cell force spectroscopy, J. Cell Sci. **121**(11), 1785–1791 (2008)

19.87 W. Grange, T. Strunz, I. Schumakovitch, H.-J. Güntherodt, M. Hegner: Molecular recognition and adhesion of individual DNA strands studied by dynamic force microscopy, Single Mol **2**(2), 75–78 (2001)

19.88 Y. Chen, G. Zeng, S.S. Chen, Q. Feng, Z.W. Chen: AFM force measurements of the gp120–sCD4 and gp120 or CD4 antigen–antibody interactions, Biochem. Biophys. Res. Commun. **407**(2), 301–306 (2011)

19.89 P.P. Lehenkari, M.A. Horton: Single Integrin Molecule Adhesion Forces in Intact Cells Measured by Atomic Force Microscopy, Biochem. Biophys. Res. Commun. **259**(3), 645–650 (1999)

19.90 M. Krieg, Y. Arboleda-Estudillo, P.-H. Puech, J. Käfer, F. Graner, D. Müller, C.-P. Heisenberg: Tensile forces govern germ-layer organization in zebrafish, Nat. Cell Biol. **10**(4), 429–436 (2008)

19.91 H. Sheng-Wen, T. Doan Van Hong, H. Ming-Hua, H. Hsyue-Jen, L. Chung-Hsing, H. Chang-Hsiang, L. Hsi-Hsin: Interactions between chitosan and cells measured by AFM, Biomed. Mat. **5**(5), 054117 (2010)

19.92 G. Weder, N. Blondiaux, M. Giazzon, N. Matthey, M. Klein, R. Pugin, H. Heinzelmann, M. Liley: Use of force spectroscopy to investigate the adhesion of living adherent cells, Langmuir **26**(11), 8180–8186 (2010)

19.93 Y.F. Dufrene: Sticky microbes: Forces in microbial cell adhesion, Trends Microbiol. **23**(6), 376–382 (2015)

19.94 M. Thie, R. Röspel, W. Dettmann, M. Benoit, M. Ludwig, H.E. Gaub, H.W. Denker: Interactions between trophoblast and uterine epithelium: Monitoring of adhesive forces, Hum. Reprod. **13**(11), 3211–3219 (1998)

19.95 V.T. Moy, E.-L. Florin, H.E. Gaub: Adhesive forces between ligand and receptor measured by AFM, Colloids Surfaces A: Physicochem. Eng. Aspects **93**, 343–348 (1994)

19.96 T.V. Ratto, K.C. Langry, R.E. Rudd, R.L. Balhorn, M.J. Allen, M.W. McElfresh: Force spectroscopy of the double-tethered concanavalin-a mannose bond, Biophys. J. **86**(4), 2430–2437 (2004)

19.97 P.-H. Puech, A. Taubenberger, F. Ulrich, M. Krieg, D.J. Muller, C.-P. Heisenberg: Measuring cell adhesion forces of primary gastrulating cells from zebrafish using atomic force microscopy, J. Cell Sci. **118**(18), 4199–4206 (2005)

19.98 G. Weder, J. Vörös, M. Giazzon, N. Matthey, H. Heinzelmann, M. Liley: Measuring cell adhesion forces during the cell cycle by force spectroscopy, Biointerphases **4**(2), 27–34 (2009)

19.99 J. Friedrichs, J.M. Torkko, J. Helenius, T.P. Teräväinen, J. Füllekrug, D.J. Muller, K. Simons, A. Manninen: Contributions of galectin-3 and -9 to epithelial cell adhesion analyzed by single cell force spectroscopy, J. Biol. Chem. **282**(40), 29375–29383 (2007)

19.100 A. Taubenberger, D.A. Cisneros, J. Friedrichs, P.-H. Puech, D.J. Muller, C.M. Franz: Revealing early steps of $\alpha 2\beta 1$ integrin-mediated adhesion to collagen type I by using single-cell force spectroscopy, Mol. Biol. Cell **18**(5), 1634–1644 (2007)

19.101 Y.F. Dufrêne: Towards nanomicrobiology using atomic force microscopy, Nat. Rev. Microbiol. **6**(9), 674–680 (2008)

19.102 S.K. Lower, M.F. Hochella, T.J. Beveridge: Bacterial recognition of mineral surfaces: Nanoscale interactions between Shewanella and α-FeOOH,

Science **292**(5520), 1360–1363 (2001)
19.103 N.P. Boks, H.J. Busscher, H.C. van der Mei, W. Norde: Bond-strengthening in staphylococcal adhesion to hydrophilic and hydrophobic surfaces using atomic force microscopy, Langmuir **24**(22), 12990–12994 (2008)
19.104 A. Razatos, Y.-L. Ong, M.M. Sharma, G. Georgiou: Molecular determinants of bacterial adhesion monitored by atomic force microscopy, Proc. Natl. Acad. Sci. **95**(19), 11059–11064 (1998)
19.105 V. Vadillo-Rodríguez, H.J. Busscher, W. Norde, J. De Vries, R.J. Dijkstra, I. Stokroos, H.C. Van Der Mei: Comparison of atomic force microscopy interaction forces between bacteria and silicon nitride substrata for three commonly used immobilization methods, Appl. Environ. Microbiol. **70**(9), 5441–5446 (2004)
19.106 S. Kang, M. Elimelech: Bioinspired single bacterial cell force spectroscopy, Langmuir **25**(17), 9656–9659 (2009)
19.107 A. Harimawan, A. Rajasekar, Y.-P. Ting: Bacteria attachment to surfaces – AFM force spectroscopy and physicochemical analyses, J. Colloid Interface Sci. **364**(1), 213–218 (2011)
19.108 D. Alsteens, A. Beaussart, S. Derclaye, S. El-Kirat-Chatel, H.R. Park, P.N. Lipke, Y.F. Dufrene: Single-cell force spectroscopy of Als-mediated fungal adhesion, Anal. Methods **5**(15), 3657–3662 (2013)
19.109 A. Beaussart, S. El-Kirat-Chatel, P. Herman, D. Alsteens, J. Mahillon, P. Hols, Y.F. Dufrêne: Single-cell force spectroscopy of probiotic bacteria, Biophys. J. **104**(9), 1886–1892 (2013)
19.110 T. Cao, H.Y. Tang, X.M. Liang, A.F. Wang, G.W. Auner, S.O. Salley, K.Y.S. Ng: Nanoscale investigation on adhesion of E-coli surface modified silicone using atomic force microscopy, Biotechnol. Bioeng. **94**(1), 167–176 (2006)
19.111 D. Bieber, S.W. Ramer, C.-Y. Wu, W.J. Murray, T. Tobe, R. Fernandez, G.K. Schoolnik: Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic Escherichia coli, Science **280**(5372), 2114–2118 (1998)
19.112 J.S. Mattick: Type IV pili and twitching motility, Annu. Rev. Microbiol. **56**(1), 289–314 (2002)
19.113 L. Craig, M.E. Pique, J.A. Tainer: Type IV pilus structure and bacterial pathogenicity, Nat. Rev. Microbil. **2**(5), 363–378 (2004)
19.114 L.L. Burrows: Pseudomonas aeruginosa twitching motility: Type IV pili in action, Annu. Rev. Microbiol. **66**(1), 493–520 (2012)
19.115 S. Lu, M. Giuliani, H. Harvey, L.L. Burrows, R.A. Wickham, J.R. Dutcher: Nanoscale pulling of type IV pili reveals their flexibility and adhesion to surfaces over extended lengths of the pili, Biophys. J. **108**(12), 2865–2875 (2015)
19.116 A. Beaussart, A.E. Baker, S.L. Kuchma, S. El-Kirat-Chatel, G.A. O'Toole, Y.F. Dufrêne: Nanoscale adhesion forces of pseudomonas aeruginosa type IV pili, ACS Nano **8**(10), 10723–10733 (2014)
19.117 L.A. Pratt, R. Kolter: Genetic analysis of Escherichia coli biofilm formation: Roles of flagella, motility, chemotaxis and type I pili, Mol. Microbiol. **30**(2), 285–293 (1998)
19.118 H. Xu, A.E. Murdaugh, W. Chen, K.E. Aidala, M.A. Ferguson, E.M. Spain, M.E. Núñez: Characterizing pilus-mediated adhesion of biofilm-forming E. coli to chemically diverse surfaces using atomic force microscopy, Langmuir **29**(9), 3000–3011 (2013)
19.119 K. Anselme: Osteoblast adhesion on biomaterials, Biomaterials **21**(7), 667–681 (2000)
19.120 F. Baino, S. Fiorilli, C. Vitale-Brovarone: Bioactive glass-based materials with hierarchical porosity for medical applications: Review of recent advances, Acta Biomaterialia **52**, 18–32 (2016)
19.121 A.J. García, P. Ducheyne, D. Boettiger: Effect of surface reaction stage on fibronectin-mediated adhesion of osteoblast-like cells to bioactive glass, J. Biomed. Mat. Res. **40**(1), 48–56 (1998)
19.122 S.M. Wiederhorn, Y.-H. Chae, C.G. Simon Jr., J. Cahn, Y. Deng, D. Day: Cell adhesion to borate glasses by colloidal probe microscopy, Acta Biomaterialia **7**(5), 2256–2263 (2011)
19.123 X. Han, D.E. Day: Reaction of sodium calcium borate glasses to form hydroxyapatite, J. Mat. Sci. Mat. Med. **18**(9), 1837–1847 (2007)
19.124 D. Day, J. White, R. Brown, K. McMenamin: Transformation of borate glasses into biologically useful materials, Glass Technol **44**(2), 75–81 (2003)
19.125 S.D. Conzone, J.G. Hemrick, D.E. Day: Glass formation and chemical durability of dysprosium lithium borate glasses intended for in vivo radiation synovectomy, Glass Sci. Technol. **74**(2), 39–45 (2001)
19.126 W. Huang, D.E. Day, K. Kittiratanapiboon, M.N. Rahaman: Kinetics and mechanisms of the conversion of silicate (45S5), borate, and borosilicate glasses to hydroxyapatite in dilute phosphate solutions, J. Mat. Sci. Mat. Med. **17**(7), 583–596 (2006)
19.127 W. Huang, M.N. Rahaman, D.E. Day, Y. Li: Mechanisms for converting bioactive silicate, borate, and borosilicate glasses to hydroxyapatite in dilute phosphate solution, Phys. Chem. Glasses-Eur. J. Glass Sci. Technol. B **47**(6), 647–658 (2006)
19.128 H.K. No, N. Young Park, S.L. Ho, S.P. Meyers: Antibacterial activity of chitosans and chitosan oligomers with different molecular weights, Int. J. Food Microbiol. **74**(1/2), 65–72 (2002)
19.129 L. Zhao, E.F. Burguera, H.H.K. Xu, N. Amin, H. Ryou, D.D. Arola: Fatigue and human umbilical cord stem cell seeding characteristics of calcium phosphate–chitosan–biodegradable fiber scaffolds, Biomaterials **31**(5), 840–847 (2010)
19.130 R.O. Hynes: Integrins: Versatility, modulation, and signaling in cell adhesion, Cell **69**(1), 11–25 (1992)
19.131 J. Jokinen, E. Dadu, P. Nykvist, J. Käpylä, D.J. White, J. Ivaska, P. Vehviläinen, H. Reunanen, H. Larjava, L. Häkkinen, J. Heino: Integrin-mediated cell adhesion to type I collagen fibrils, J. Biol. Chem. **279**(30), 31956–31963 (2004)
19.132 C.M. Franz, A. Taubenberger, P.-H. Puech, D.J. Muller: Studying integrin-mediated cell adhesion at the single-molecule level using AFM

force spectroscopy, Sci. Signal. **2007**(406), pl5 (2007)
19.133 E.P. Wojcikiewicz, X. Zhang, A. Chen, V.T. Moy: Contributions of molecular binding events and cellular compliance to the modulation of leukocyte adhesion, J. Cell Sci. **116**(12), 2531–2539 (2003)
19.134 E. Ponte, E. Bracco, J. Faix, S. Bozzaro: Detection of subtle phenotypes: The case of the cell adhesion molecule csA in Dictyostelium, Proc. Natl. Acad. Sci. **95**(16), 9360–9365 (1998)
19.135 D.A. Steeber, T.F. Tedder: Adhesion molecule cascades direct lymphocyte recirculation and leukocyte migration during inflammation, Immunol. Res. **22**(2), 299–317 (2000)
19.136 T.A. Springer: Adhesion receptors of the immune system, Nature **346**(6283), 425–434 (1990)
19.137 P. Kubes, S.M. Kerfoot: Leukocyte Recruitment in the Microcirculation: the Rolling Paradigm Revisited, Physiology **16**(2), 76–80 (2001)
19.138 W. Risau: Differentiation of endothelium, FASEB J **9**(10), 926–933 (1995)
19.139 X. Zhang, E.P. Wojcikiewicz, V.T. Moy: Dynamic adhesion of T lymphocytes to endothelial cells revealed by atomic force microscopy, Exp. Biol. Med. **231**(8), 1306–1312 (2006)
19.140 D.C. Radisky: Epithelial-mesenchymal transition, J. Cell Sci. **118**(19), 4325–4326 (2005)
19.141 E. Tomaskovic-Crook, E.W. Thompson, J.P. Thiery: Epithelial to mesenchymal transition and breast cancer, Breast Cancer Res. **11**(6), 1–10 (2009)
19.142 M.A. Huber, N. Kraut, H. Beug: Molecular requirements for epithelial–mesenchymal transition during tumor progression, Curr. Opin. Cell Biol. **17**(5), 548–558 (2005)
19.143 H. Oka, H. Shiozaki, K. Kobayashi, M. Inoue, H. Tahara, T. Kobayashi, Y. Takatsuka, N. Matsuyoshi, S. Hirano, M. Takeichi, T. Mori: Expression of E-Cadherin Cell Adhesion Molecules in Human Breast Cancer Tissues and Its Relationship to Metastasis, Cancer Res. **53**(7), 1696–1701 (1993)
19.144 T. Okegawa, R.-C. Pong, Y. Li, J.-T. Hsieh: The role of cell adhesion molecule in cancer progression and its application in cancer therapy, Acta Biochim. Polonica – Engl. Edition **51**, 445–458 (2004)
19.145 C. Ounkomol, S. Yamada, V. Heinrich: Single-cell adhesion tests against functionalized microspheres arrayed on AFM cantilevers confirm heterophilic E- and N-cadherin binding, Biophys. J. **99**(12), L100–L102 (2010)
19.146 P. Panorchan, M.S. Thompson, K.J. Davis, Y. Tseng, K. Konstantopoulos, D. Wirtz: Single-molecule analysis of cadherin-mediated cell-cell adhesion, J. Cell Sci. **119**(1), 66–74 (2006)
19.147 R. Omidvar, M. Tafazzoli-shadpour, M.A. Shokrgozar, M. Rostami: Atomic force microscope-based single cell force spectroscopy of breast cancer cell lines: An approach for evaluating cellular invasion, J. Biomech. **47**(13), 3373–3379 (2014)
19.148 G. Weder, O. Guillaume-Gentil, N. Matthey, F. Montagne, H. Heinzelmann, J. Vörös, M. Liley: The quantification of single cell adhesion on functionalized surfaces for cell sheet engineering, Biomaterials **31**(25), 6436–6443 (2010)
19.149 E. Potthoff, O. Guillaume-Gentil, D. Ossola, J. Polesel-Maris, S. LeibundGut-Landmann, T. Zambelli, J.A. Vorholt: Rapid and serial quantification of adhesion forces of yeast and Mammalian cells, PLoS One **7**(12), e52712 (2012)
19.150 E. Potthoff, D. Franco, V. D'Alessandro, C. Starck, V. Falk, T. Zambelli, J.A. Vorholt, D. Poulikakos, A. Ferrari: Toward a rational design of surface textures promoting endothelialization, Nano Lett **14**(2), 1069–1079 (2014)
19.151 P. Dörig, P. Stiefel, P. Behr, E. Sarajlic, D. Bijl, M. Gabi, J. Vörös, J.A. Vorholt, T. Zambelli: Force-controlled spatial manipulation of viable mammalian cells and micro-organisms by means of FluidFM technology, Appl. Phys. Lett. **97**(2), 023701 (2010)
19.152 O. Guillaume-Gentil, T. Zambelli, J.A. Vorholt: Isolation of single mammalian cells from adherent cultures by fluidic force microscopy, Lab Chip **14**(2), 402–414 (2014)
19.153 P. Stiefel, T. Zambelli, J.A. Vorholt: Isolation of optically targeted single bacteria by application of fluidic force microscopy to aerobic anoxygenic phototrophs from the phyllosphere, Appl. Environ. Microbiol. **79**(16), 4895–4905 (2013)
19.154 O. Guillaume-Gentil, R.V. Grindberg, R. Kooger, L. Dorwling-Carter, V. Martinez, D. Ossola, M. Pilhofer, T. Zambelli, J.A. Vorholt: Tunable single-cell extraction for molecular analyses, Cell **166**(2), 506–516 (2016)
19.155 O. Guillaume-Gentil, E. Potthoff, D. Ossola, P. Dörig, T. Zambelli, J.A. Vorholt: Force-controlled fluidic injection into single cell nuclei, Small **9**(11), 1904–1907 (2013)
19.156 A. Meister, M. Gabi, P. Behr, P. Studer, J. Vörös, P. Niedermann, J. Bitterli, J. Polesel-Maris, M. Liley, H. Heinzelmann, T. Zambelli: FluidFM: Combining atomic force microscopy and nanofluidics in a universal liquid delivery system for single cell applications and beyond, Nano Lett. **9**(6), 2501–2507 (2009)
19.157 P. Stiefel, F.I. Schmidt, P. Dorig, P. Behr, T. Zambelli, J.A. Vorholt, J. Mercer: Cooperative vaccinia infection demonstrated at the single-cell level using FluidFM, Nano Lett **12**(8), 4219–4227 (2012)
19.158 S. Di Talia, J.M. Skotheim, J.M. Bean, E.D. Siggia, F.R. Cross: The effects of molecular noise and size control on variability in the budding yeast cell cycle, Nature **448**(7156), 947–951 (2007)
19.159 E. Gil-Santos, D. Ramos, J. Martinez, M. Fernandez-Regulez, R. Garcia, A. San Paulo, M. Calleja, J. Tamayo: Nanomechanical mass sensing and stiffness spectrometry based on two-dimensional vibrations of resonant nanowires, Nat. Nanotechnol. **5**(9), 641–645 (2010)
19.160 A. Boisen, S. Dohn, S.S. Keller, S. Schmid, M. Tenje: Cantilever-like micromechanical sensors, Rep. Prog. Phys. **74**(3), 036101 (2011)
19.161 B. Ilic, D. Czaplewski, M. Zalalutdinov, H.G. Craighead, P. Neuzil, C. Campagnolo, C. Batt: Single

cell detection with micromechanical oscillators, J. Vacuum Sci. Technol. B **19**(6), 2825–2828 (2001)
19.162 D. Ramos, J. Tamayo, J. Mertens, M. Calleja: Photothermal excitation of microcantilevers in liquids, J. Appl. Phys. **99**(12), 124904 (2006)
19.163 K.Y. Gfeller, N. Nugaeva, M. Hegner: Rapid biosensor for detection of antibiotic-selective growth of Escherichia coli, Appl. Environ. Microbiol. **71**(5), 2626–2631 (2005)
19.164 T.P. Burg, M. Godin, S.M. Knudsen, W. Shen, G. Carlson, J.S. Foster, K. Babcock, S.R. Manalis: Weighing of biomolecules, single cells and single nanoparticles in fluid, Nature **446**(7139), 1066–1069 (2007)
19.165 J.A. Efe, R.J. Botelho, S.D. Emr: The Fab1 phosphatidylinositol kinase pathway in the regulation of vacuole morphology, Curr. Opin. Cell Biol. **17**(4), 402–408 (2005)
19.166 N. Maloney, G. Lukacs, J. Jensen, M. Hegner: Nanomechanical sensors for single microbial cell growth monitoring, Nanoscale **6**(14), 8242–8249 (2014)
19.167 N. Nugaeva, K.Y. Gfeller, N. Backmann, H.P. Lang, M. Düggelin, M. Hegner: Micromechanical cantilever array sensors for selective fungal immobilization and fast growth detection, Biosens. Bioelectron. **21**(6), 849–856 (2005)
19.168 K. Park, J. Jang, D. Irimia, J. Sturgis, J. Lee, J.P. Robinson, M. Toner, R. Bashir: 'Living cantilever arrays' for characterization of mass of single live cells in fluids, Lab Chip **8**(7), 1034–1041 (2008)
19.169 K. Park, L.J. Millet, N. Kim, H. Li, X. Jin, G. Popescu, N.R. Aluru, K.J. Hsia, R. Bashir: Measurement of adherent cell mass and growth, Proc. Natl. Acad. Sci. USA **107**(48), 20691–20696 (2010)
19.170 A.K. Bryan, V.C. Hecht, W. Shen, K. Payer, W.H. Grover, S.R. Manalis: Measuring single cell mass, volume, and density with dual suspended microchannel resonators, Lab Chip **14**(3), 569–576 (2014)
19.171 W.H. Grover, A.K. Bryan, M. Diez-Silva, S. Suresh, J.M. Higgins, S.R. Manalis: Measuring single-cell density, Proc. Natl. Acad. Sci. USA **108**(27), 10992–10996 (2011)
19.172 M. Godin, F.F. Delgado, S. Son, W.H. Grover, A.K. Bryan, A. Tzur, P. Jorgensen, K. Payer, A.D. Grossman, M.W. Kirschner, S.R. Manalis: Using buoyant mass to measure the growth of single cells, Nat. Methods **7**(5), 387–390 (2010)
19.173 A.K. Bryan, A. Goranov, A. Amon, S.R. Manalis: Measurement of mass, density, and volume during the cell cycle of yeast, Proc. Natl. Acad. Sci. USA **107**(3), 999–1004 (2010)
19.174 M. Carrera, R.O. Zandomeni, J.L. Sagripanti: Wet and dry density of Bacillus anthracis and other Bacillus species, J. Appl. Microbiol. **105**(1), 68–77 (2008)
19.175 K. Inoue, M. Nishimura, B.B. Nayak, K. Kogure: Separation of marine bacteria according to buoyant density by use of the density-dependent cell sorting method, Appl. Environ. Microbiol. **73**(4), 1049–1053 (2007)
19.176 D.C. Jinks, R. Guthrie, E.W. Naylor: Simplified procedure for producing Bacillus subtilis spores for the Guthrie phenylketonuria and other microbiological screening tests, J. Clin. Microbiol. **21**(5), 826–829 (1985)
19.177 H.I. Adler, W.D. Fisher, A. Cohen, A.A. Hardigree: Miniature escherichia coli cells deficient in dna, Proc. Natl. Acad. Sci. USA **57**(2), 321–326 (1967)
19.178 J. Drews: Drug discovery: A historical perspective, Science **287**(5460), 1960–1964 (2000)
19.179 J.P. Overington, B. Al-Lazikani, A.L. Hopkins: How many drug targets are there?, Nat. Rev. Drug. Discov. **5**(12), 993–996 (2006)
19.180 D.C. Swinney, J. Anthony: How were new medicines discovered?, Nat. Rev. Drug. Discov. **10**(7), 507–519 (2011)
19.181 E. Gregori-Puigjane, V. Setola, J. Hert, B.A. Crews, J.J. Irwin, E. Lounkine, L. Marnett, B.L. Roth, B.K. Shoichet: Identifying mechanism-of-action targets for drugs and probes, Proc. Natl. Acad. Sci. USA **109**(28), 11178–11183 (2012)
19.182 T. Lambert: Antibiotics that affect the ribosome, Rev. Sci. Tech. **31**(1), 57–64 (2012)
19.183 C. Ma, X. Yang, P.J. Lewis: Bacterial transcription as a target for antibacterial drug development, Microbiol. Mol. Biol. Rev. **80**(1), 139–160 (2016)
19.184 D.L. Johnson, C.B. Stone, D.C. Bulir, B.K. Coombes, J.B. Mahony: A novel inhibitor of Chlamydophila pneumoniae protein kinase D (PknD) inhibits phosphorylation of CdsD and suppresses bacterial replication, BMC Microbiol. **9**, 218 (2009)
19.185 S.M. Knudsen, M.G. von Muhlen, D.B. Schauer, S.R. Manalis: Determination of bacterial antibiotic resistance based on osmotic shock response, Anal. Chem. **81**(16), 7087–7090 (2009)
19.186 G. Longo, S. Kasas: Effects of antibacterial agents and drugs monitored by atomic force microscopy, Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. **6**(3), 230–244 (2014)
19.187 G. Longo, L. Alonso-Sarduy, L.M. Rio, A. Bizzini, A. Trampuz, J. Notz, G. Dietler, S. Kasas: Rapid detection of bacterial resistance to antibiotics using AFM cantilevers as nanomechanical sensors, Nat. Nanotechnol. **8**(7), 522–526 (2013)
19.188 S. Kasas, F.S. Ruggeri, C. Benadiba, C. Maillard, P. Stupar, H. Tournu, G. Dietler, G. Longo: Detecting nanoscale vibrations as signature of life, Proc. Natl. Acad. Sci. **112**(2), 378–381 (2015)
19.189 S. Kasas, F.S. Ruggeri, C. Benadiba, C. Maillard, P. Stupar, H. Tournu, G. Dietler, G. Longo: Detecting nanoscale vibrations as signature of life, Proc. Natl. Acad. Sci. USA **112**(2), 378–381 (2015)
19.190 Y.F. Dufrêne: Atomic force microscopy in microbiology: New structural and functional insights into the microbial cell surface, mBio **5**(4), e01363-14 (2014)
19.191 Y.F. Dufrene: Towards nanomicrobiology using atomic force microscopy, Nat. Rev. Microbiol. **6**(9), 674–680 (2008)
19.192 G.E. Fantner, R.J. Barbero, D.S. Gray, A.M. Belcher: Kinetics of antimicrobial peptide activity mea-

sured on individual bacterial cells using high-speed atomic force microscopy, Nat. Nanotechnol. **5**(4), 280–285 (2010)

19.193 K.C. Huang, R. Mukhopadhyay, B. Wen, Z. Gitai, N.S. Wingreen: Cell shape and cell-wall organization in Gram-negative bacteria, Proc. Natl. Acad. Sci. USA **105**(49), 19282–19287 (2008)

19.194 D. Alsteens, C. Verbelen, E. Dague, D. Raze, A.R. Baulard, Y.F. Dufrene: Organization of the mycobacterial cell wall: A nanoscale view, Pflugers Arch **456**(1), 117–125 (2008)

19.195 E.C. Hett, E.J. Rubin: Bacterial growth and cell division: a mycobacterial perspective, Microbiol. Mol. Biol. Rev. **72**(1), 126–156 (2008), table of contents

19.196 G. Longo, L.M. Rio, A. Trampuz, G. Dietler, A. Bizzini, S. Kasas: Antibiotic-induced modifications of the stiffness of bacterial membranes, J. Microbiol. Methods **93**(2), 80–84 (2013)

19.197 P. Polyakov, C. Soussen, J. Duan, J.F. Duval, D. Brie, G. Francius: Automated force volume image processing for biological samples, PLoS One **6**(4), e18887 (2011)

19.198 G. Francius, P. Polyakov, J. Merlin, Y. Abe, J.M. Ghigo, C. Merlin, C. Beloin, J.F. Duval: Bacterial surface appendages strongly impact nanomechanical and electrokinetic properties of Escherichia coli cells subjected to osmotic stress, PLoS One **6**(5), e20066 (2011)

20. Contact-free Mechanical Manipulation of Biological Materials

Jörg Schnauß, Josef A. Käs, David M. Smith

In biotechnology and medicine, controlled studies on biological material are fundamental for developing new methodologies and therapeutic approaches. To explore the nature of biological processes and to test inherent material properties, it has become increasingly clear that new experimental methods must be developed in order to allow precise manipulations and quantification of biological materials on the microscopic level. Traditional methods often rely on physical contact with the sample, which can induce drastic artifacts in soft biological systems. To bypass this limitation, tools for contact-free manipulation were developed, which even enable the induction of whole-cell deformations to explore their mechanical properties. These approaches facilitate extensive investigations of single molecules, molecular ensembles, cells and even tissues, potentially reducing the need for animal studies. In this rapidly changing field, it is nearly impossible to provide a comprehensive overview of all available techniques since new methods are constantly being developed. In this chapter, we highlight many of the predominant approaches, aiming to investigate cellular as well as subcellular mechanical properties and processes.

20.1 **Interaction of Laser Light with Biological Material** 618
20.1.1 Conservation of Momentum – Gradient and Scattering Force 618
20.1.2 Quasiparallel Laser Light – Dual-Beam Laser Trap 619
20.1.3 Focused Laser Light – Optical Tweezers 620
20.2 **Optical and Magnetic Tweezers for Single-Molecule Manipulations** 621
20.2.1 Optical Tweezers 622
20.2.2 Magnetic Tweezers 623
20.2.3 Mechanical Measurements of Biomolecular Ensembles 624
20.3 **Optical and Electric Forces for the Manipulation of Whole Cells** ... 629
20.3.1 The Optical Stretcher 629
20.3.2 Cell Deformation via Optical Tweezers .. 634
20.4 **Hydrodynamic Shear Forces for the Manipulation of Single Cells** ... 636
20.4.1 Flow-Induced Shear Deformation of Cells 636
20.4.2 High-Throughput Mechanical Phenotyping of Cell Populations in Real Time 637
20.5 **Conclusion and Outlook** 638
References 638

Exploring mechanical properties of biological matter or quantifying cellular processes requires controlled manipulations on the microscopic or even nanoscopic scale. Traditional methods such as atomic force microscopy (AFM) are on the forefront of these techniques, but rely on physical contact with the sample. While this approach is sufficient for hard structures, soft biological systems are easily biased by the physical contact and drastic artifacts can be induced. Cells, for instance, are highly heterogeneous structures and results crucially depend on the cellular substructure that is attached to the measurement tool. To reliably quantify biological matter, it has become necessary to use contact-free techniques which are also compatible with the study of biological processes. Several methods have been developed to induce stress-dependent strains on the desired structures. Combined with optical microscopy, deformations can be visualized to derive stress-strain relations, enabling the determination of mechanical properties or the quantification of force-dependent processes.

To controllably manipulate biological structures, methods based on optical effects can induce forces in the pico- and nanonewton range, which even enable whole-cell deformations. Other approaches employ external magnetic or electrical fields to controllably cause strains, while smart designs of microfluidic setups induce large shear deformations of the desired structures at high throughput rates.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_20

20.1 Interaction of Laser Light with Biological Material

Since the days of *van Leeuwenhoek*, the inventor of the microscope, light has been the most powerful and intuitive tool for the investigation of cells. Moreover, light microscopy still remains at the forefront of investigative techniques today. Approaches such as stimulated emission depletion (STED) and statistical optical reconstruction microscopy (STORM) go far beyond the resolution limit of Abbe's criteria and are close to achieving molecular resolution [20.1–4]. The fact that light also carries momentum and angular momentum is less intuitive since it is not on a magnitude that can impact our daily lives, however as described herein, this duality renders light an equally well-suited tool to actively manipulate cells in addition to simply seeing them. The forces exerted on biological matter are in the pico- to nanonewton range and may appear small, however these forces are perfectly suited to noninvasively hold and manipulate individual biological cells. This was already recognized by *Ashkin* [20.5, 6] and *Burns* et al. [20.7] shortly after they discovered the fundamental principles of optical trapping.

Because of their average size of several micrometers, direct physical manipulation of whole biological cells is a challenge. Optical traps have therefore attracted the attention of biologists, providing access from the subnano- to the micrometer scale [20.6, 8].

Compared with other micromanipulation tools such as atomic force microscopes (AFMs) or micropipettes, optical traps are able to generate lower forces, in the pico- to nanonewton range, enabling the holding, moving, orienting, and repositioning of micro- and submicrometer objects as well as dynamic force measurements. Additionally, the manipulation is contact-free, which makes this technique perfectly suited for biophysical applications that need to avoid contamination.

Optical forces become very significant when going to small scales. With a simple *gedankenexperiment*, the highest possible light pressure can be estimated for a given light intensity: consider a ray of light, consisting of single photons that carry a certain momentum, hitting a mirror and getting reflected. According to conservation of momentum, the mirror has to experience a change in momentum twice as high as the momentum of all reflected photons. The force exerted is then given by the momentum change per unit time, resulting in a force of 6.6 nN for a light beam with a power of 1W. Besides the momentum change due to reflection, the momentum change of the photons due to refraction also gives rise to a force. This is the primary effect utilized in optical traps. Forces that arise in this manner are smaller than those arising due to reflection and depend on multiple parameters.

Because of the dual wave–particle nature of light, the particle-like picture of photons is true only within certain limits. The full theoretical explanation for the interaction of small dielectric particles with laser radiation strongly depends on the ratio of the particle diameter d to the wavelength λ of the laser light used. Bead diameters $d \gg \lambda$ are described by the Mie regime of ray optics, whereas for $d \ll \lambda$ the Rayleigh regime applies. The case $d \approx \lambda$ is highly nontrivial and requires the exact solution of the Maxwell equations within the given boundary conditions.

In optical traps, the wavelengths used typically range between 600 and 1300 nm, whereas the minimum diameter of eukaryotic cells is about 10 µm. This falls within the Mie regime, also known as the ray-optics regime, and thus the laser beam can be decomposed into individual rays that propagate in a straight line and are therefore described by geometric optics.

Regardless of the regime applied, laser-based particle trapping works only as long as the refractive index of the particle is higher than that of the surrounding medium, which holds for biological materials since their biomolecular content leads to higher refractive indices than water. Only special beam geometries (a donut-shaped mode, for instance) enable particle trapping with a lower refractive index in the region of lowest intensity, which will not be discussed within the presented framework [20.9].

20.1.1 Conservation of Momentum – Gradient and Scattering Force

In the limit $d \gg \lambda$ (Mie regime), the ability of light to exert forces on matter can be explained by simple transfer of momentum. For a vanishing rest mass, the relativistic energy-momentum relation predicts the momentum $\boldsymbol{p}$ carried by a photon as

$$\boldsymbol{p} = \frac{E}{c}\boldsymbol{e}_r \equiv n\frac{E}{c_0}\boldsymbol{e}_r \,, \tag{20.1}$$

where $E = h\nu$ is the energy of a photon of frequency ν, h is the Planck constant, c is the speed of light in the medium, c_0 the speed of light in vacuum, $n = c_0/c$ is the refractive index, and $\boldsymbol{e}_r$ is the unit vector in the direction of propagation.

As a ray of light, consisting of N photons, hits a particle of low opacity, it gets partially reflected or refracted. These interactions effectively change the momentum of each photon, and conservation of mo-

mentum requires a resulting momentum of

$$\begin{aligned}\boldsymbol{p}_{\text{res}} &= N\boldsymbol{p}_{\text{in}} - N\boldsymbol{p}_{\text{out}} = Nn\frac{E}{c_0}(\boldsymbol{e}_{\text{in}} - \boldsymbol{e}_{\text{out}}) \\ &= Nn\frac{E}{c_0}\boldsymbol{e}_{\text{res}} \end{aligned} \tag{20.2}$$

to be picked up by the particle. Here, the vector $\boldsymbol{e}_{\text{out}}$, is the superposition of the reflected and refracted fractions of the light rays, weighted by their respective intensities. Newton's second law allows the calculation of the resulting force, resulting in

$$\boldsymbol{F}_{\text{res}} = \frac{\mathrm{d}\boldsymbol{p}_{\text{res}}}{\mathrm{d}t} = n\frac{N}{c_0}\frac{\mathrm{d}E}{\mathrm{d}t}\boldsymbol{e}_{\text{res}} = n\frac{P}{c_0}\boldsymbol{e}_{\text{res}}\,, \tag{20.3}$$

where $\boldsymbol{e}_{\text{res}}$ is the vector as defined in (20.2), and P is the power of a ray of light consisting of N photons per unit time.

To describe an extended coherent light source composed of infinite single rays, it is appropriate to switch to an integral formulation of (20.3) using the intensity distribution $I(\boldsymbol{r})$, which represents the number of photons with energy E per unit area and time. The resulting force is then given by

$$\boldsymbol{F}_{\text{res}} = \frac{n}{c_0}\int I(\boldsymbol{r})\boldsymbol{e}_{\text{res}}(\boldsymbol{r})\mathrm{d}A\,. \tag{20.4}$$

The integral has to be performed over the whole illuminated surface of the trapped particle, and it is important to realize that the resulting vector $\boldsymbol{e}_{\text{res}}$ depends on the angle of incidence of the laser. In addition, $\boldsymbol{e}_{\text{res}}$ varies over the surface as does the intensity distribution $I(\boldsymbol{r})$. In general, appropriate expressions for $\boldsymbol{e}_{\text{res}}(\boldsymbol{r})$ and $I(\boldsymbol{r})$ that sufficiently characterize the system physically make the equation easily solvable, but nontrivial, especially for irregularly shaped particles. For many experimental situations, the assumption of a homogeneous, spherical particle is sufficient and largely avoids the need for numerical calculations. In the following discussion of parallel and focused laser intensity distributions, particles are therefore assumed to be spherical and homogeneous.

20.1.2 Quasiparallel Laser Light – Dual-Beam Laser Trap

Optical forces can be utilized to effectively trap particles in a dual-beam laser setup consisting of two counter-propagating, nonfocused Gaussian laser beams, which are slightly divergent. Gaussian beams have a light intensity gradient, with the highest intensity in the beam center (on the optical axis) and a decreasing intensity towards the outer off-axis positions. The beam width w increases with increasing distance z from the optical fiber according to

$$w(z) = w_0\sqrt{1 + \left(\frac{\lambda z}{n_1\pi w_0^2}\right)^2}\,. \tag{20.5}$$

Here, w_0 is the initial beam width (= mean field diameter of the optical fiber), λ is the wavelength of the laser light used, and n_1 is the refractive index of the surrounding medium. In order to understand the effects leading to a stable trapping of particles, one must estimate a bundle of parallel light rays at a distance z.

Consider a sphere of radius ρ located at a distance ρ_0 from the optical axis of a parallel Gaussian laser beam with a beam width w (Fig. 20.1). The intensity distribution $I(\boldsymbol{r}) \equiv I(\theta, \varphi, \rho_0)$ illuminating the particle is then given in spherical coordinates by

$$\begin{aligned} I(\theta, \varphi, \rho_0) &= \frac{2P}{w(z)^2\pi} \\ &\times \exp\left(-\frac{2\rho^2\sin^2(\theta) + \rho_0^2 - 2\rho_0\rho\sin(\theta)\sin(\varphi)}{w(z)^2}\right)\,. \end{aligned} \tag{20.6}$$

Equation (20.6) is the intensity distribution as required in (20.4).

To determine the vector, $\boldsymbol{e}_{\text{res}}$, consider a single ray of power P hitting a sphere at an angle of incidence θ with an incident momentum per second of n_1P/c_0. The

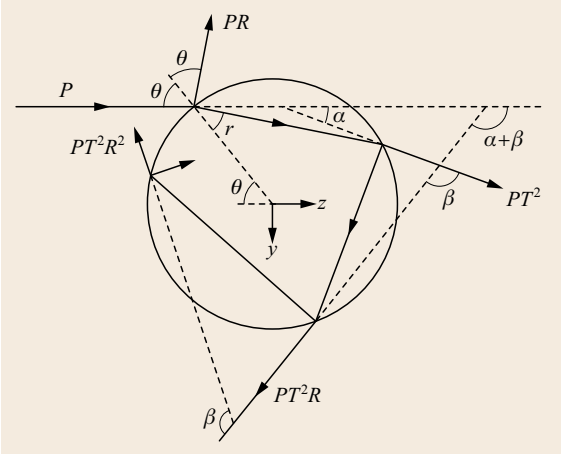


Fig. 20.1 A ray of power P hits a sphere at an angle θ where it partially reflects and partially refracts. These scattered rays make angles relative to the incident forward ray direction of $\pi + 2\theta, \alpha, \alpha + \beta, \ldots, \alpha + n\beta, \ldots$, respectively. Note the geometric relations $\alpha = 2\theta - 2r$ and $\beta = \pi - 2r$

total force on the sphere is the sum of contributions due to the reflected ray of power PR and the infinite number of emergent refracted rays of successively decreasing power $PT^2, PT^2R, \ldots, PT^2R^n, \ldots$, whereas $T(\theta)$ and $R(\theta)$ are the angle-dependent Fresnel coefficients for transmission and reflection.

The subsequent calculation of all appearing forces [20.5] yields a resulting force $\boldsymbol{F}_{\mathrm{res}}$, which is commonly separated into components parallel and perpendicular to the direction of propagation: the scattering force F_{scat} and the gradient force F_{grad}, respectively.

$$\boldsymbol{F}_{\mathrm{res}} = \sum_{\text{all photons}} \frac{\mathrm{d}\boldsymbol{p}_{\mathrm{res}}}{\mathrm{d}t} = \begin{pmatrix} F_{\mathrm{grad}} \\ F_{\mathrm{scat}} \end{pmatrix} , \tag{20.7}$$

where the components F_{grad} and F_{scat} are given by the analytically unsolvable expressions

$$F_{\mathrm{grad}} = \frac{n_1}{c_0} \int_0^{\pi/2} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi I(\theta, \varphi, \rho_0) \rho^2 \sin^2(\theta) \sin(\varphi) \times \left(R(\theta) \sin(2\theta) - \frac{T(\theta)^2 [\sin(2\theta - 2r) + R(\theta) \sin(2\theta)]}{1 + R(\theta)^2 + 2R(\theta) \cos(2r)} \right) \tag{20.8}$$

and

$$F_{\mathrm{scat}} = \frac{n_1}{c_0} \int_0^{\pi/2} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi I(\theta, \varphi, \rho_0) \rho^2 \sin(\theta) \cos(\varphi) \times \left(1 + R(\theta) \cos(2\theta) - \frac{T(\theta)^2 [\cos(2\theta - 2r) + R(\theta) \cos(2\theta)]}{1 + R(\theta)^2 + 2R(\theta) \cos(2r)} \right) . \tag{20.9}$$

In (20.8) and (20.9), n_1 stands for the refractive index of the surrounding medium, c_0 is the speed of light in vacuum, $R(\theta)$ and $T(\theta) \equiv 1 - R(\theta)$ are Fresnel coefficients describing reflection and transmission, and r is the refraction angle according to Snell's law.

A subsequent analysis of the expressions of the perpendicular components of the resulting force shows that the scattering force arising from a divergent laser beam pushes a trapped sphere away from the laser source, whereas the gradient force pulls it onto the beam axis.

Dual-beam optical traps consist of two counter-propagating laser beams, causing the scattering forces to cancel each other out in the middle. Gradient forces, however, are added up so that the particle is stably trapped on the beam axis.

20.1.3 Focused Laser Light – Optical Tweezers

For beams of complex shape, such as the highly convergent beams used in so-called optical tweezers, the calculation of arising forces becomes more complicated since the angle of incidence on the surface depends on both the polar coordinate θ as well as the relative position of the sphere from the focal point. Nevertheless, the concept of gradient forces and scattering forces can still be applied and at least allows a qualitative understanding of the underlying mechanism of trapping.

Consider a tightly focused laser beam, as illustrated in Fig. 20.2, and a spherical particle near the focal point [20.5]. The trap's underlying working principle for a dielectric sphere can be described in terms of the total force due to a representative pair of rays a and b of the converging beam when neglecting reflection. Using this simplification, the forces F_{a} and F_{b} are caused entirely by refraction and are pointing in the direction of the momentum change due to conservation of momentum (Fig. 20.2a,b).

In experiments, reflection cannot be excluded. Therefore, scattering forces appear in addition to the gradient forces (denoted as F_{a} and F_{b} in Fig. 20.2). For the successful trapping of a particle, the gradient force must exceed the scattering force. Otherwise, the particle will be pushed out of the focus in the direction of light propagation. In order to avoid this, an objective with a high numerical aperture is employed: the higher the aperture is, the tighter the focus, and, therefore, the higher the ratio $F_{\mathrm{grad}}/F_{\mathrm{scat}}$ becomes.

There are several factors influencing the efficiency of a focused laser beam trap. In addition to the laser power, the numerical aperture of the objective and the employed laser mode, the effective refractive index,

$$m = \frac{n_{\mathrm{medium}}}{n_{\mathrm{particle}}} , \tag{20.10}$$

has the greatest effect. Hence, the most convenient way to improve trap efficiency is to increase the effective refractive index.

Because of the complex set of parameters influencing the trapping force, the nontrivial and necessary numerical calculation, and the fact that the particles of interest often don't fulfill the Mie regime criteria $d \gg \lambda$, the real trapping force of optical tweezers has to be determined experimentally. A very common force-calibration method is based on the observation of the Brownian motion of a trapped particle, which assumes

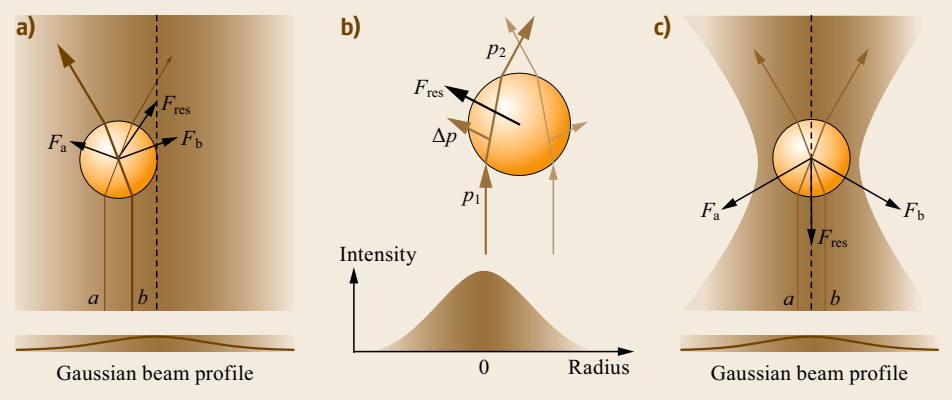


Fig. 20.2 **(a)** In a Gaussian laser beam the rays are refracted when entering and exiting a particle. This change of direction causes a change of momentum resulting in optical forces. In a straight beam these arising forces push the particle down the stream. **(b)** A restoring gradient force towards the beam axis occurs when the particle is off-center. **(c)** In the case of a tightly focused beam this gradient force points towards the beam waist and against the scattering force. These two forces are balanced downstream of the waist causing a stable trapped state (after [20.10])

a harmonic potential generated by the optical tweezers. The fluctuation $x(t)$ of a small particle within a viscous medium satisfies the well-known Langevin equation for the overdamped oscillator,

$$\gamma \frac{\mathrm{d}x}{\mathrm{d}t} + kx = F(t) \,, \tag{20.11}$$

where γ is the hydrodynamic drag coefficient, k is the spring constant of the harmonic potential, and $F(t)$ is the stochastic force experienced by the particle due to the thermal motion of the molecules in the solution. The power spectrum for $x(t)$ can be derived by taking the Fourier transform and the modulus squared, which gives a Lorentzian curve

$$S(f) = \frac{k_\mathrm{B}T}{\gamma\pi^2(f^2 + f_\mathrm{c}^2)} \,, \tag{20.12}$$

where f is the frequency, k_B is the Boltzmann constant, T is the sample temperature, and f_c is the corner frequency. By fitting the power spectrum of the measured positions $x(t)$ to the Lorentzian curve (20.12), the corner frequency f_c provides the desired spring constant k using the relation

$$f_\mathrm{c} = \frac{k}{2\pi\gamma} \,. \tag{20.13}$$

The harmonic potential $V(x)$, generated by optical tweezers, is therefore given by $V(x) = 1/2\,kx^2$, which is a good approximation for small displacements. Once calibrated, the displacement of a bead from the center of the optical trap can be associated with a force, thus allowing the performance of force measurements in the pico- to nanonewton range.

The described procedure has to be understood as a basic recipe for force calibration. The hydrodynamic drag coefficient γ depends on the viscosity of the surrounding medium, which is, in turn, temperature dependent. Because of the heating of the laser, the temperature of the medium varies locally, which has a direct effect on Brownian motion, and, consequently, on the fluctuation of the observed particle. More advanced calibration methods account for this and other errors [20.11, 12].

20.2 Optical and Magnetic Tweezers for Single-Molecule Manipulations

When Kepler looked through his telescope to observe comets roughly 400 years ago, he suggested that the deflection of their tails was caused by solar ray pressure. Although Newton predicted and Bartoli and Maxwell theoretically deduced forces induced by radiation, it was not until 1901 that Lebedev experimentally verified electromagnetic pressure. A few years later, *Debye* presented his dissertation regarding radiation pressure on spheres of different materials [20.13]. However, until the development of lasers as light sources with sufficiently high intensities, this principle could not be used to exert forces on microscopic objects. In 1970, *Ashkin*,

who was working at the Bell Telephone Laboratories, observed that freely suspended particles are accelerated by the forces of radiation pressure from visible laser light [20.8]. In 1986, he published his paper on the *Observation of a single beam gradient force optical trap for dielectric particles*, therein describing the first instance of using light to stably hold micro- and nanoscale objects on the size scale of cells in place [20.14]. This broad concept of manipulating microscopic matter with external fields – whether electromagnetic or magnetic – underpins the techniques that are nowadays known as *optical* and *magnetic tweezers*.

20.2.1 Optical Tweezers

With the invention of optical traps in the 1980s, it became possible to generate and measure forces in the unprecedented range of pico- and nanonewtons. For the first time, microscopic particles such as living cells or bacteria could be individually manipulated. Today, optical tweezers are the most common type of optical traps, and due to their simplicity, they are used in a wide range of different applications.

Their construction essentially only requires a laser and a lens, although applying several lenses and optical devices is useful for controlling the form and position of the laser beam. A conventional optical microscope with a high-aperture objective is sufficient to produce a diffraction-limited focal spot that can act as a trap and already has imaging potential. Furthermore, an important criterion for stable realization of the trap is a minimization of room-temperature variations, acoustic noise, and mechanical vibrations.

For the practical realization of optical tweezers, lasers in the near infrared, e.g., 800 or 1024 nm, are typically used, since biological samples show low absorption at these wavelengths. Currently, diode or fiber lasers are displacing common Nd:YAG or Ti:sapphire lasers due to their lower costs and maintenance needs.

In the simplest setup, the laser beam is broadened by a beam expander, i. e., a telescopic lens system, to fully illuminate the back aperture of the objective before being coupled into the light path of the microscope by a dichroic mirror. Some setups use piezoelectric microscope stages to move the trap with respect to the sample. To facilitate relocation of the trap without moving the stage, beam-steering devices such as acousto-optical deflectors (AODs) or galvanometrically controlled mirrors can be inserted into the beam path. The more common AODs are crystals with a high optical density in which high-frequency acoustic waves are generated and the incoming laser beam is diffracted due to the acousto-optical effect. By tuning the frequency and amplitude of the acoustic waves, the output angle and intensity of the laser beam can be rapidly controlled with great accuracy. Fast switching of the beam deflection angle even allows for simulating multiple traps. Besides controlling the direction of the beam, the amplitude can be attenuated using acousto-optical modulators (AOMs), which work on the same principles as AODs [20.15, 16].

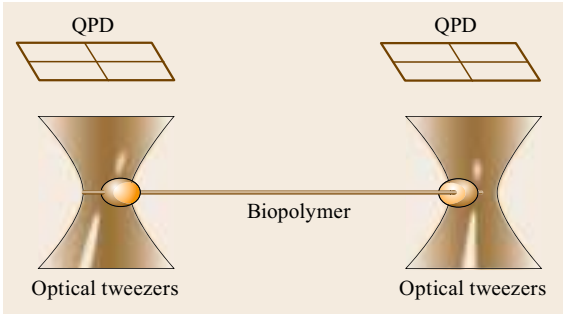


Fig. 20.3 The diffraction pattern of the optical tweezers can be recorded by QPDs and changes if the trapped particle is deflected from the trap center. Force calibrations of the trap allow an attribution of these changes to pulling forces, which can be used to derive stress-strain relations

These additional components allow precise movements of the trapping laser and accordingly the trapped particle, which can be recorded via an external camera. Attaching these particles, for instance to biopolymeric structures, allows an experimenter to controllably deform them in order to investigate their mechanical features and responses against external stimuli [20.17, 18].

If a CCD (charge-coupled device) camera is not sufficient as a position detection system, the setup can be expanded by a quadrant photodiode (QPD) system. After penetrating the sample, the laser light is recollimated by a second objective and projected onto the QPD. A trapped particle causes a diffraction pattern that only depends on its position with respect to the trap center [20.19]. External forces pulling the particle out of the trap center cause a change of the diffraction pattern. Calibrations of the tweezers facilitate the determination of the according forces (Fig. 20.3). Thus, stress-strain profiles for the probed structures (e.g., biopolymers) can be recorded to evaluate their mechanical properties or even monitor dynamic processes [20.18, 20–22].

Trapping of Macromolecules by Dynamic Temperature Fields

Recently, a highly novel trapping design has succeeded in eliminating the direct interaction of the incident light with the probe molecule, instead indirectly inducing the process through the plasmonically driven generation of

thermal fields [20.23–25]. This innovative methodology depends on the coupling of two fundamental physical principles working in unison – thermophoresis and surface plasmon resonance (SPR) – which are dynamically controlled to trap and direct the motions of nanoparticles and macrobiomolecules. Here, the basic trapping chamber is integrated with plasmonically active gold nanostructures using microfabrication techniques. The simplest setup resembles a circular cavity of approximately 5–6 μm in diameter, formed in a 50 nm gold film (Fig. 20.4), via microsphere lithography and thermal evaporation [20.25]. When focused laser light in the visible range (532 nm) is incident upon the edge of the gold structure, localized surface plasmon resonance leads to the conversion of optical energy to heat around the nanoscale ridge structure. This in turn creates a significant, but localized temperature gradient of several 10 K per μm within the surrounding solution. Viewed over the entire trap structure, this gradient can be seen as an inhomogeneous thermal field. For a large nanoparticle such as a 200 nm polystyrene bead [20.24, 25] or a long strand of lambda-DNA (deoxyribonucleic acid) [20.23] in solution, such a field generates a net directional force along one direction of the gradient, an effect termed thermophoresis [20.26]. This arises from a temperature gradient along the surface of the particle inducing an interfacial fluid flow, which acts to move the particle through the solution. In most cases, this leads to particle movement towards the cold region, although in fewer cases, movement in the opposite direction, termed *negative thermophoresis*, is observed.

In this setup, the laser-induced heating of the edges of the chamber leads to a gradient which grows cooler towards its center, causing a particle to become effectively trapped. By altering the laser power, and subsequently the magnitude of the gradient, the diameter of the trap can be controlled [20.25]. A more recent design included a dynamic feedback loop, where the position of the trapped particle is detected with a CCD camera, and the incident laser at the edge of the chamber is steered with the help of an AOD to optimize the thermal gradient [20.23]. This enabled a fast switching of the gradient in response to particle diffusion, leading to a higher trapping efficiency while effectively decreasing the necessary laser power, and the resulting temperature increment at the trapped particle. This ability to precisely tailor the thermal field offers a dual advantage. First, this allowed for smaller particles (with diameters down to 28 nm) and biological macromolecules (here, lambda-DNA) which would normally be difficult to hold in a static thermal gradient, to be stably trapped for long times. Additionally, a dynamic control over the overall thermal field opens the possibility for this methodology to be utilized for on-the-fly manipulation of the trapped objects.

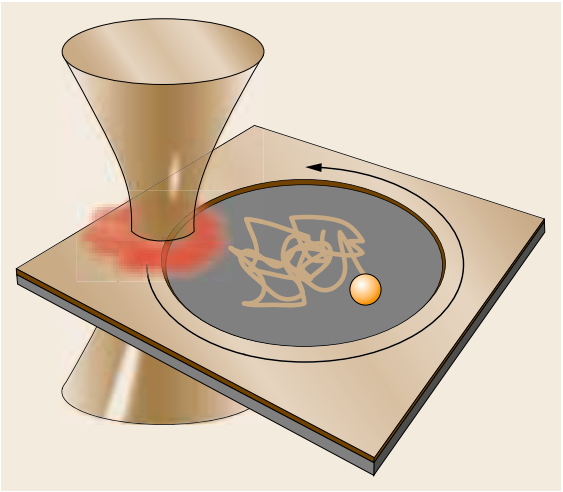


Fig. 20.4 Optically induced thermophoretic trap. A small circular chamber of a few μm in diameter is fabricated into a 50 nm thick gold film. Focused laser light incident upon the nanostructured edge causes local heating due to surface plasmon resonance. The resulting thermal gradient leads to a thermophoretic force on the particle towards the cooler areas. The resulting field from local heating around the edge of the chamber leads to an effective trapping of the particle

20.2.2 Magnetic Tweezers

Optical tweezers have some inherent limitations when used to probe biological samples. The laser light that heats the sample – possibly even to biologically incompatible temperatures which can lead to denaturation of proteins or biopolymers – may interact with other particles within the sample, or cause photodamage. A different, contact-free method for single-molecule or whole-cell experiments consists of so-called magnetic tweezers. Experiments involve attaching magnetic particles, e.g., beads comprised of paramagnetic nanoparticles, to the structures which are tested (Fig. 20.5). These particles can be manipulated by a setup exposing them to an external magnetic field $\boldsymbol{B}$, which induces a magnetic momentum $\boldsymbol{m}(\boldsymbol{B})$ proportional to the magnetic field $\boldsymbol{m}(\boldsymbol{B}) = V\chi\boldsymbol{B}/\mu_0$ with μ_0 being the vacuum permeability, V the volume of the particle, and χ the magnetic susceptibility. An induced magnetic moment saturates at a material-dependent value $\boldsymbol{m}_{\text{saturation}}$ in strong magnetic fields, and induced forces $\boldsymbol{F}$ can be derived from the potential $U = -1/2\,\boldsymbol{m}(\boldsymbol{B})\cdot\boldsymbol{B}$ for the according magnetic momentum [20.27].

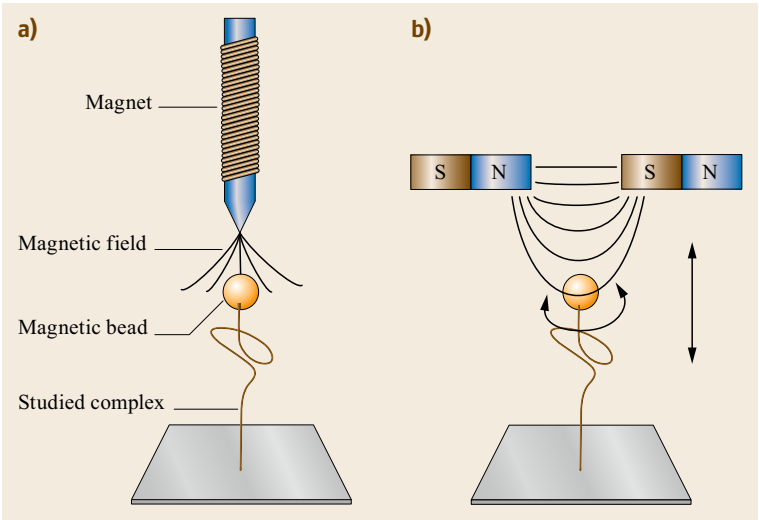


Fig. 20.5a,b Magnetic fields can be used to effectively manipulate magnetic particles. **(a)** A single magnet is sufficient to pull on a magnet bead inducing a stress-dependent strain in the studied structure. **(b)** Using at least two magnets allows not only a pull but also induction of torques on structures

Because of the highly specific interaction of the particle with the magnetic field, stresses are applied solely at the test structure (Fig. 20.5) leaving the rest of the sample practically unaffected.

Forces in the pico- and nanonewton range can be exerted to induce strains in the sample. Measurements are conducted via visual inspection and image acquisition to monitor the displacement of the bead as a response to the imposed magnetic field. Thus, the spatial and temporal resolution of the bead position depend highly on the properties of the recording camera. To evaluate not only the x- and y-position of the bead but also its vertical deflection or z-position, the diffraction rings around beads can be analyzed. These rings occur as an optical effect if the bead moves out of the focal plane of the objective. To precisely determine the z-position, the diffraction pattern has to be compared and normalized to a reference bead with known distances from the focal plane. During this calibration, the bead is fixed while moving the objective by defined distances, which later allows an attribution of specific diffraction patterns to an according z-position with a precision of a few nanometers [20.28]. When using an electromagnet-based setup, the field strength and its direction can be precisely adjusted by the amplitude and phase of the current for the magnets, which can be controlled by a feedback loop depending on the coordinates of the bead. Thus, the bead can be either held at a specific position or precisely moved by alterations of the magnetic field.

Monitoring and evaluating the exact position of the bead at a given time can be used to derive stress-strain relations. These facilitate the determination of the mechanical properties of the structure and further enable, for instance, the evaluation of binding events between subdomains. In addition to stretching forces, magnetic tweezers are also able to induce torsional strains. The magnetic particles themselves gain their special characteristics from embedded paramagnetic nanoparticles. Since the distribution of these nanoparticles within the bead inherently develops anisotropies, the arising magnetic moment is never perfectly aligned with the external magnetic field. Thus, torsional forces arise, aiming to align $\boldsymbol{m}$ and $\boldsymbol{B}$, which are usually sufficient to twist the tested molecule. Exerting torques is also possible with optical tweezers employing a circularly polarized laser beam and birefringent beads. However, magnetic tweezers are able to measure a multitude of the molecules in parallel since they are not limited to one focal spot.

20.2.3 Mechanical Measurements of Biomolecular Ensembles

The ability for optical and magnetic tweezers to accurately measure forces on the magnitude relevant to single biomolecules has made them invaluable tools for the quantitative characterization of the fundamental mechanisms and machinery underpinning the vast complexity of cells. Active force generation as well as processes driven by entropic or enthalpic mechanisms, respectively, can be readily quantified to complement the existing picture of cellular processes [20.29].

Measurement of Passive Force Generation

The predominant picture of biologically generated forces relies on active, energy-dissipating processes based on molecular motors. However, a series of studies employing optical tweezers has revealed that biomolecular ensembles can generate biologically relevant forces

without the need to convert energy into mechanical work. Experiments with pairs of microtubules or actin filaments, respectively, have proven that these biopolymers contract in crowded environments by entropic means – so-called depletion forces – to achieve a maximized overlap of their contours [20.21]. These biopolymers were each attached to a bead one micrometer in size, which were trapped by optical tweezers, similar to the configuration shown in Fig. 20.3. Emergent retraction forces deflected the beads off the trap center, corresponding to the magnitude of the contractile force on the order of 0.1 pN, independent of the overlap lengths of the biopolymers. Upon addition of a third rod, forces have been shown to be pairwise additive. Scaling up this additivity to a multifilament scale, like in the case of actin bundles, revealed very different dynamics within the contractile process when compared to the pairwise case. The magnitude of the contractile forces has not been reported to be constant as in the pair-wise case, but decayed exponentially during contraction, depending on the overlap of the constituting filaments. In this bundled case, maximal forces even reached up to 3.5 pN, a regime which corresponds to the activity of single molecular motors [20.18].

Besides depletion force-induced contractions, recent studies emphasize the role of crosslinking proteins for force generation. These proteins interconnect biopolymers and utilize their parallel alignment at high concentrations to maximize available binding sites. If several of these crosslinking proteins are confined, for instance between two axially shifted microtubules, their thermal motion leads to the generation of force aiming to drive the microtubules to a maximized overlap, i. e., to maximally available binding sites. When intentionally deflecting these rods by optical tweezers, available binding sites are reduced and the system exerts restoring forces similarly to a gas spring [20.22]. These forces are even sufficient to stall active molecular motors, a scenario which might explain the interplay of antiparallel microtubules in the midzone of the mitotic spindle forming during cell division [20.30].

Crosslinkers can also alter properties of biopolymer arrangements drastically and even add their own contributions to the system. An inherent property of most naturally occurring crosslinking proteins is that they constantly bind and unbind to the according biopolymer with discrete on- and off-rates. This so-called transient binding leads to rearranging effects and facilitates time-dependent responses of the entire structure. Actin bundles crosslinked by alpha-actinin and attached to a polystyrene bead, for instance, predominantly show elastic responses when deformed on a short time scale via optical tweezers. Deforming these bundles on longer time scales, however, reveals plasticity effects due to crosslinker rearrangements stabilizing the deformed configuration over time [20.17]. At the point of deformation, the bead is deflected off the trap center, since the stressed bundle aims to restore its equilibrium position. The involved crosslinkers rearrange in a time-dependent manner and stabilize the deformed configuration, which in turn slowly releases the stress, allowing the bead to slowly relax into the trap center again.

Optical tweezers-based experiments can be used to probe the mechanical properties of a wide range of biopolymers in reconstituted systems. So far, only studies on cytoskeletal proteins have been described, but this versatile tool is also widely employed to investigate DNA structures and DNA-related processes. Attaching the desired DNA structure to beads enables optical-tweezers-assisted force-clamp measurements or stretching experiments to quantify the according mechanical properties of the probe molecule. Force-clamp measurements are especially suitable to determine forces for DNA condensation processes. An inherently important process is the compacting of DNA via histones in order to condense this carrier of genetic information into the tiny space inside the nucleus. These histones are protein complexes able to exert sufficiently high forces to fold DNA appropriately in order to achieve a high packing density. Force-clamp measurements have revealed that histone activity is stalled only for relatively high forces around 40 pN and that two characteristic force regimes arise for the histone-like protein from *Thermotoga maritima* [20.31].

Measurement and Manipulation of Actively Driven Molecular Motors

Classically, the initial groundbreaking work for single-molecule characterization of biological materials with optical tweezers was based on the study of individual molecular motors such as kinesin [20.32, 33] and myosin [20.34, 35]. These translocate along microtubules or actin filaments, respectively, while converting chemical energy in the form of adenosine triphosphate (ATP) to kinetic motion. In early studies, motor proteins were attached to the surface of microscopic beads consisting of either silica or polystyrene, which could be manipulated by an optical trap during the course of translocation along the corresponding filament type (Fig. 20.6). By applying an optical force against the direction of motion until the bead-motor complex stalled [20.33], or conversely by reducing the force until motors were able to escape the trap [20.34], the force generated by single motors could be deter-

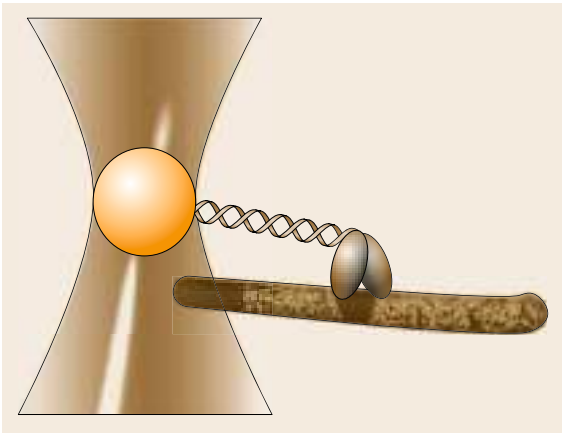


Fig. 20.6 Manipulation of a translocating molecular motor with an optical tweezers setup. A molecular motor (e.g., kinesin or myosin) undergoes directed motion along its corresponding filament type (microtubule or actin, respectively). The motor is physically tethered to a bead, which can be actively manipulated in order to determine physical properties of the motors such as stalling force

mined with single pN resolution. Measurements revealed that individual kinesin motors are capable of exerting roughly 5−6 pN of force during translocation along a microtubule, while motile subunits of myosin II generate between 1−5 pN of force along actin filaments [20.35].

Similarly, the ATP-dependent translocation of DNA-specific motors has been heavily investigated with the help of magnetic tweezers. One set of studies in particular focused on the DNA enzyme EcoR124I [20.36, 37]. This is a multisubunit motor where a core unit recognizes and binds a specific DNA sequence, then the two flanking subunits translocate in a unidirectional manner along the DNA strand. This draws the DNA towards the central unit, thereby creating pairs of DNA loops. The result is an overall decrease in the end-to-end distance, which can be analyzed with the help of a standard magnetic tweezers setup. Here, the translocation of a single flanking subunit along a DNA strand was investigated in order to determine properties such as the translocation rate, dependence on ATP concentration and generation of helical torque [20.36] as well as step size and ATP coupling efficiency [20.37]. A long DNA segment was anchored on one end to the surface and bound on the other end to a magnetic bead, which was used to stretch and supercoil the DNA molecule with the help of two external magnets. By monitoring the end-to-end distance during single translocation events, a translocation speed of 550±30 DNA base pairs per second was found at a saturating ATP concentration of 4 mM, and the quantification of supercoil unwinding demonstrated a helical pitch for the motor, matching that of the DNA double-helix. Later studies confirmed a basic step-size of one base with approximately one ATP molecule consumed during each step.

The ability to apply torsional forces to beads in a magnetic trap has also been instrumental in uncovering the chemical details of molecular motors exhibiting a rotational motion rather than directed linear translocation. Most famously, the coupling of mechanical work to the chemical synthesis of ATP molecules by the rotary F1-ATPase molecular motor has been determined on the single-molecule level [20.38, 39]. This motor is part of the ATP synthase complex responsible for generating ATP in biological systems, by coupling proton gradients to chemical synthesis. When isolated in vitro, the F1 fragment hydrolyzes ATP molecules at the rate of three catalytic events per full counterclockwise (viewed from above) turn. By attaching a magnetic bead to the central rotary γ subunit, it was shown that rotations in the reverse (clockwise) direction, mechanically driven by magnetic forces, led to the generation of ATP with up to 77% mechanochemical coupling efficiency [20.39].

Measurement of Specific Ligand Recognition

Specific receptor-ligand recognition events can also be characterized on a single-molecule level through dynamic force spectroscopy enabled by optical tweezers. In particular, the specificity of monoclonal antibodies (mAb) to their target ligand can be investigated by optical tweezers-enabled dynamic force spectroscopy in order to precisely determine binding kinetics and determine possible off-target cross-reactions [20.40]. This was carried out for the mAb HPT-101, which is known to bind to specific markers on proteins that are potential early indicators for Alzheimer's disease. Previous studies in steady-state, equilibrium conditions using the enzyme-linked immunosorbent assay (ELISA) have indicated that HPT-101 binds exclusively to the double-phosphorylated tau-proteins. While no indication of off-target binding to the corresponding mono-phosphorylated variant has been seen in ELISA, this does not reveal quantitative details of the dynamic binding process where all cross-reactions would be evident.

Here, a two-bead setup was used, where one is trapped in optical tweezers, while the other is fixed on the tip of a micropipette and can be spatially manipulated, e.g., retracted with a known velocity. The mAb is immobilized on the bead in the laser trap, and the target region of the tau-protein – either the double- or one of the two mono-phosphorylated derivatives – is immobilized on the pipette-fixed bead (Fig. 20.7).

By initially bringing the beads into contact to establish mAb-peptide binding and then retracting the pipette-bound bead from the trap until rupture is observed, the frequency of binding events was measured. This unexpectedly revealed that off-target binding to the two mono-phosphorylated derivatives does indeed occur at a comparable frequency (0.20 ± 0.12 and 0.15 ± 0.04) to the standard double-phosphorylated target (0.31 ± 0.08). The apparent discrepancy to earlier ELISA results was resolved by determining the average bond lifetime through force-induced unbinding events. By measuring the average force required to rupture the mAb-peptide bond through the application of a force ramp (with an applied loading rate of $56 \pm 7\,\mathrm{pNs}^{-1}$), the average bond lifetime was calculated. This showed that the bond between the mAb and double-phosphorylated target has a lifetime of approximately 10 and 40 times that of the two off-target mono-phosphorylated variants, explaining the lack of observed binding at steady-state conditions. A further study utilized the same range of tweezers-based techniques to scan the sequence of the target region and identify distinct amino acids that were essential to the specific binding of the mAb [20.41]. Such a precise kinetic analysis is not possible with equilibrium conditions in standard ELISA assays, and offers important information that can be used to selectively optimize antibody specificity in a deterministic manner.

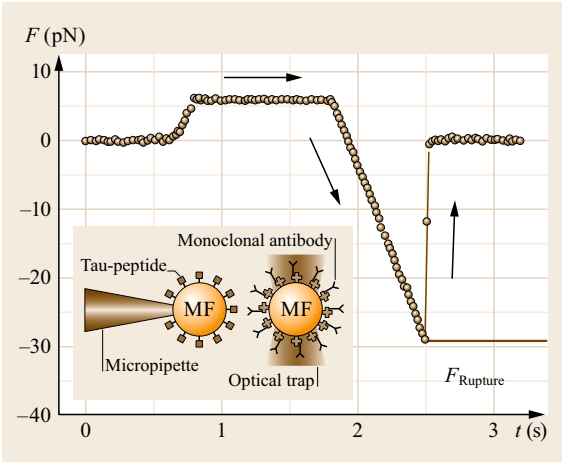


Fig. 20.7 Specific ligand recognition characterized by an optical tweezers setup. (*Inset*) The ligand of interest is immobilized on a micropipette-bound bead which can be freely manipulated, while the receptor, in this case an antibody, is on a separate bead held by a force-calibrated optical trap. Bringing the beads into contact induces ligand-receptor binding, and subsequent retraction to the point of bond rupture allows the quantitative examination of bond kinetics [20.40]

Bottom-Up Construction of Tools for Micromechanical Characterization

More recently, advances made in the last decade for the bottom-up, structural and mechanical programming of materials, starting from the single-molecule level, have facilitated a new range of complementary tools to enrich these micromechanical techniques. Most prominent amongst these is the advent of structural DNA nanotechnology, which enables the construction of discrete objects of defined size and shape, with molecular precision. Approaches such as DNA origami [20.43] or tile-based DNA construction [20.44, 45] are based on the self-assembly of sequence-defined sets consisting of anywhere from a few to several hundred synthetically produced DNA oligonucleotides. Through the hybridization of strands with complementary sequences based on the Watson-Crick nucleobase pairing of adenine to thymine and guanine to cytosine, complex objects can be designed and constructed with structural and functional resolution on the size scale of a single base-pair (3.4 Å). Beyond their potential for biomedical applications [20.46], these hold a particular importance for the micromechanical techniques described in this chapter, since their underlying structural properties, e.g.,

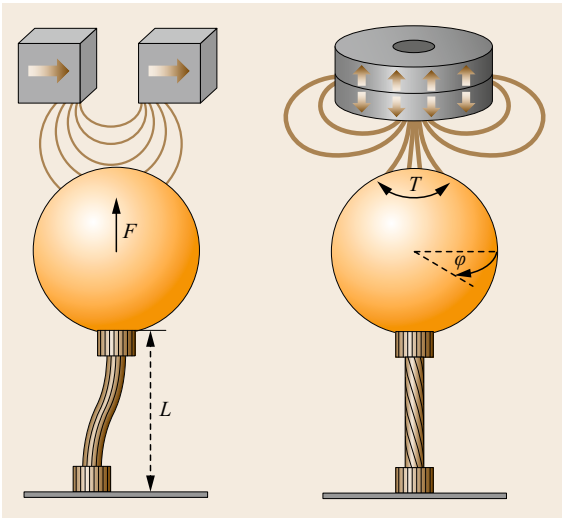


Fig. 20.8 Probing the bending and torsional rigidity of a rod-like DNA nanostructure with magnetic tweezers. Rod-like DNA origami structures are fixed on one end to a surface and on the other to a magnetic bead. By applying shear or torsional force from external magnets, the bending ($l_{p,b}$) and torsional ($l_{p,t}$) persistence lengths can be determined [20.42]

bending [20.47, 48] or torsional rigidity [20.42], can be deterministically controlled, opening possibilities for

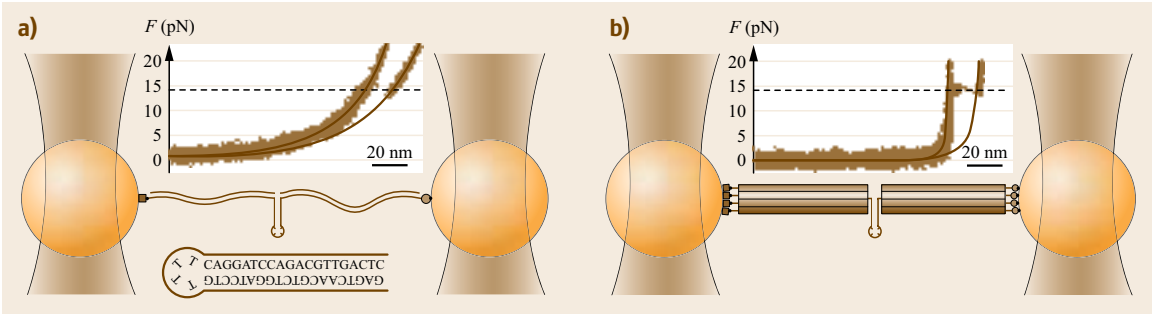


Fig. 20.9a,b Rigid DNA-based beams for noise suppression in optical tweezers-assisted single-molecule force spectroscopy. **(a)** Flexible double-stranded DNA linkers between optically manipulated beads and a macromolecule of interest (here, 20-base-pair DNA hairpin) lead to a high amount of noise and inability to precisely determine conformational transitions in the low-force (single-pN) regime. **(b)** By contrast, rigid beam connectors constructed from DNA origami suppress Brownian fluctuations, thereby increasing the resolution of observed transitions [20.49]

their use as mechanical standards or thermal noise suppressors.

One recent study [20.42] points towards the ability to use these bottom-up methods for fabricating precisely tailored mechanical calibration standards. Semi-flexible, rod-like structures were formed using the DNA origami method, consisting of parallel bundles of double-stranded DNA (dsDNA) arranged in a tube-like configuration. Two such structures were characterized in this study: a so-called six-helix bundle (6HB of six parallel helices, measuring 371 nm long and approximately 8 nm in diameter, and a four-helix bundle (4HB) with a length of 478 nm and a diameter of 5 nm. The bending and torsional rigidity of each species was determined by magnetic tweezers (Fig. 20.8), and compared to the standard of a dsDNA segment with a length of 428 nm. Compared to dsDNA, both rod-like DNA origami *n*HBs (n = number of helices) showed a significant increase in rigidity. Bending rigidity was seen to approximately scale with n^2, with the persistence length $l_{p,b}$ increasing from 50 nm for dsDNA to 740 ± 140 nm for 4HBs and 1880 ± 270 nm for 6HBs. The torsional rigidity had a weaker scaling proportional to n, with $l_{p,t}$ increasing from 97 ± 4 nm for dsDNA to 390 ± 30 nm for 4HBs and 530 ± 20 nm for 6HBs. This ability to deterministically *program* the material properties of discrete objects can be applied to craft mechanical rulers for the precise calibration of magnetic and optical tweezers systems.

Rigid DNA structures can also be utilized to suppress inherent thermal noise in single-molecule force measurements by optical tweezers. Measuring conformational transitions in single proteins or other macromolecules often involves the necessity to measure small length changes on the order of single nanometers, and force regimes less than 10 pN. Usually, the molecules of interest are attached at a distance to the trapped beads by dsDNA segments or some other predominantly flexible linker. Measurements at this high resolution are limited in the case of conventional force spectroscopy by optical or magnetic tweezers due to inherent thermal noise arising from Brownian fluctuations of beads and linkers. A recent study [20.49] showed that rigid beam-like structures from DNA could suppress noise in this low-force regime (Fig. 20.9). 250 nm beams consisting of 10 parallel double helices (10HB) in a roughly block-like configuration were constructed using the DNA origami method, and found to be completely rigid. As a proof of concept, a 20-base DNA hairpin, known to have an unzipping force of approximately 15 pN, was used as a standard probe molecule. Force-extension curves for the unzipping transition were compared for both the rigid beam linkers as well as for more flexible dsDNA linkers (modeled as a worm-like chain with $l_p = 50$ nm). Noise suppression by the rigid DNA origami linker at loads of 14 pN provided greater detail of information about the energy landscape of the hairpin transition. This included transiently populated substrates and sharper barriers which were masked with the flexible DNA linkers. This highlights the ability of these bottom-up methods to not only give cleaner, less noisy data, but also enable the acquisition of basic physical information which is otherwise unavailable through conventional means.

Novel DNA-based architectures can also be designed for assisting in mechanochemical analysis of molecules, where chemical processes are coupled to mechanical changes detectible by force spectroscopy methods. In one such study, an interconnected chain of seven DNA origami segments was used as a means to improve upon not only the detection limit and time resolution of mechanochemical sensing events, but also

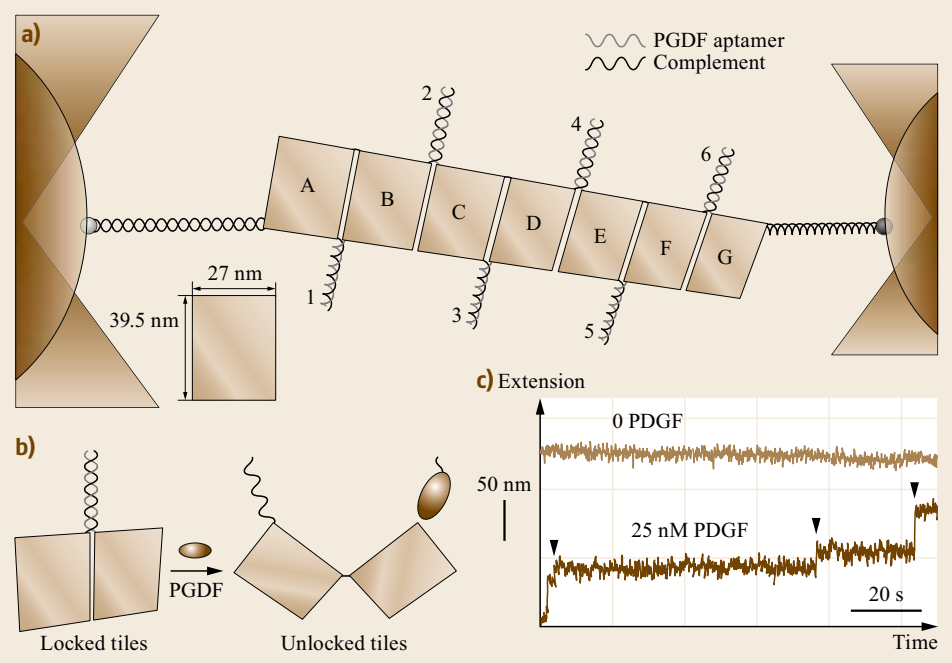


Fig. 20.10a–c Enhanced mechanochemical sensing with the aid of DNA origami. **(a)** A DNA origami structure with seven interconnected segments is maintained in a *closed* conformation by aptamer-based locks for a target molecule while held at a constant force by a dual-trap optical tweezers setup. **(b)** Upon attachment of the target molecule (here, PDGF) to the aptamer, the lock is opened, creating an outstretched configuration which **(c)** can be detected as jumps in the relative force-distance curves of the two beads [20.50]

enable simultaneous multiplexed sensing of multiple targets [20.50]. Aptamer-based recognition elements for platelet-derived growth factor (PDGF) were integrated as locking elements between neighboring segments in such a way that binding of the molecule would release the lock, thereby inducing a mechanical change (Fig. 20.10a,b). When tethered between two beads held at a constant force of 8 pN in an optical tweezers setup, this associated elongation of the overall structure could be detected by distinct jumps of approximately 40 nm in the force-extension curves (Fig. 20.10c). Compared to a mechanochemical sensor with only a single recognition element, the integration of six parallel elements on the DNA origami-based backbone reduced the lower detection limit by a factor of 10 from 100 pM to 10 pM, while simultaneously reducing the necessary measurement time from 30 to 10 min. The same approach was additionally used to enable multiplexed sensing of multiple targets. In this case, PGDF recognition elements and an additional element unlocked by the hybridization of a complementary DNA sequence were coupled into the segmented DNA origami structure in a way so that each conferred a distinguishable elongation of either 40 or 15 nm, respectively. With this variation, the simultaneous detection of PDGF at 25 nM and the target DNA sequence at 5 μM in the solution was possible.

20.3 Optical and Electric Forces for the Manipulation of Whole Cells

Moving beyond single molecules and biomolecular ensembles, the contact-free interaction of external fields – in particular the aforementioned electromagnetic and inhomogeneous electric fields – offers unique opportunities for manipulating discrete biological objects such as whole cells with unprecedented precision. This has been responsible for carving out cutting-edge breakthroughs for rheological characterization of biomaterials beyond the bulk scale, and down to the single-cell or even subcellular level of detail.

20.3.1 The Optical Stretcher

While *Ashkin* is known for the invention of the single-beam gradient force laser trap, he also proposed a two-beam laser trap using divergent laser light [20.8]. Since the momentum of a photon is changed not only by the particle as a whole, but at each transition from one medium to another (Fig. 20.11), forces actually act on the surface of a trapped particle according to (20.3). The transition of a ray of light from a medium of refractive index n_1 to the particle of refractive index n_2 is accom-

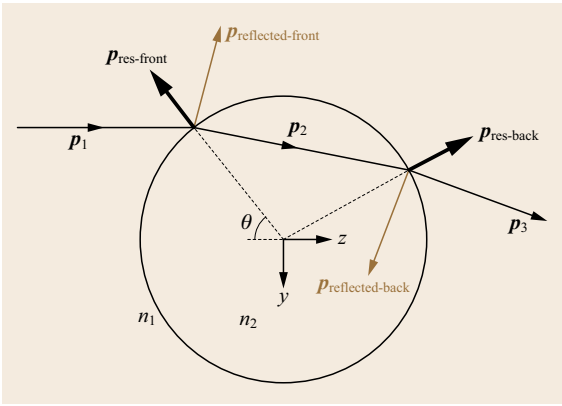


Fig. 20.11 Surface forces exerted by the transition of a single ray passing a spherical particle. By using Snell's law, it can be shown that the resulting momenta are always perpendicularly oriented to the transmitted surfaces, from higher to lower refractive index, independent of the particle's shape

panied by a resulting momentum of

$$\boldsymbol{p}_{\text{res-front}} = N(\boldsymbol{p}_1 - T\boldsymbol{p}_2 - R\boldsymbol{p}_{\text{reflected-front}}) \qquad (20.14)$$

and as the ray exits, another resulting momentum of

$$\boldsymbol{p}_{\text{res-back}} = TN(\boldsymbol{p}_2 - T\boldsymbol{p}_3 - R\boldsymbol{p}_{\text{reflected-back}}) \qquad (20.15)$$

occurs, where the influence of reflection on the front and on the back side is included by $\boldsymbol{p}_{\text{reflected}}$.

In these terms, the resulting force in (20.7), acting on the whole particle, has to be understood as the sum over all infinitesimal surface forces, (20.14) and (20.15).

In 2000, *Guck* et al. presented their two-beam laser trap – dubbed the Optical Stretcher (OS) – which realized that a two-beam laser trap could not only trap cells but also deform them thereby probing their biomechanical properties [20.51].

Two counter-propagating, divergent laser beams induce surface forces on a spherical object (as illustrated in Fig. 20.12) and give rise to a stress profile as shown in Fig. 20.13a. Scattering and gradient forces act in such a way that a cell entering the vicinity of the counter-propagating beams is centered and stably held in the middle of the trap. Once stably trapped, the resulting force acting on the whole particle is zero even though surface forces still act upon it [20.52]. Generally, whenever light enters or exits a dielectric medium it exerts a force pointing away from the optically denser medium normal to its surface. These surface forces depend on various parameters such as the applied laser power, the refractive indices of the cell and the surrounding medium, the cell size, the beam width at the center of the trap, and the wavelength of the laser beam.

Using appropriate laser powers yields surface forces, which are sufficient to effectively stretch a cell (Fig. 20.13b) [20.53]. This effect enables contact-free, whole-cell elasticity measurements and allows a determination of the mechanical properties of suspended, biological material in general [20.54, 55]. Once a cell is trapped, the magnitude of the stretch stress can be varied simply by tuning the intensity of the laser beam. Captured by a CCD camera, the deformation is analyzed using edge-detection software, thereby permitting the extraction of numerical data and the calculation of the cells' mechanical properties. The extent of a cell's deformation at a certain applied stress is known as its deformability. More specifically, the term *optical deformability* describes the elongation of the cell along the laser axis, normalized by the acting maximal forces. This deformability gives information not only about the cell's overall mechanical properties but also its underlying, cytoskeletal cellular structures, which can be used to compare different cells. Assuming similar refractive indices and sizes, a cell that deforms more at a given applied stress is softer than a cell that deforms less.

Compared to focused laser-beam traps, divergent laser-beam traps have a great advantage of exposing the trapped biological material to lower laser intensities, and therefore avoiding radiation-induced damage of living samples. The typical light intensities of optical

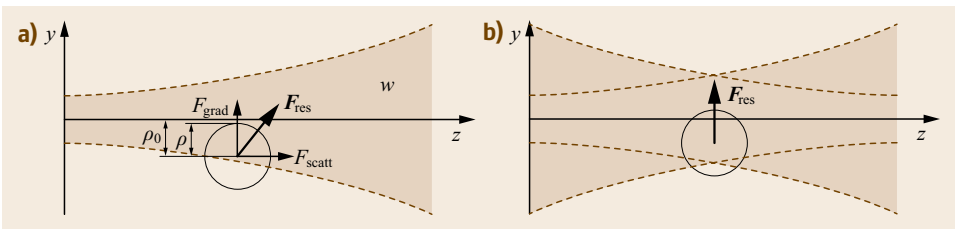


Fig. 20.12 **(a)** Forces exerted by a single Gaussian laser beam. The gradient force pulls the sphere onto the beam axis, whereas the scattering force pushes it away from the laser source. **(b)** A sphere in a divergent, dual-beam trap. The scattering forces of the two lasers cancel each other out in the middle, whereas the gradient forces are summed up, leading to stable trapping

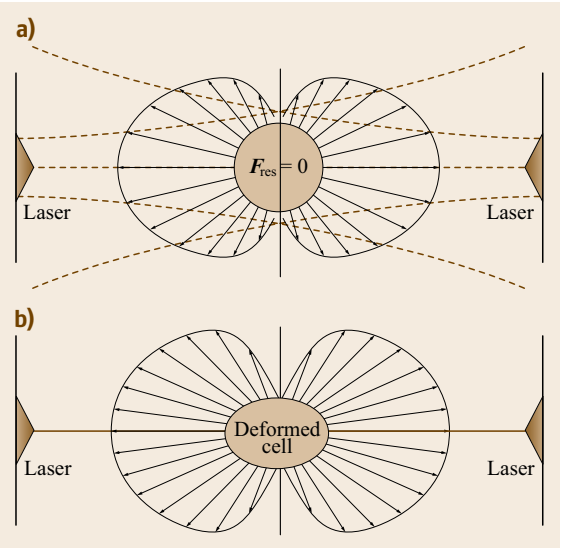


Fig. 20.13 **(a)** Numerical calculation of the stress profile of a spherical particle centered on the optical axis inside the optical stretcher, assuming parallel light beams. When light enters a dielectric medium, forces are induced normal to the surface and pointing away from the optically denser medium. In spite of a vanishing resulting force, stretch forces are applied on the surface. **(b)** At appropriate laser powers, exerted surface forces yield a stretch deformation of the trapped cell along the direction of the laser axis, with the highest stretching forces on the beam axis and virtually no forces near the poles (after [20.53])

tweezers, which would not lead to damage of biological samples, can be sufficiently used to trap particles, but induced forces are too weak to lead to deformations as the basis for rheology experiments. As compared with other techniques for single-cell elasticity measurements such as AFM, the OS allows one to measure up to 250 cells per hour, providing better statistical results.

Contact-Free Cell Rheology with the OS

Structural changes of the cytoskeleton cause changes in the viscoelastic response of the cell. This can be used to investigate diverse characteristics, as previously mentioned. Optical deformability is, however, an OS-specific quantity. To make it comparable with results obtained from other elasticity measurement techniques, a model for calculating the complex shear modulus was developed.

When a deforming force is applied to a material, it will respond elastically, viscously, or viscoelastically. The latter is an intermediate form between purely elastic or viscous behavior, and is observed in cells. The different responses are illustrated in Fig. 20.14.

For linear elastic deformations, the tensile creep compliance $D(t)$ can be described in terms of stress $\sigma(t)$ and the relative deformation (strain) $\gamma(t) = \Delta x/x$

$$\sigma(t) = E\gamma(t) = \frac{1}{D(t)} , \tag{20.16}$$

where $E(t)$ is the Young's Modulus. Linear viscous deformation is given by

$$\sigma(t) = \eta \frac{\mathrm{d}\gamma(t)}{\mathrm{d}t} . \tag{20.17}$$

There are various models for viscoelastic behavior. The general relationship between stress and strain can be expressed by

$$\sum_{i=0}^{n} a_i \partial_t^i \gamma(t) = \sum_{j=0}^{m} b_j \partial_t^j \sigma(t) . \tag{20.18}$$

In typical stretch experiments, cells are exposed to a step stress and the time-dependent strain is recorded. Terminating (20.18) at the second-order expansion of the strain and at first-order expansion of the stress and neglecting zeroth-order strain terms leads to

$$a_1 \partial_t \gamma(t) + a_2 \partial_t^2 \gamma(t) = \sigma(t) + b_1 \partial_t \sigma(t) , \tag{20.19}$$

which has been found to be sufficient to fit the typical response of a cell to a temporal stress [20.56]. A temporal step stress can be described by

$$\sigma(t) = F_\mathrm{G} \sigma_0 \theta(t)\theta(t_1 - t) , \tag{20.20}$$

where F_G is a geometric correction factor that takes into account the geometry of the cell and the shape of the applied stress field with respect to the peak stress σ_0 [20.57]. Here, t is the time from the initiation of the step stress, and t_1 is the duration of the applied stress. The deformation behavior of a cell is primarily determined by its actin network, which is assembled as an isotropic cortex in suspended cells and can therefore be modeled as a thick shell [20.57]. Combining (20.19) and (20.20), the strain becomes

$$\gamma(t) = F_\mathrm{G}\sigma_0 \left(\frac{b_1}{a_1} - \frac{a_2}{a_1^2}\right)\left[1 - \exp\left(-\frac{a_1}{a_2}t\right)\right] + \frac{F_\mathrm{G}\sigma_0}{a_1}t \tag{20.21}$$

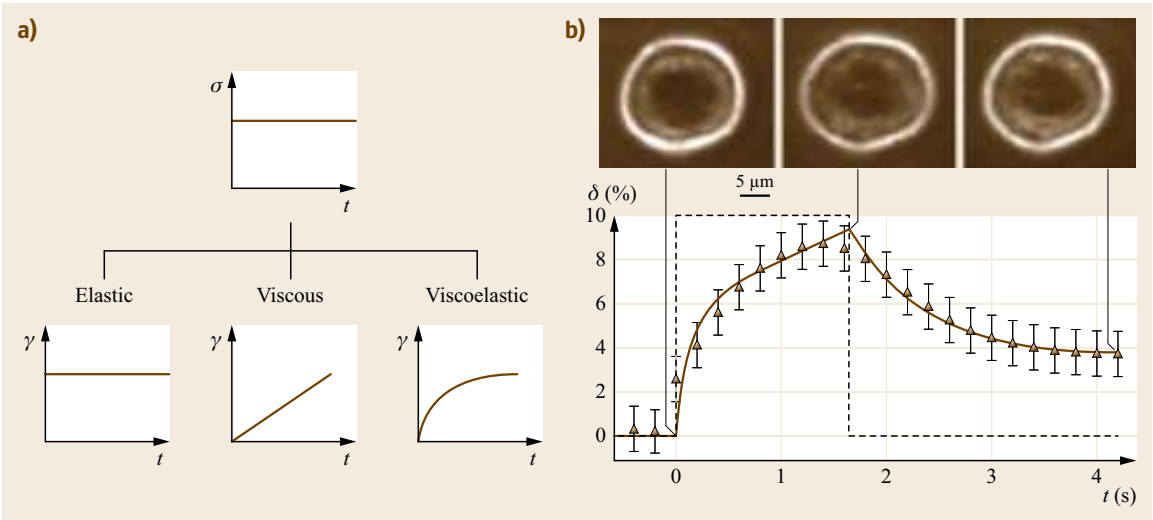


Fig. 20.14 **(a)** Different responses to a constant stress σ. Elastic materials show an immediate response while viscous materials expand in proportion to time. Viscoelasticity, as a combination of elastic and viscous behavior, displays a retardation of the response to the applied stress. **(b)** Typical deformation behavior of a cell when a step stress is applied. After increase of the laser power from 0.1 to 1.7 W at $t = 0$ s, an instantaneous deformation of the cell is observed, followed by a slower increase to a maximum strain value. Reducing the power back to trapping levels, the strain decreases but does not completely relax back to the initial value. The solid line shows a fit of the strain function ((20.21)–(20.22)) to the deformation data and the dashed line represents the laser power signature. Data points were binned to five points per second. The three phase contrast images show the cell prior (0 s), during (until 1.7 s), and after application (up to 4.2 s) of the step stress

for $0 < t < t_1$, the interval over which the stress is applied, and

$$\gamma(t) = F_G\sigma_0\left(\frac{b_1}{a_1} - \frac{a_2}{a_1^2}\right)\left[1 - \exp\left(-\frac{a_1}{a_2}t\right)\right] \times \exp\left[-\frac{a_1}{a_2}(t - t_1)\right] + \frac{F_G\sigma_0}{a_1}t_1 \tag{20.22}$$

for $t > t_1$.

In order to obtain the constants a_1, a_2 and b_1, the function is fitted to the measured data of the time-dependent strain. The time-dependent shear modulus, $G(t)$, is defined by the ratio of mean stress to mean strain. The strain behavior, $G(t)$, is given by

$$G(t) = \frac{1}{2(1+\mu)}\left[\frac{a_1b_1 - a_2}{b_1^2}\exp\left(-\frac{t}{b_1}\right) + \frac{a_2}{b_1}\delta(t)\right], \tag{20.23}$$

where μ is the Poisson's ratio and considered to be $\mu = 0.45$ for a shell-like actin cortex [20.58]. $G(t)$ can be converted via Fourier transformation into a frequency-dependent complex shear modulus $G^*(\omega) = G'(\omega) + \mathrm{i}G''(\omega)$ with

$$G'(\omega) = \frac{1}{2(1+\mu)}\left(\frac{\omega^2(a_1b_1 - a_2)}{1 + \omega^2b_1^2}\right), \tag{20.24}$$

and

$$G''(\omega) = \frac{1}{2(1+\mu)}\left(\frac{\omega a_1 + \omega^3a_2b_1}{1 + \omega^2b_1^2}\right). \tag{20.25}$$

The storage modulus, $G'(\omega)$, yields the elastic component of the cell's response to the applied stress. The loss modulus, $G''(\omega)$, gives the viscous component. From these functions, one can obtain characteristic material constants [20.56] such as the so-called rubber plateau,

$$G_{\mathrm{Plateau}} = \lim_{\omega\to\infty} G'(\omega) = \frac{1}{2(1+\mu)}\frac{a_1b_1 - a_2}{b_1^2}, \tag{20.26}$$

the fluid-to-solid transition frequency,

$$\omega_{\mathrm{f\text{-}s}} = \frac{1}{2}\left(\frac{a_1}{a_2} - \frac{1}{b_1}\right) - \sqrt{\frac{1}{4}\left(\frac{a_1}{a_2} - \frac{1}{b_1}\right) - \frac{a_1}{a_2b_1}}, \tag{20.27}$$

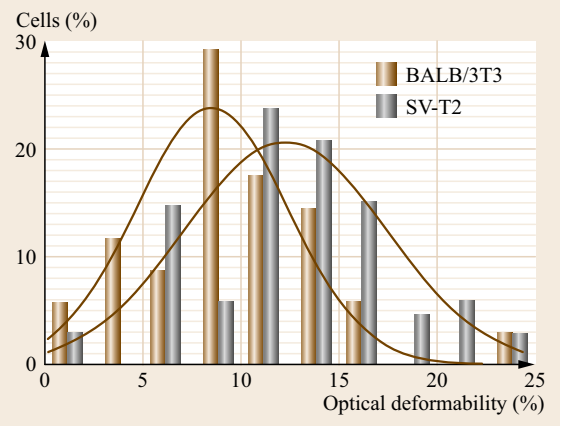


Fig. 20.15 The deformation of fibroblasts in an Optical Stretcher. A BALB/3T3 fibroblast deforms by 6.48 ± 0.36% along the laser axis, as determined by image analysis, when the light power is increased from 0.1 to 1.7 W in both beams. Measuring large numbers of cells reveals that the optical deformability of malignantly transformed SV-T2 fibroblasts is significantly shifted to higher values compared to normal BALB/3T3 fibroblasts (after [20.61])

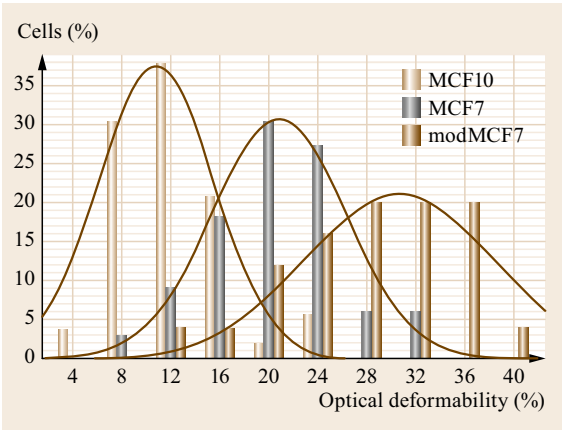


Fig. 20.16 Optical deformability of normal, cancerous, and metastatic breast epithelial cells. The three populations of the MCF cell lines are clearly distinguishable in the histograms of the measured optical deformability (after [20.61])

and the inner-dynamical transition frequency,

$$\omega_t = \frac{1}{2}\left(\frac{a_1}{a_2} - \frac{1}{b_1}\right) + \sqrt{\frac{1}{4}\left(\frac{a_1}{a_2} - \frac{1}{b_1}\right) - \frac{a_1}{a_2 b_1}}\,. \tag{20.28}$$

High-throughput biomechanical measurements became increasingly important when the mechanical properties of cells were found to be a very sensitive cell marker. A relatively low number of cells is sufficient to distinguish different populations without the use of molecular markers. In addition, subcellular structures such as the nucleus or the cell membrane can be targeted individually [20.59, 60].

The OS has been used, for instance, to identify differences in the biomechanical properties of normal and cancerous primary tissue as well as cell lines. Differences in optical deformability of benign fibroblasts (BALB/3T3) and malignantly transformed fibroblasts (SV-T2), for instance, clearly show that SV-T2 cells have a higher deformation and are therefore softer than the BALB/3T3 cells (Fig. 20.15) [20.50]. This finding corresponds to measurements of the cytoskeletons of these two cell lines, showing a reduced actin cortex for the cancerous SV-T2 cells, which accounts for the softer behavior.

Regarding human cancer types, breast cancer is the most common form of this disease in women. Using cell models, measurements with the OS have shown that breast cancer cells in different states of aggressiveness can be distinguished by their optical deformability, as illustrated in Fig. 20.16. More importantly, this trend is persistent for primary samples extracted from breast cancer patients, which give rise to a wide range of medical applications, including diagnosis and drug screening.

Since mechanical properties are well-defined quantities, this opens a wide field of applications. In the future, the OS might be used for tumor staging and might help to develop individual chemotherapies for cancer patients. A reasonable addition would be the monitoring of the cellular optical properties, since refractive index distributions can strongly vary between cell populations. This extra information is rather necessary in particular for experiments with primary cells due to the lack of literature values [20.62].

Thermorheology of Living Cells

Using laser light to trap and stretch biological matter also leads to a significant temperature increase within the sample, which nearly instantaneously influences the biomechanical responses of cells [20.63, 64]. Laser-induced heating from absorption is an inherent property of optical traps and optically driven cell deformations change about 10% per Kelvin. These changes are mainly caused by temperature-induced viscosity variations as known for polymeric solutions in general. Thus, the cellular, viscoelastic components, which determine the overall mechanical parameters, are highly sensitive to temperature changes.

More information on the temperature dependence of the cell's cytoskeletal components and its regulating processes are needed to be able to precisely attribute

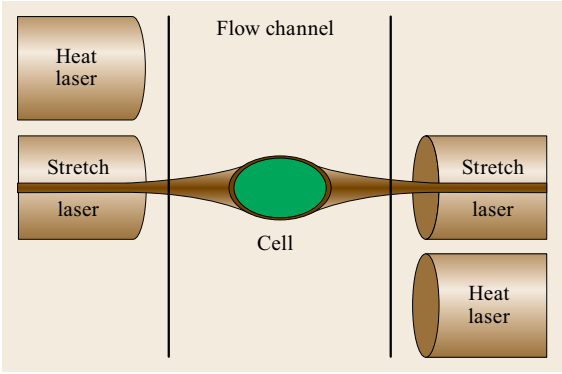


Fig. 20.17 The OS can be equipped with additional heating lasers, which allow the temperature of the medium around the cell to be tuned, thereby facilitating thermorheological measurements. Cells are delivered via a microfluidic system pumping cells in suspension through a capillary. Once the software recognizes a cell, the flow is stopped and the cell is subsequently trapped and stretched. The deformation is recorded via optical microscopy and afterwards the flow is switched on again to position the next cell to be measured (after [20.35])

experimental findings to specific effects. To achieve a decoupling of temperature-induced effects and purely rheological properties, the OS setup can be extended with two additional heating lasers (Fig. 20.17). It is very reasonable to use the OS instead of single-beam optical traps since its nonfocused beams drastically decrease local heating compared to strongly focused laser beams (about a factor of 400). By tuning the power of the heating lasers, the medium around the cell can be precisely adjusted before measuring. Randomizing the heating pattern for the measurement of an entire cell population yields unbiased, temperature-dependent deformation curves, which allow a distinction between temperature and rheological effects.

First experiments illustrated that time-temperature superposition (TTS) principles are applicable to creep compliance curves to create a master curve for MCF-10A cells [20.65]. In the consideration that cells are highly heterogeneous structures based on different types of polymers, it is rather surprising that deformation curves can be matched by this fundamental model. However, other cell lines have proven to have more complex properties and cannot be described by simple thermorheological behavior [20.66]. These effects can arise, for instance, due to a thermally activated process or active cellular responses and more complex thermorheological descriptions have to be applied.

In general, simple as well as complex thermorheology are new approaches for contact-free cell rheology. Differences in the applicability of the TTS give rise to new insights of biomechanical measurements without any assumptions for underlying models. Results for various cell lines have proven that the creep deformation of suspended cells strongly depends on temperature and that there are qualitative differences between cell types. Potentially, this concept can even be extended to determine whether effects are based on passive, stochastic processes or active, directed mechanisms since they behave in a contrary manner upon temperature changes.

Dielectrophoretic Cell Stretching

Optical stretching is an ideal technique to induce whole-cell deformations at a relatively high throughput without the need for extensive preparation for each cell. However, some experiments might require stable temperatures, which cannot be achieved with the OS since the stretch lasers will always lead to a temperature increase during the measurement.

A potential workaround is provided by a technique that uses forces induced in a dielectrophoretic field to stretch whole cells and also allows a characterization of their mechanical properties by stress-strain relations. Here, stretching is based on the fact that cells are dielectric particles that are polarized in an electric field established by an alternating current. Within this field, cells develop a dipole moment and interact with the inhomogeneous electric field. Depending on the external field and the induced dipoles, cells can either experience translational motion (dielectrophoresis), rotation, or stretch deformation [20.67]. The amplitude of the stretching depends on the size of the cell and its dielectric properties in regard to the medium. By carefully choosing the same experimental conditions, the induced deformations can be used to determine and compare mechanical properties of different cell types. However, since deformations are taking place on different time scales than in the OS, results can be hardly compared and potentially even display discordant trends [20.68].

20.3.2 Cell Deformation via Optical Tweezers

Optical tweezers are used for an increasing number of biological experiments and are now an appropriate technique for cell manipulation and characterization. The most common applications cover the deformation of whole cells in conjunction with force measurements as well as the manipulation of subcellular components, probing of the cell interior via microrheology, or the positioning of cells in combination with spectroscopy techniques in order to follow processes within single cells.

To study the mechanical behavior of a cell using optical tweezers, beads can be attached to the cell membrane diametrically on each side. If one of the beads is

attached to the substrate and the second is held by optical tweezers, constant force or constant-deformation measurements can be realized. *Suresh* et al. used such a setup to measure the shear modulus of red blood cells (RBCs) during the development of the malaria parasite *Plasmodium falciparum* and discovered a tenfold increase in the rigidity of infected RBCs compared with noninfected cells [20.69, 70]. However, one can also optically trap both attached beads and use them to quantify forces generated by cells. With this geometry, both the mechanical properties of the cells as well as their force fluctuations can be investigated in parallel. Additionally, the influences of cytoskeletal drugs such as blebbistatin (inhibiting myosin activity) or nocodazole (depolymerizes microtubules) on the cell stiffness can be readily tested and evaluated [20.71].

The use of optical tweezers is not limited to studying cells as a whole, but also allows for the manipulation of subcellular components. Kress et al. attached beads to filopodia, thin finger-like protrusions from the cell body, of macrophages. Amongst others, filopodia are implicated in a process called endocytosis, where they attach to objects and retract them to the cell membrane to be engulfed and taken up into the cell. Upon measuring step sizes and forces of filopodial retraction, it could be inferred that molecular motors are involved in this process [20.72].

However, even without beads being attached to the cell membrane, optical tweezers can be used to measure the cortical tension by monitoring the position of the membrane as well as recording the dynamics of bleb growth [20.73]. By automating this technique, optical tweezers can even be used as scanning devices monitoring, for instance, the fluctuation of membranes. This allows the employment of active and passive microrheology to directly compare membrane responses and thermal fluctuations on a single cell level [20.74].

By increasing the laser power to relatively high values compared to trapping conditions, the shape of cells can even be manipulated and processes such as protrusion can be guided, as shown for neuronal cells. More precisely, the intracellular processes are affected by highly focused laser light on the leading edge of a growth cone, which results in growth of the nerve cell towards the laser focus. The induced growth has been verified experimentally for wavelengths in the range of 800–900 nm [20.75].

Another method dubbed laser tracking microrheology (LTM) makes use of the QPD position detection system of an optical tweezers setup. In the passive mode, the Brownian motion of a tracer particle, for instance, a micrometer-sized bead or an intracellular granule, is monitored with subnanometer resolution and sampling rates in the range of 100 kHz at low laser powers [20.76, 77]. At higher powers, tracer particles can be trapped and moved such that out-of-equilibrium viscoelastic properties can be inferred.

Using optical tweezers to hold individual cells in place and combining the measurement with Raman spectroscopy employing the same laser facilitates noninvasive characterization. This technique yields highly reproducible biomolecular fingerprints representing DNA, protein and lipid content along with conformational information [20.78] and can be used, for instance, to distinguish neoplastic from normal lymphocytes [20.79]. Laser (or optical) tweezers Raman spectroscopy (LTRS/OTRS) even allows for marker-free kinetic studies of biochemical processes inside a single cell. Singh et al. monitored the response of single yeast cells to heat treatment and hyperosmotic stress caused by high glucose concentrations. The examined yeast, *Saccharomyces cerevisiae*, then switches to fermentation mode, and the products of which (glycerol and ethanol) could be observed in real time [20.80].

These examples underline the wide use of optical tweezers to characterize individual cells or even subcellular components and processes. Both mechanical properties as well as chemical composition can be measured and monitored in a contact-free, noninvasive way in order to study the effect of drugs on the intracellular processes of living cells, which also allows for the separation of cells based on these characteristics.

The possibility to hold and manipulate several objects at the same time opens the door for a large number of new applications such as the study of the collective behavior of trapped objects. However, time-sharing or splitting and recombining the laser beam poses limits on the number and power of trap centers.

In 1997 Grier and Dufresne pursued the idea of creating multiple traps using phase holograms and patented their invention as holographic optical tweezers (HOT). With holographic optical tweezers, a novel device had been developed, allowing for the creation of an array of optical tweezers for synchronous trapping of particles in arbitrary geometric patterns [20.81]. The basis of holographic optical tweezers are holograms that modify the collimated laser beam. In general, the hologram would affect both the amplitude and the phase. However, since changing the amplitude would diminish the power available for trapping, algorithms have been developed that produce an equivalent phase-shift pattern, a so-called kinoform. When being focused through the objective, the pattern is Fourier-transformed into a pattern of multiple traps.

The first HOTs were realized using microstructured diffractive optical elements. However, because of their lack of flexibility, they have been succeeded in research applications by computer-addressed spatial light

modulators. These are liquid-crystal devices where the orientation of the molecules can be locally controlled, varying the optical path length and resulting in the desired phase shift [20.82–84]. The technique of holographic optical tweezers is a relatively new technology that opens the door to a number of new applications by parallelization of particle trapping. Whereas a single-beam trap is only able to hold a single or very few cells, the application of HOTs allows for the positioning and organization of many cells into a desired pattern.

In this way, it is possible to create a stable three-dimensional (3-D) configuration of cells where each object is located at a predefined position. *Jordan* et al. placed *E. coli* in a gelatin matrix, switched off the lasers after the positioning of the objects and provided appropriate nutrients such that single cells were able to survive for several days [20.85, 86]. This demonstration of cellular positioning using multiple beams included observation of the effects of both the position and number of neighboring cells upon one another and points to the wide range of future biological applications. These include the possibility of organizing specific cell lines such as stem cells into a novel or complex geometry, e.g., so-called stem cell niches. These geometric arrangements of cells may facilitate the study of the collective behavior of many-body structures and act as motifs for promoting tissue differentiation.

While trapping cells in suspension with conventional optical tweezers might prove difficult for cells much larger than the focal spot, HOTs offer the possibility to create a more stable trap configuration by using multiple beams around the periphery of the object. By applying Raman spectroscopy the samples can be scanned back and forth across the focus of the stationary excitation laser beam, resulting in space-resolved Raman spectra of an entire cell [20.87].

20.4 Hydrodynamic Shear Forces for the Manipulation of Single Cells

Optical forces are a versatile tool to investigate cellular properties on a single cell level. They allow a quantification of biological processes by scanning parts of the cell or even facilitate the induction of whole-cell deformations to record the stress-strain relation thereby determining a cell's mechanical properties. However, optical forces also inherently induce a temperature increase in the sample, which has to be accounted for in the analysis. Furthermore, these measurements are too slow for a wide range of applications in biology or medicine and cannot compete with throughput rates of standard flow cytometers. Rates of 250 cells per hour – as achieved with an automated OS setup – are sufficient to probe concepts such as marker-free, mechanical phenotyping of cell populations [20.88]. A major advantage of this approach is that cells do not have to be treated before being measured, but to transfer this concept into a clinical frame, higher throughput rates are needed to gain sufficient statistics in a reasonable time frame.

To achieve these clinically relevant rates, new techniques were developed inducing whole-cell deformation by shear flows facilitating rates up to 2000 cells per second [20.89] with a real-time analysis possible at slower rates of up to 500 cells per second [20.90]. This throughput is sufficient for flow cytometry, which is independent of additional markers. In contrast to experiments based on optical forces, only one deformation mode is tested neglecting possible time and frequency dependencies of the cellular mechanical properties. Thus, these techniques nicely complement each other facilitating a further advancement of the concept of marker-free, mechanical phenotyping of cells.

20.4.1 Flow-Induced Shear Deformation of Cells

Within microfluidic systems, cells can be carried by the surrounding fluid. Under laminar flows they migrate in a predictable manner across streamlines, which can ultimately lead to an ordering effect in longitudinal direction driving the cells to the axial center of the channel [20.89]. This ordering and initial focusing can be used to deliver suspended cells uniformly to a chamber where they are deformed in contact-free manner by an additional, opposing flow (Fig. 20.18a). Resulting high strains induce large whole-cell deformations allowing a derivation of stress-strain relations reflecting the cellular mechanical properties. The uniform delivery is crucial to preserve the uniformity in hydrodynamic stresses, which enables a comparison of the deformed cells. Furthermore, cells never come into contact with the channel itself, thereby bypassing any adherence or clogging effects. Furthermore, this principle has the advantage that measurements are independent of heterogeneities of subcellular structures since the whole cell is deformed. These heterogeneities can cause severe reproducibility issues for methods relying on physical contact such as AFM.

After a cell has entered the chamber where the opposing flows meet, its deformation is recorded via an optical microscope equipped with a high-speed camera. Depending on the camera, throughputs up to 2000 cells per second can be achieved, which allows the user to acquire statistically significant data. Subsequently, automated image analysis routines can be employed

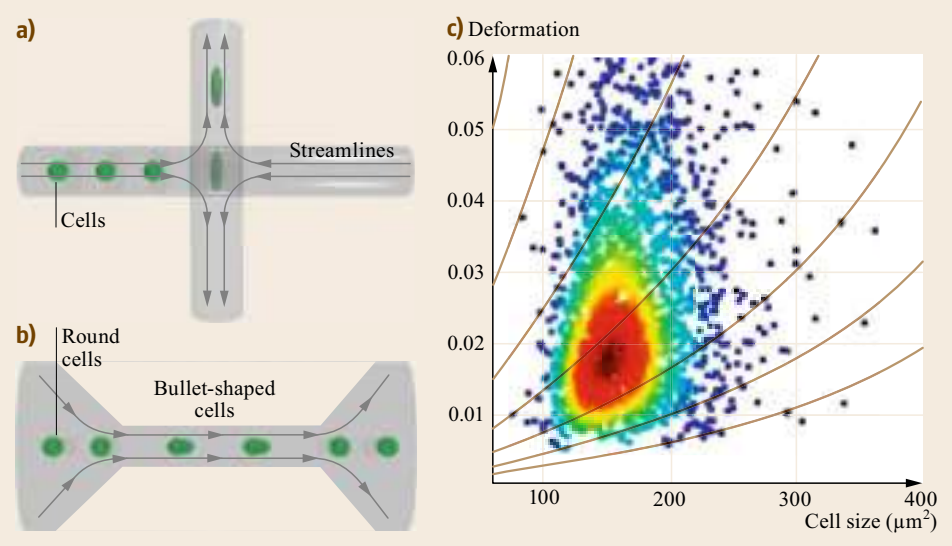


Fig. 20.18a–c Shear forces induced by flows can cause cell deformations detectable via standard microscopy techniques at high throughput rates. **(a)** A flow transports cells through a channel and aligns them in the middle region. At the intersection, an opposing flow leads to a detectable deformation attributable to the cells' mechanical properties. Stretched cells are subsequently transported through the release channels (after [20.89]). **(b)** The RT-DC setup employs geometrical constrictions in the flow channel leading to substantial shear forces. Cells are deformed into a bullet-shape profile, which can be analyzed and attributed to cell deformability (after [20.90]). **(c)** Flow-induced cell deformations can be recorded in the form of scatter plots similar to data illustrations of other standard types of flow cytometers

to identify the cell boundaries to quantify the according deformability, usually presented in the form of scatter plots similar to those also used for standard flow cytometers (representative example shown in Fig. 20.18c). Since measurements can be completely automated, user bias is minimized and experiments are quickly repeatable. This illustrates the translational possibility of mechanical phenotyping for diagnostic tools.

Microfluidic approaches, in general, allow a tuning of the magnitude of the induced deformation by altering the hydrodynamic stresses. This variability is important to probe cells under different conditions in order to investigate, for instance, time-dependent deformabilities necessary to gain a complete rheological profile. For slow deformations, optical stretching can be used to complement the rather fast hydrodynamic measurement techniques.

20.4.2 High-Throughput Mechanical Phenotyping of Cell Populations in Real Time

The previously described methods have one major drawback, which still impedes their clinical use. Although data acquisition can be relatively fast due to high throughput rates, their evaluation has to be performed off-line due to the enormous amount of data, which even limits the total number of analyzed cells.

A new technique called real-time deformability cytometry (RT-DC) resolves this issue by analyzing the data on-the-fly [20.90]. Within this setup, high shear stresses are induced by constricting the channel geometry (Fig. 20.18b). Arising pressure gradients are sufficient to deform even relatively stiff cells into a bullet-like shape, which facilitates the extraction of Young's modulus while also considering different initial cell sizes [20.91]. The underlying analytical model decouples cell deformation and size, which is infeasible for methods based on inertial forces.

Data is acquired by a high-speed camera synchronized with a light source for illumination and directly analyzed by an image-processing algorithm, which detects and determines the contour of a cell in real time [20.90]. Using these values, the cell deformation and size can be quantified and displayed in form of a scatter plot (representative example shown in Fig. 20.18c). This method allows a continuous analysis of cells at rather high rates ($>$ 1000 cells/s) and enables measurements of more than 100 000 cells per run.

RT-DC readily enables the measurement of, for instance, the influence of cytoskeletal drugs [20.92] or the progression in the cell cycle [20.90]. Furthermore, even cell mixtures such as blood can be analyzed showing

a typical fingerprint when decomposing the different cell populations. This fingerprint has been reported to change upon infections, which can be potentially used as a marker-free diagnostic tool for diseases [20.90]. Essentially, all bodily fluids are easily experimentally accessible since cells are already in the suspended state and do not require extensive preparations. The high throughput rates potentially even allow the user to sort the different cell populations in sufficient quantities after the deformation measurement.

20.5 Conclusion and Outlook

Methods for contact-free measurements of biological materials have greatly advanced during the last decade and now enable the exploration of the biological microcosmos with high precision. Manipulations and quantifications have been reduced down to the molecular level and can readily identify nanoscopic details of biopolymeric structures. However, not only has the high resolution led to significant new insights, but modern techniques have also attained high accuracy in measuring whole-cell deformability to determine according mechanical properties. These recent advances have promoted the concept of using cell mechanics as a potential marker-free method for mechanical phenotyping. Pathological as well as physiological changes are reflected in an alteration of cell mechanics and potentially allow the application of this principle as a diagnostic tool, a concept which has been already initially proven for blood analysis [20.90, 93], cancer diagnostics [20.61, 94, 95], and the tracking of stem cell differentiation [20.93]. Because of the increasing throughput, some methods are rather close to translation to clinical frames, and furthermore have the potential to drastically reduce animal studies by shifting the focus towards cell-based assays as a first approach.

References

20.1 M.J. Rust, M. Bates, X. Zhuang: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), Nat. Methods **3**, 793–796 (2006)

20.2 B. Huang, W. Wang, M. Bates, X. Zhuang: Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy, Science **319**, 810–813 (2008)

20.3 S.W. Hell, J. Wichmann: Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy, Opt. Lett. **19**, 780–782 (1994)

20.4 S.W. Hell: Toward fluorescence nanoscopy, Nat. Biotechnol. **21**, 1347–1355 (2003)

20.5 A. Ashkin: Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime, Biophys. J. **61**, 569–582 (1992)

20.6 A. Ashkin, J. Dziedzic: Optical trapping and manipulation of viruses and bacteria, Science **235**, 1517–1520 (1987)

20.7 M.M. Burns, J.-M. Fournier, J.A. Golovchenko: Optical binding, Phys. Rev. Lett. **63**, 1233–1236 (1989)

20.8 A. Ashkin: Acceleration and trapping of particles by radiation pressure, Phys. Rev. Lett. **24**, 156–159 (1970)

20.9 K.T. Gahagan, G.A. Swartzlander: Optical vortex trapping of particles, Opt. Lett. **21**, 827–829 (1996)

20.10 Y. Kimura, P.R. Bianco: Single molecule studies of DNA binding proteins using optical tweezers, Analyst **131**, 868–874 (2006)

20.11 A. Sasso, G. Pesce: Optical tweezers calibration: A quantitative tool for local viscosity investigation, Proc. SPIE **5830**, 368 (2005)

20.12 S.F. Tolić-Nørrelykke, E. Schäffer, J. Howard, F.S. Pavone, F. Jülicher, H. Flyvbjerg: Calibration of optical tweezers with positional detection in the back focal plane, Rev. Sci. Instrum. **77**, 103101 (2006)

20.13 P. Debye: Der Lichtdruck auf Kugeln von beliebigem Material, Ann. Phys. **335**, 57–136 (1909)

20.14 A. Ashkin, J.M. Dziedzic, J.E. Bjorkholm, S. Chu: Observation of a single-beam gradient force optical trap for dielectric particles, Opt. Lett. **11**, 288–290 (1986)

20.15 K. Svoboda, S.M. Block: Biological applications of optical forces, Annu. Rev. Biophys. Biomol. Struct. **23**, 247–285 (1994)

20.16 K.C. Neuman, S.M. Block: Optical trapping, Rev. Sci. Instrum. **75**, 2787–2809 (2004)

20.17 D. Strehle, J. Schnauß, C. Heussinger, J. Alvarado, M. Bathe, J. Käs, B. Gentry: Transiently crosslinked F-actin bundles, Eur. Biophys. J. **40**, 93–101 (2011)

20.18 J. Schnauß, T. Golde, C. Schuldt, B.U.S. Schmidt, M. Glaser, D. Strehle, T. Händler, C. Heussinger, J.A. Käs: Transition from a linear to a harmonic potential in collective dynamics of a multifilament actin bundle, Phys. Rev. Lett. **116**, 108102 (2016)

20.19 F. Gittes, C.F. Schmidt: Interference model for back-focal-plane displacement detection in optical tweezers, Opt. Lett. **23**, 7–9 (1998)

20.20 A. Ward, F. Hilitski, W. Schwenger, D. Welch, A.W.C. Lau, V. Vitelli, L. Mahadevan, Z. Dogic: Solid friction between soft filaments, Nat. Mater. **14**, 583–588 (2015)

20.21 F. Hilitski, A.R. Ward, L. Cajamarca, M.F. Hagan, G.M. Grason, Z. Dogic: Measuring cohesion between macromolecular filaments one pair at

a time: Depletion-induced microtubule bundling, Phys. Rev. Lett. **114**, 138102 (2015)
20.22 Z. Lansky, M. Braun, A. Lüdecke, M. Schlierf, P.R. ten Wolde, M.E. Janson, S. Diez: Diffusible crosslinkers generate directed forces in microtubule networks, Cell **160**, 1159–1168 (2015)
20.23 M. Braun, A.P. Bregulla, K. Günther, M. Mertig, F. Cichos: Single molecules trapped by dynamic inhomogeneous temperature fields, Nano Lett. **15**, 5499–5505 (2015)
20.24 M. Braun, A. Würger, F. Cichos: Trapping of single nano-objects in dynamic temperature fields, Phys. Chem. Chem. Phys. **16**, 15207 (2014)
20.25 M. Braun, F. Cichos: Optically controlled thermophoretic trapping of single nano-objects, ACS Nano **7**, 11200–11208 (2013)
20.26 D. Braun, A. Libchaber: Trapping of DNA by thermophoretic depletion and convection, Phys. Rev. Lett. **89**, 188103 (2002)
20.27 M. Tanase, N. Biais, M. Sheetz: Magnetic tweezers in cell biology. In: *Cell Mechanics*, ed. by Y.-L. Wang, D.E. Discher (Elsevier, Amsterdam, Boston 2007) pp. 473–493
20.28 I.D. Vilfan, J. Lipfert, D.A. Koster, S.G. Lemay, N.H. Dekker: Magnetic tweezers for single-molecule experiments. In: *Handbook of Single-Molecule Biophysics*, ed. by P. Hinterdorfer, A. van Oijen (Springer, New York 2009) pp. 371–369
20.29 J. Schnauß, M. Glaser, C. Schuldt, T. Golde, T. Händler, B.U.S. Schmidt, S. Diez, J.A. Käs: Motor-free force generation in biological systems, diffusion-fundamentals.org **23**, 1–15 (2015)
20.30 M. Braun, Z. Lansky, F. Hilitski, Z. Dogic, S. Diez: Entropic forces drive contraction of cytoskeletal networks, BioEssays **38**, 474–481 (2016)
20.31 C. Wagner, C. Olbrich, H. Brutzer, M. Salomo, U. Kleinekathöfer, U.F. Keyser, F. Kremer: DNA condensation by TmHU studied by optical tweezers, AFM and molecular dynamics simulations, J. Biol. Phys. **37**, 117–131 (2011)
20.32 S.M. Block, L.S.B. Goldstein, B.J. Schnapp: Bead movement by single kinesin molecules studied with optical tweezers, Nature **348**, 348–352 (1990)
20.33 K. Svoboda, S.M. Block: Force and velocity measured for single kinesin molecules, Cell **77**, 773–784 (1994)
20.34 R.M. Simmons, J.T. Finer, H.M. Warrick, B. Kralik, S. Chu, J.A. Spudich: Force on single actin filaments in a motility assay measured with an optical trap. In: *Mechanism of Myofilament Sliding in Muscle Contraction*, ed. by H. Sugi, G.H. Pollack (Springer, New York 1993) pp. 331–337
20.35 F. Huber, J. Schnauß, S. Rönicke, P. Rauch, K. Müller, C. Fütterer, J. Käs: Emergent complexity of the cytoskeleton: From single filaments to tissue, Adv. Phys. **62**, 1–112 (2013)
20.36 R. Seidel, J. van Noort, C. van der Scheer, J.G.P. Bloom, N.H. Dekker, C.F. Dutta, A. Blundell, T. Robinson, K. Firman, C. Dekker: Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I, Nat. Struct. Mol. Biol. **11**, 838–843 (2004)
20.37 R. Seidel, J.G.P. Bloom, C. Dekker, M.D. Szczelkun: Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I, EMBO J. **27**, 1388–1398 (2008)
20.38 H. Itoh, A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida, K. Kinosita: Mechanically driven ATP synthesis by F1-ATPase, Nature **427**, 465–468 (2004)
20.39 Y. Rondelez, G. Tresset, T. Nakashima, Y. Kato-Yamada, H. Fujita, S. Takeuchi, H. Noji: Highly coupled ATP synthesis by F1-ATPase single molecules, Nature **433**, 773–777 (2005)
20.40 T. Stangner, C. Wagner, D. Singer, S. Angioletti-Uberti, C. Gutsche, J. Dzubiella, R. Hoffmann, F. Kremer: Determining the specificity of monoclonal antibody HPT-101 to tau-peptides with optical tweezers, ACS Nano **7**, 11388–11396 (2013)
20.41 T. Stangner, S. Angioletti-Uberti, D. Knappe, D. Singer, C. Wagner, R. Hoffmann, F. Kremer: Epitope mapping of monoclonal antibody HPT-101: A study combining dynamic force spectroscopy, ELISA and molecular dynamics simulations, Phys. Biol. **12**, 66018 (2015)
20.42 D.J. Kauert, T. Kurth, T. Liedl, R. Seidel: Direct mechanical measurements reveal the material properties of three-dimensional DNA origami, Nano Lett. **11**, 5558–5563 (2011)
20.43 P.W.K. Rothemund: Folding DNA to create nanoscale shapes and patterns, Nature **440**, 297–302 (2006)
20.44 P. Yin, R.F. Hariadi, S. Sahu, H.M.T. Choi, S.H. Park, T.H. LaBean, J.H. Reif: Programming DNA tube circumferences, Science **321**, 824–826 (2008)
20.45 C. Schuldt, J. Schnauß, T. Händler, M. Glaser, J. Lorenz, T. Golde, J.A. Käs, D.M. Smith: Tuning synthetic semiflexible networks by bending stiffness, Phys. Rev. Lett. **117**, 197801 (2016)
20.46 D. Smith, V. Schüller, C. Engst, J. Rädler, T. Liedl: Nucleic acid nanostructures for biomedical applications, Nanomedicine **8**, 105–121 (2013)
20.47 D. Schiffels, T. Liedl, D.K. Fygenson: Nanoscale structure and microscale stiffness of DNA nanotubes, ACS Nano **7**, 6700–6710 (2013)
20.48 M. Glaser, J. Schnauß, T. Tschirner, B.U.S. Schmidt, M. Moebius-Winkler, J.A. Käs, D.M. Smith: Self-assembly of hierarchically ordered structures in DNA nanotube systems, New J. Phys. **18**, 55001 (2016)
20.49 E. Pfitzner, C. Wachauf, F. Kilchherr, B. Pelz, W.M. Shih, M. Rief, H. Dietz: Rigid DNA beams for high-resolution single-molecule mechanics, Angew. Chem. Int. Ed. **52**, 7766–7771 (2013)
20.50 D. Koirala, P. Shrestha, T. Emura, K. Hidaka, S. Mandal, M. Endo, H. Sugiyama, H. Mao: Single-molecule mechanochemical sensing using DNA origami nanostructures, Angew. Chem. **126**, 8275–8279 (2014)
20.51 J. Guck, R. Ananthakrishnan, T.J. Moon, C.C. Cunningham, J. Käs: Optical deformability of soft biological dielectrics, Phys. Rev. Lett. **84**, 5451–5454 (2000)
20.52 J. Guck, R. Ananthakrishnan, H. Mahmood, T.J. Moon, C.C. Cunningham, J. Käs: The optical stretcher: A novel laser tool to micromanipulate

cells, Biophys. J. **81**, 767–784 (2001)
20.53 P.B. Bareil, Y. Sheng, A. Chiou: Local scattering stress distribution on surface of a spherical cell in optical stretcher, Opt. Express **14**, 12503–12509 (2006)
20.54 A.E. Ekpenyong, C.L. Posey, J.L. Chaput, A.K. Burkart, M.M. Marquardt, T.J. Smith, M.G. Nichols: Determination of cell elasticity through hybrid ray optics and continuum mechanics modeling of cell deformation in the optical stretcher, Appl. Opt. **48**, 6344 (2009)
20.55 L. Boyde, K.J. Chalut, J. Guck: Interaction of Gaussian beam with near-spherical particle: An analytic-numerical approach for assessing scattering and stresses, J. Opt. Soc. Am. A **26**, 1814 (2009)
20.56 F. Wottawah, S. Schinkinger, B. Lincoln, S. Ebert, K. Müller, F. Sauer, K. Travis, J. Guck: Characterizing single suspended cells by optorheology, Acta Biomater. **1**, 263–271 (2005)
20.57 R. Ananthakrishnan, J. Guck, F. Wottawah, S. Schinkinger, B. Lincoln, M. Romeyke, J. Kas: Modelling the structural response of an eukaryotic cell in the optical stretcher, Curr. Sci. **88**, 1434–1440 (2005)
20.58 R.E. Mahaffy, C.K. Shih, F.C. MacKintosh, J. Käs: Scanning probe-based frequency-dependent microrheology of polymer gels and biological cells, Phys. Rev. Lett. **85**, 880–883 (2000)
20.59 U. Delabre, K. Feld, E. Crespo, G. Whyte, C. Sykes, U. Seifert, J. Guck: Deformation of phospholipid vesicles in an optical stretcher, Soft Matter **11**, 6075–6088 (2015)
20.60 K.J. Chalut, M. Höpfler, F. Lautenschläger, L. Boyde, C.J. Chan, A. Ekpenyong, A. Martinez-Arias, J. Guck: Chromatin decondensation and nuclear softening accompany nanog downregulation in embryonic stem cells, Biophys. J. **103**, 2060–2070 (2012)
20.61 J. Guck, S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H.M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Käs, S. Ulvick, C. Bilby: Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence, Biophys. J. **88**, 3689–3698 (2005)
20.62 S. Grosser, A.W. Fritsch, T.R. Kießling, R. Stange, J.A. Käs: The lensing effect of trapped particles in a dual-beam optical trap, Opt. Express **23**, 5221 (2015)
20.63 S. Ebert, K. Travis, B. Lincoln, J. Guck: Fluorescence ratio thermometry in a microfluidic dual-beam laser trap, Opt. Express **15**, 15493 (2007)
20.64 F. Wetzel, S. Rönicke, K. Müller, M. Gyger, D. Rose, M. Zink, J. Käs: Single cell viability and impact of heating by laser absorption, Eur. Biophys. J. **40**, 1109–1114 (2011)
20.65 T.R. Kießling, R. Stange, J.A. Käs, A.W. Fritsch: Thermorheology of living cells – Impact of temperature variations on cell mechanics, New J. Phys. **15**, 45026 (2013)
20.66 B.U.S. Schmidt, T.R. Kießling, E. Warmt, A.W. Fritsch, R. Stange, J.A. Käs: Complex thermorheology of living cells, New J. Phys. **17**, 73010 (2015)
20.67 V.L. Sukhorukov, H. Mussauer, U. Zimmermann: The effect of electrical deformation forces on the electropermeabilization of erythrocyte membranes in low- and high-conductivity media, J. Membr. Biol. **163**, 235–245 (1998)
20.68 I. Guido, M.S. Jaeger, C. Duschl: Dielectrophoretic stretching of cells allows for characterization of their mechanical properties, Eur. Biophys. J. **40**, 281–288 (2011)
20.69 S. Suresh, J. Spatz, J.P. Mills, A. Micoulet, M. Dao, C.T. Lim, M. Beil, T. Seufferlein: Connections between single-cell biomechanics and human disease states: Gastrointestinal cancer and malaria, Acta Biomater. **1**, 15–30 (2005)
20.70 M. Dao, C.T. Lim, S. Suresh: Mechanics of the human red blood cell deformed by optical tweezers, J. Mech. Phys. Solids **51**, 2259–2280 (2003)
20.71 F. Schlosser, F. Rehfeldt, C.F. Schmidt: Force fluctuations in three-dimensional suspended fibroblasts, Philos. Trans. R. Soc. B **370**, 20140028 (2014)
20.72 H. Kress, E.H.K. Stelzer, D. Holzer, F. Buss, G. Griffiths, A. Rohrbach: Filopodia act as phagocytic tentacles and pull with discrete steps and a load-dependent velocity, Proc. Natl. Acad. Sci. **104**, 11633–11638 (2007)
20.73 J. Peukes, T. Betz: Direct measurement of the cortical tension during the growth of membrane blebs, Biophys. J. **107**, 1810–1820 (2014)
20.74 H. Turlier, D.A. Fedosov, B. Audoly, T. Auth, N.S. Gov, C. Sykes, J.-F. Joanny, G. Gompper, T. Betz: Equilibrium physics breakdown reveals the active nature of red blood cell flickering, Nat. Phys. **12**, 513–519 (2016)
20.75 A. Ehrlicher, T. Betz, B. Stuhrmann, D. Koch, V. Milner, M.G. Raizen, J. Käs: Guiding neuronal growth with light, Proc. Natl. Acad. Sci. **99**, 16024–16028 (2002)
20.76 S. Yamada, D. Wirtz, S.C. Kuo: Mechanics of living cells measured by laser tracking microrheology, Biophys. J. **78**, 1736–1747 (2000)
20.77 B.D. Hoffman, G. Massiera, K.M. van Citters, J.C. Crocker: The consensus mechanics of cultured mammalian cells, Proc. Natl. Acad. Sci. **103**, 10259–10264 (2006)
20.78 C. Xie, M.A. Dinno, Y.-Q. Li: Near-infrared Raman spectroscopy of single optically trapped biological cells, Opt. Lett. **27**, 249–251 (2002)
20.79 J.W. Chan, D.S. Taylor, T. Zwerdling, S.M. Lane, K. Ihara, T. Huser: Micro-raman spectroscopy detects individual neoplastic and normal hematopoietic cells, Biophys. J. **90**, 648–656 (2006)
20.80 G.P. Singh, C.M. Creely, G. Volpe, H. Grötsch, D. Petrov: Real-time detection of hyperosmotic stress response in optically trapped single yeast cells using raman microspectroscopy, Anal. Chem. **77**, 2564–2568 (2005)
20.81 E.R. Dufresne, D.G. Grier: Optical tweezer arrays and optical substrates created with diffractive optics, Rev. Sci. Instrum. **69**, 1974–1977 (1998)
20.82 D.G. Grier, Y. Roichman: Holographic optical trapping, Appl. Opt. **45**, 880–887 (2006)

20.83 D.G. Grier: A revolution in optical manipulation, Nature **424**, 810–816 (2003)
20.84 M. Reicherter, J. Liesener, T. Haist, H.J. Tiziani: Advantages of holographic optical tweezers, Proc. SPIE **5143**, 76–83 (2003)
20.85 P. Jordan, J. Leach, M. Padgett, P. Blackburn, N. Isaacs, M. Goksör, D. Hanstorp, A. Wright, J. Girkin, J. Cooper: Creating permanent 3D arrangements of isolated cells using holographic optical tweezers, Lab Chip **5**, 1224–1228 (2005)
20.86 J. Leach, G. Sinclair, P. Jordan, J. Courtial, M. Padgett, J. Cooper, Z. Laczik: 3D manipulation of particles into crystal structures using holographic optical tweezers, Opt. Express **12**, 220–226 (2004)
20.87 C. Creely, G. Volpe, G. Singh, M. Soler, D. Petrov: Raman imaging of floating cells, Opt. Express **13**, 6105–6110 (2005)
20.88 T.R. Kießling, M. Herrera, K.D. Nnetu, E.M. Balzer, M. Girvan, A.W. Fritsch, S.S. Martin, J.A. Käs, W. Losert: Analysis of multiple physical parameters for mechanical phenotyping of living cells, Eur. Biophys. J. **42**, 383–394 (2013)
20.89 D.R. Gossett, H.T.K. Tse, S.A. Lee, Y. Ying, A.G. Lindgren, O.O. Yang, J. Rao, A.T. Clark, D. Di Carlo: Hydrodynamic stretching of single cells for large population mechanical phenotyping, Proc. Natl. Acad. Sci. **109**, 7630–7635 (2012)
20.90 O. Otto, P. Rosendahl, A. Mietke, S. Golfier, C. Herold, D. Klaue, S. Girardo, S. Pagliara, A. Ekpenyong, A. Jacobi, M. Wobus, N. Töpfner, U.F. Keyser, J. Mansfeld, E. Fischer-Friedrich, J. Guck: Real-time deformability cytometry: On-the-fly cell mechanical phenotyping, Nat. Methods **12**, 199–202 (2015)
20.91 A. Mietke, O. Otto, S. Girardo, P. Rosendahl, A. Taubenberger, S. Golfier, E. Ulbricht, S. Aland, J. Guck, E. Fischer-Friedrich: Extracting cell stiffness from real-time deformability cytometry: Theory and experiment, Biophys. J. **109**, 2023–2036 (2015)
20.92 C.J. Chan, A.E. Ekpenyong, S. Golfier, W. Li, K.J. Chalut, O. Otto, J. Elgeti, J. Guck, F. Lautenschläger: Myosin II activity softens cells in suspension, Biophys. J. **108**, 1856–1869 (2015)
20.93 A.E. Ekpenyong, G. Whyte, K. Chalut, S. Pagliara, F. Lautenschläger, C. Fiddler, S. Paschke, U.F. Keyser, E.R. Chilvers, J. Guck, W. Lam: Viscoelastic properties of differentiating blood cells are fate- and function-dependent, PLoS ONE **7**, e45237 (2012)
20.94 J. Runge, T.E. Reichert, A. Fritsch, J. Käs, J. Bertolini, T.W. Remmerbach: Evaluation of single-cell biomechanics as potential marker for oral squamous cell carcinomas: A pilot study, Oral Dis. **20**, e120–e127 (2014)
20.95 A. Fritsch, M. Höckel, T. Kiessling, K.D. Nnetu, F. Wetzel, M. Zink, J.A. Käs: Are biomechanical changes necessary for tumour progression?, Nat. Phys. **6**, 730–732 (2010)

21. Nano-Particles for Biomedical Applications

Paolo Decuzzi, Alessandro Coclite, Aeju Lee, Anna Lisa Palange, Daniele Di Mascolo, Ciro Chiappini, Hélder A. Santos, Maria Laura Coluccio, Gerardo Perozziello, Patrizio Candeloro, Enzo Di Fabrizio, Francesco Gentile

Nanoparticles (NPs) are extremely small particulates with an average size that ranges from a micron or less to a few nanometers. The large majority of NPs necessitate nanotechnology methods for their production. The size of NPs may vary over a significant range, which underlies their scientific potential in that NPs may help cross the bridge between bulk materials and molecular structures. More importantly, NPs are (nano)tech products and thus, in contrast to natural systems, they can be designed and engineered. On directly interacting with cells, including the structures of cells, their machinery and their waste products, NPs represent an unprecedented tool for addressing specific biological problems. In this chapter, we will briefly review some recent advances in nanoparticle research for biomedical applications, ranging from mesoporous silicon particles to gold and silver nanoparticles and polymeric nanocarriers for therapeutic, diagnosis, or theranostic (therapeutics + diagnosis) applications. We will offer a description of how, at the current state of the art, similar nanomedicine platforms are realized.

21.1 **Overview** 644
21.2 **Rational Design of Nanoconstructs** 646
21.2.1 A Hierarchical Multiscale Approach for Vascular Transport of Nanoconstructs 647
21.3 **Multifunctional Polymeric Nanoconstructs** 650
21.3.1 Spherical Polymeric Nanoconstructs (SPNs) for Combination Therapy and Biomedical Imaging 650
21.3.2 Discoidal Polymeric Nanoconstructs (DPNs) for Maximizing Accumulation Within Tumor Vasculature 650
21.4 **Sensing and Drug Delivery with Porous Silicon Nanomaterials** 655
21.4.1 Fabrication 656
21.4.2 Surface Chemistry 657
21.4.3 Biointerface 658
21.4.4 Drug Delivery 659
21.4.5 Sensing 661
21.5 **Nanomedicine Synthesis by Microfluidics Technology** 662
21.5.1 Why Microfluidics for the Synthesis of Nanomedicines? 662
21.5.2 Microfluidic Production of Nanocarriers Using PDMS Microfluidic Devices 664
21.5.3 Microfluidic Production of Nanocarriers Using Glass Capillary Microfluidic Devices 665
21.5.4 Summary 668
21.6 **Electroless Formation of Metal Nanoparticle Aggregates** 669
21.6.1 DLA Methods Can Predict Electroless Formation of Metal Nanoparticles 673
21.7 **2-D and 3-D Optical Nanostructures** 677
21.7.1 Integration of Optical Nanostructures in Microfluidic Devices 677
21.7.2 Three-Dimensional Optical Nanostructure 679
21.7.3 Two-Dimensional Optical Nanostructure 679
References 681

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_21

21.1 Overview

Nanoparticles (NPs) are extremely small particulates with an average size that ranges from a micron or less to a few nanometers [21.1–5]. This is not an operational or functional definition and in the present interpretation the sole condition for an object to be considered a nanoparticle is its size. Notice that the radius of the first electron orbit in the hydrogen atom is $\approx 0.5\,\text{Å} = 0.05\,\text{nm}$ [21.6] and the atomic radius of unstrained silicon is $\approx 1\,\text{Å} = 0.1\,\text{nm}$ [21.6], thus smaller NPs in the low nanometer range comprise a finite and generally low number of atoms and are obtained using bottom-up techniques including supramolecular assembly techniques [21.7, 8], and chemical [21.9, 10] or electrochemical [21.11, 12] deposition techniques. Larger micrometric or submicrometric NPs are produced using top-down methods in which the intended structures are obtained from a bulk constituent that is shaped into its final form using patterning, lithography or imprinting techniques, or a combination of these [21.13], or by conveniently removing the excess of material using wet or dry etching methods or porosification techniques [21.14]. In some cases, bottom-up and top-down techniques may be used in cooperation to attain the maximum control over the size and shape of the NPs, and this is described below. Thus, despite not being a strict requisite, the vast majority of NPs necessitate nanotechnology/nanofabrication methods. There are a few exceptions, such as diatoms, which are unicellular organisms with a porous silica (SiO_2) skeleton that is regular in size with an average length scale of some microns, and a spherical, spheroidal or ellipsoidal shape, depending on the family. Diatoms resemble micro/nanovectors and in fact most of the diatoms are used as sponges or molecular filters [21.15, 16]. But, even in this case, nanotechnology methods are necessary to image and verify the structures and assess their efficacy.

The size of NPs may vary over a significant range, and in this resides their scientific potential in that NPs may help cross the bridge between bulk materials and atomic or molecular structures. More importantly, NPs are nanotechnology products, thus, in contrast to biological/natural systems, they can be designed and engineered. NPs are not really particles; they are more akin to engineering systems in which size, shape, material, physical and chemical properties, structure and level of hierarchy can be precisely tailored or tuned to obtain systems with an elevated technology content. The complexity of these systems may depend less on their individual constituents taken in isolation and more on the fact that many of these constituents cooperate in a complex system. Interactions among these elements at the nanoscale yield collective phenomena with improved abilities in contrast to isolated components of those systems at conventional continuum scales. Consider a very practical example. For spherical metal particles, the surface-to-volume ratio (SVR) scales with $3/r$, where r is the radius of the particles. Thus, for a 50 nm particle, $SVR = 0.06\,\text{nm}$, which means that one has $0.06\,\text{nm}^2$ of surface per nm^3 of volume. For a 50 mm particle, $SVR = 6\times10^{-5}\,\text{nm}$, with a thousand-fold decrease. Thus smaller nanometric particles have larger surface-to-volume ratios and larger surface energy reactivity. The increasing proportion of surface atoms with decreasing particle size compared with bulk metals makes small metal particles highly reactive catalysts, as surface atoms possess more energy than bulk atoms. This in turn translates into a decrease of melting point that for nanomaterials may be two to three times lower compared to macromaterials. Moreover, it is widely reported [21.17] and very well understood [21.18–20] that metal nanoparticles with an average particle size of 50 mm interact with an electromagnetic field to yield site-specific increments of that field, which in cascade yields an amplification of the Raman spectroscopy signal of molecules adsorbed to those particles, in an effect that is called surface-enhanced Raman spectroscopy. Thus metal nanoparticles are preferable to bulk metal materials in at least two aspects. In general, the scale-dependent properties of nanomaterials are:

1. Chemical properties, including reactivity and catalysis [21.21]
2. Thermal properties, i. e., melting temperature and heat amplification/attenuation [21.22, 23]
3. Mechanical properties, i. e., adhesion, capillary forces, and increased mechanical strength at the nanoscale [21.24, 25]
4. Optical properties, i. e., absorption and scattering of light at the nanoscale
5. Electrical properties [21.26, 27]
6. Magnetic properties and superparamagnetic effects [21.28].

Consider now the case of biology, biomedicine and life sciences. The aims and scopes of such disciplines are various and diverse in nature and range from understanding the chemical/biological mechanisms of life [21.29]; to deepening the knowledge of anatomy, pathology and physiology of the body system; to studying the human body and how it is affected by disease, including studying at cellular and molecular levels; to the development of systems that may prevent or alter the

course of a disease or treat it, either by improving the diagnosis efficiency of conventional analytical/imaging techniques or the methods of administration and the therapeutic efficacy and the methods and timing of administration of drugs in the systemic circulation. Very roughly, cells have a size of ten microns, biomarkers (proteins and molecules involved in cellular activity) or viruses have a size of few nanometers or less [21.30–32], and the circuits involved in the cardiovascular system, the pulmonary circulation and systemic circulation, have cross-sections that range from several microns to few nanometers [21.33]. Thus, most of the items in biology have a characteristic length scale that falls in the same dimensional range as nanotechnology, and much of nanotechnology can be applied to biology and medicine. On directly interacting with cells, including the structures of cells, their machinery and their waste products, NPs represent an unprecedented tool for addressing specific biological problems. Excellent properties associated with nanoparticles of controlled geometry have opened up exciting opportunities for new material designs and will potentially revolutionize current biology and medicine practice. Following this motivation, nanomedicine originated at the intersection between biomedicine, nanotechnology and nanoparticles.

Nanomedicine can be defined as the design of diagnostics and/or therapeutics on the nanoscale, which provides advantages due to the high degree of coincident transport and delivery of the active species with mediation of their navigation within the biological systems for the treatment, prevention and diagnosis of diseases [21.34]. The biological transport processes, anatomically and down to the cellular and subcellular levels, are affected by the physical attributes of the nanocarriers, including their size, shape, and flexibility, as well as their chemical characteristics, including for instance the incorporation of active ligands for recognition by and triggering of biological receptors. Therefore, it is of critical importance to utilize procedures that prepare nanostructures with high degrees of uniformity, and with control over their physical and chemical traits. Nanoparticles can be constructed from various materials (e.g., polymers, lipids, metals) and can host a wide range of active components, including chemotherapeutics, contrast agents, proteins and nucleic acids, for various biomedical applications. Incorporation of certain functionalities can modulate the responsiveness (assembly/disassembly) of the nanoparticles in biological environments under different pH, enzymatic, oxidative, and reductive conditions, etc., or in response to external stimuli, such as variation in temperature, irradiation with near-infrared (IR) or ultraviolet-visible (UV-Vis) light, activation with magnetic fields, and application of ultrasound vibrations, etc.

In the remaining part of this section, we will briefly review some recent advances rather than classical results in nanoparticle research for biomedical applications, spanning mesoporous silicon particles, gold and silver nanoparticles, magnetic rather than paramagnetic effects, and polymeric nanocarriers for therapeutic, diagnosis, or theranostic (therapeutics + diagnosis) applications. This will provide the background necessary to understand and locate the paragraphs that form the present chapter.

Although intravenous injection is the main route of administration for nanoparticles, there are other ways to deliver them via less invasive ways, such as dermal/transdermal, oral and mucosal delivery. In all cases, these nanoparticles have to be equipped with smart components to allow their delivery beyond the different biological barriers, such as skin, mucus, blood, and extracellular matrix, in addition to the cellular and subcellular barriers. In the activity led by *Mauro Ferrari* [21.35, 36], mesoporous silicon particles were used for the delivery of different types of nanoparticles into primary endothelial cells, where the avoidance of biological barriers is left to a multistage strategy, whereby larger mesoporous particles are transported with a certain efficiency in close proximity to the endothelium, and a second generation of smaller nanoparticles is released from the pores to traverse easily, due to their small size, the endothelium barrier with their therapeutic payload. In [21.37], the same team used biomimetic leukolike vectors, i. e., nanoporous silicon particles coated with cellular membranes purified from leukocytes, to deliver drugs in the systemic circulation with no or minimum clearance by the immune system.

In the described applications, Ferrari et al. used mesoporous or nanoporous silicon as the main constituent for their carriers, principally because a similar porosified material is biocompatible, biodegradable [21.38], and offers a system of pores to load/unload therapeutics drugs rather than imaging contrast agents as in [21.39].

In contrast to porous silicon, gold nanoparticles are stable over time, can be easily functionalized, and may interact with an external electromagnetic (EM) radiation field to produce localized surface plasmon resonance and an amplification of that field [21.40, 41]. Gold nanoparticles (GNPs) with controlled geometrical, optical, and surface chemical properties are the subject of intensive studies of applications in biology and medicine. GNPs have been demonstrated in a variety

of applications, including genomics and biosensorics, immunoassays and clinical chemistry, photothermolysis of cancer cells and tumors, targeted delivery of drugs and antigens, and optical bioimaging of cells and tissues with state-of-the-art nanophotonic detection systems. In similar applications, gold nanoparticles have been produced in a variety of different, sometimes exotic shapes, including nanospheres, nanorods, sharpened nanorods, nanoshells, nanocages/frames, hollow nanospheres, tetrahedra/octahedra/cubes/icosahedra, rhombic dodecahedra, octahedra, concave nanocubes, tetrahexahedra, rhombic dodecahedra, obtuse triangular bipyramids, trisoctahedra, and nanoprisms. A similar diversity is justified by different EM amplification factors, transport properties in biological fluids and drug-loading capacities associated with different shapes. Similarly to gold NPs, silver NPs are used in diagnostic applications and therapeutic applications. The high antimicrobial activity of silver NPs has attracted much interest from scientists and technologists who want to develop nanosilver-based disinfectant products [21.42, 43]. Thus, in most of the therapeutic applications, it is the antimicrobial property that is being intensely explored. In a recent approach, the advantages of gold and silver have been united in bimetallic nanoparticles, as for example in [21.44].

Magnetic and superparamagnetic NPs have also been heavily investigated as sensitive probes for magnetic resonance imaging and therapeutic applications, tissue repair, immunoassay, detoxification of biological fluids, hyperthermia, drug delivery, and cell separation [21.45, 46]. In some cases, magnetic nanoparticles have been used to convert electromagnetic energy into heat [21.47] for applications such as disease treatment and remote control of single-cell functions. Recently, an approach based on magnetic particles and protein-assisted functionalization of graphene has been proposed for the production of hierarchical composites of functionalized graphene in polymeric matrices, exhibiting transparency and electron conductivity [21.48]. This shows that while materials science methods are effective in bionanotechnology, the opposite may be also true, and biology methods may be used to obtain supramolecular structures in which the organization of the structures depends on a biological driver, such as DNA assembly [21.49] in an approach that reconsiders the interactions of nanostructure with biological systems and highlights new directions in the design of biodegradable and multifunctional nanomedicine. Hybridization between organic and inorganic systems at the nanoscale is perhaps the new frontier in nanomedicine.

Polymeric nanoparticles have received great interest due to the versatility in which their structures can be modified to package and deliver their cargoes to the desired site of action or to respond to specific physiological or external stimuli [21.34, 50]. Polymeric nanoparticle-based therapeutics show great promise in the treatment of a wide range of diseases, due to the flexibility in which their structures can be modified. Advances in polymerization chemistries and the application of reactive, efficient and orthogonal chemical modification reactions have enabled the engineering of multifunctional polymeric nanoparticles with precise control over the architectures of the individual polymer components, to direct their assembly and subsequent transformations into nanoparticles of selective overall shapes, sizes, internal morphologies, external surface charges and functionalizations [21.34].

In what follows, we will cover specific nanoparticle-based systems for drug delivery and diagnosis, where the reported cases are thoroughly analyzed to offer a description of how, at the present state of technology advancement, a nanomedicine platform is realized, from a concept to a working prototype, passing through its design and fabrication.

21.2 Rational Design of Nanoconstructs

The intravenous administration of nanoconstructs for the precise delivery of imaging and therapeutic agents holds potential for the detection and treatment of deadly diseases, such as cancer and cardiovascular and neurodegenerative diseases [21.51, 52]. Nanoconstructs are man-made nanoscale objects that can be loaded with multiple different drug molecules as well as contrast agents. They are made out of biocompatible and biodegradable materials, such as lipids, polymers, and combinations thereof [21.53], and when transported by the blood flow can reach any location within the circulatory system and release their cargo [21.54, 55].

Over the last decade, a plethora of nanoconstructs have been demonstrated exhibiting different sizes, shapes, surface properties and, more recently, mechanical stiffnesses [21.13, 56, 57]. The characteristic size of a nanoconstruct can vary from a few tens of nanometers, as in the case of dendrimers and micelles, up to several hundreds of nanometers, as in the case of

large polymeric nanoparticles. Furthermore, although most nanoconstructs exhibit a spherical shape, multiple approaches have been developed for fabricating nanoparticles with discoidal, cylindrical, elliptical, stellar and other shapes. As for the surface properties, nanoconstructs with positive, neutral and negative electrostatic charges have been proposed. Also, multiple chemical moieties have been conjugated on the surfaces of nanoparticles to provide specific targeting properties. Finally, a few laboratories have recently started manipulating the mechanical stiffness of nanoconstructs leading to nanoparticles as soft as cells or as rigid as silica beads.

These four parameters – size, shape, surface and stiffness – have been shown to significantly affect the vascular and extravascular behavior of systemically injected nanoconstructs [21.13, 57]. However, a complete understanding of their effects on vascular margination and adhesion, crossing of the fenestrated endothelium, and diffusion into the diseased tissue of nanoconstructs is still missing. Indeed, given the range of variability for the design parameters, the number of possible combinations to be experimentally explored is impractically large. Consequently, the rational design of nanoconstructs requires the integration of experimental research activities with computational modeling, as schematically depicted in Fig. 21.1. The integration of in silico, in vitro and in vivo research modules, which is known as the integrated approach for the rational design, aims at identifying ideal combinations of the four design parameters – 4S parameters – that would maximize accumulation within the diseased tissue while minimizing the nonspecific sequestration by filtering organs, such as the liver, spleen and lungs [21.56].

Multiple computational techniques have been proposed for modeling transport problems including direct numerical simulations (DNS) [21.58], immersed finite element methods (IFEM) [21.59, 60], immersed boundary methods (IBM) [21.61], and isogeometric analysis (IA) [21.62, 63]. These approaches are all based on the numerical integration of the Navier–Stokes equations and their direct applications to modeling the vascular transport of multiple particles with complex shapes and mechanical properties is still challenging. However, particle-based techniques, such as the lattice Boltzmann method (LBM), provide simpler implementation schemes, higher flexibility and can be readily translated into parallel computing, allowing for a more efficient analysis of complex, multiparticle flows [21.64–66]. LBM uses the mesoscopic Boltzmann equation to determine macroscopic fluid dynamics and mass transport and has already been employed for modeling the vascular transport of particle suspensions and deformable cells [21.67, 68].

21.2.1 A Hierarchical Multiscale Approach for Vascular Transport of Nanoconstructs

A combined immersed boundary (IB) and lattice Boltzmann (LB) model has been developed for solving the transport of nanoparticles into vascular flows [21.68]. As an example, this computational framework allows for precise elucidation of the mechanisms that regulate the transport of particles in moderate Reynolds planar Couette flows. Under this condition, three forces govern the problem:

1. Inertial drift due to shear slip
2. Inertial drift due to rotational slip
3. Lubrication effect due to the presence of the wall.

Through direct numerical simulation, it has been shown that the difference between the actual particle horizontal velocity and the local undisturbed flow velocity (shear slip velocity, u_{Slip}) is responsible for the generation of a lateral force aimed at reducing such a difference. This force drives the particle towards flow regions where $u_{\text{Slip}} = 0$. Specifically, a particle would move towards a flow region of higher (slower) velocities if it lags (leads) the flow or, in other words, if $u_{\text{Slip}} < 0\ (> 0)$.

Such a behavior is depicted in Fig. 21.2a, where the linear slip velocity is plotted for particles in a planar Couette flow at Reynolds (Re) equal to 20. One particle is released at a quarter of the channel height, $y/H_{t=0} = 0.25$, and the other particle is released at three quarters of the channel height, $y/H_{t=0} = 0.75$.

The second contribution to the particle lateral drift, given by inertial drift due to rotational slip and described by *Joseph* and *Ocando* [21.58], shows that the rotational slip velocity is responsible for a lateral force. In particular, the angular slip velocity, $\omega_{\text{Slip}} = \omega + 1/2\gamma$, where ω is the angular velocity and γ is the local shear rate. Notice that $-1/2\gamma$ is the fluid angular velocity so that the nonzero angular slip velocity at the equilibrium implies a nonzero component of the lateral force generated by the rotational slip; see Fig. 21.2a.

Finally, the lubrication effect generates a lateral force pushing the particle away from the wall. The flow through the small meniscus at the particle-wall interface is strongly reduced as compared to the flow above the particle. Figure 21.3 demonstrates this mechanism, showing the profiles of a local slip velocity obtained at $tU_{\text{Max}}/L_{\text{ref}} = 10, 50$, and 500 along the line passing by the centroid of the particle. This local slip velocity is evaluated as the difference between the undisturbed linear Couette velocity profile and the computed horizontal velocity profile, namely $u_{\text{Shear}} - u$. Once the equilibrium position is reached, the local slip velocity

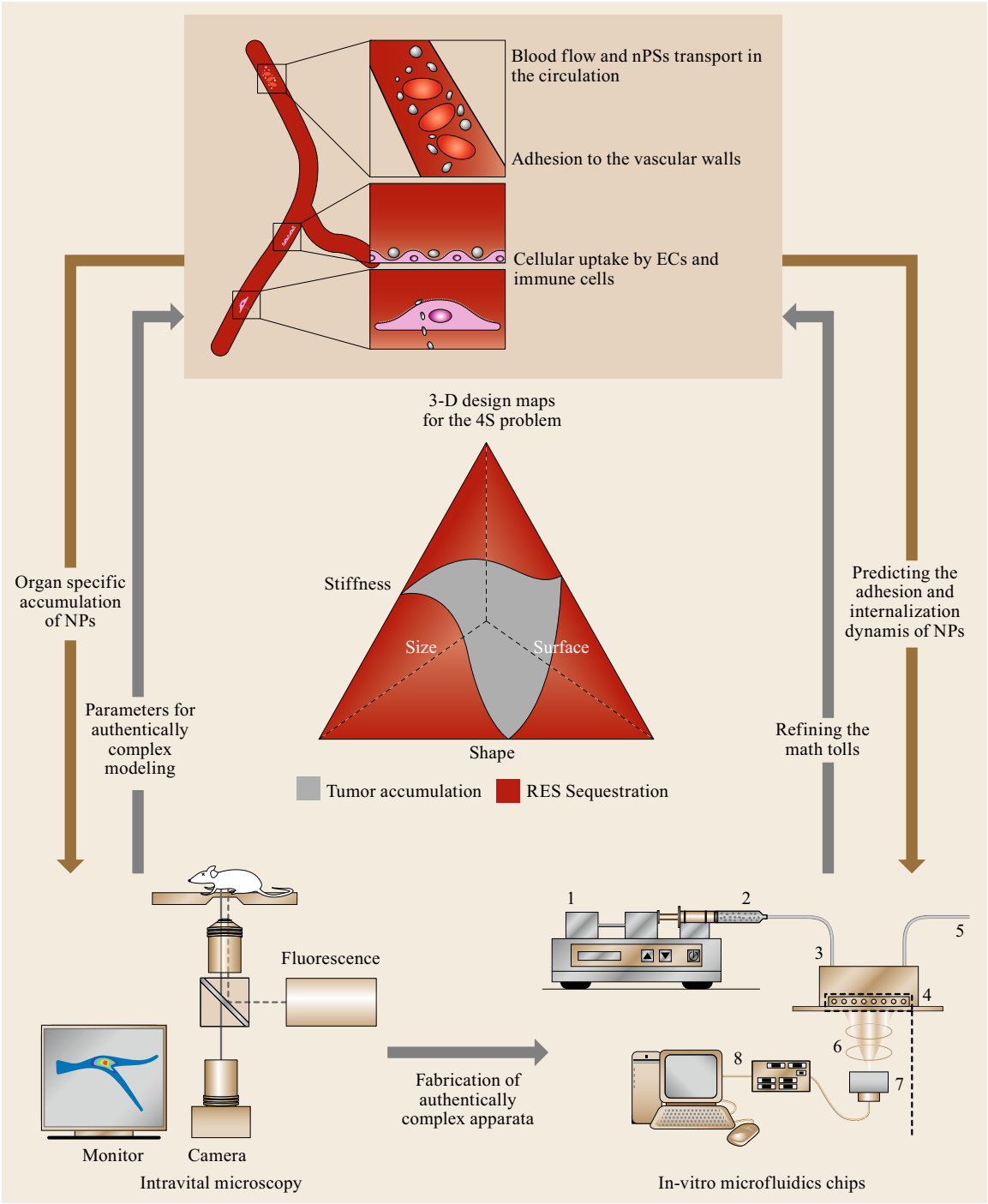


Fig. 21.1 Schematic representation of the research modules in the integrated framework for the rational design of nanoconstructs. Numbers 1–8 represent: 1) syringe pump; 2) syringe; 3) inlet; 4) flow chamber; 5) outlet; 6) microscope; 7) camera; 8) computer (after [21.56])

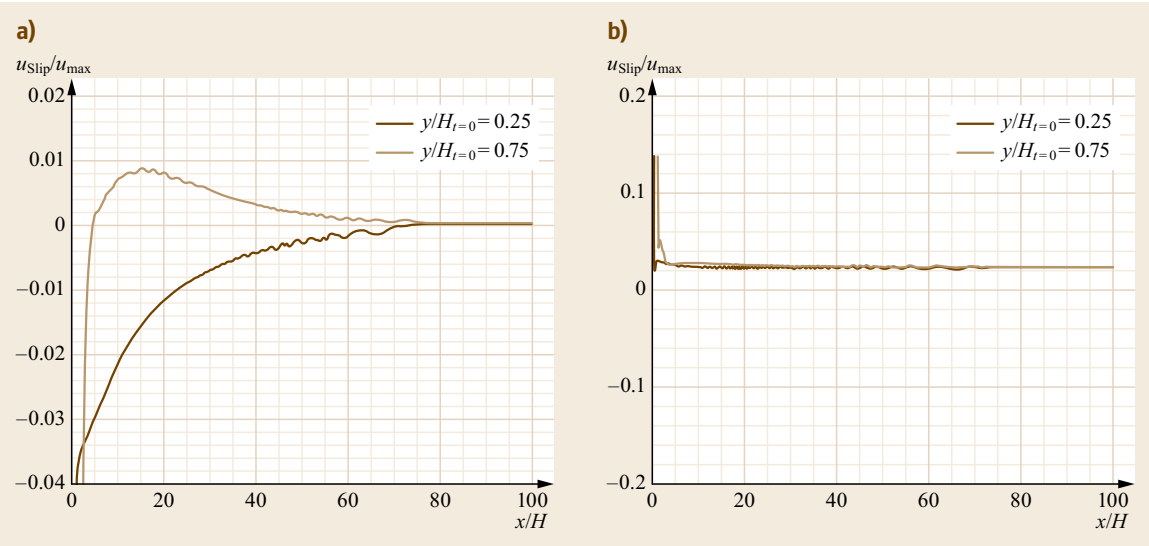


Fig. 21.2a,b *Circular particle lateral migration.* **(a)** Linear slip velocity distributions. **(b)** Angular slip velocity distributions (after [21.68])

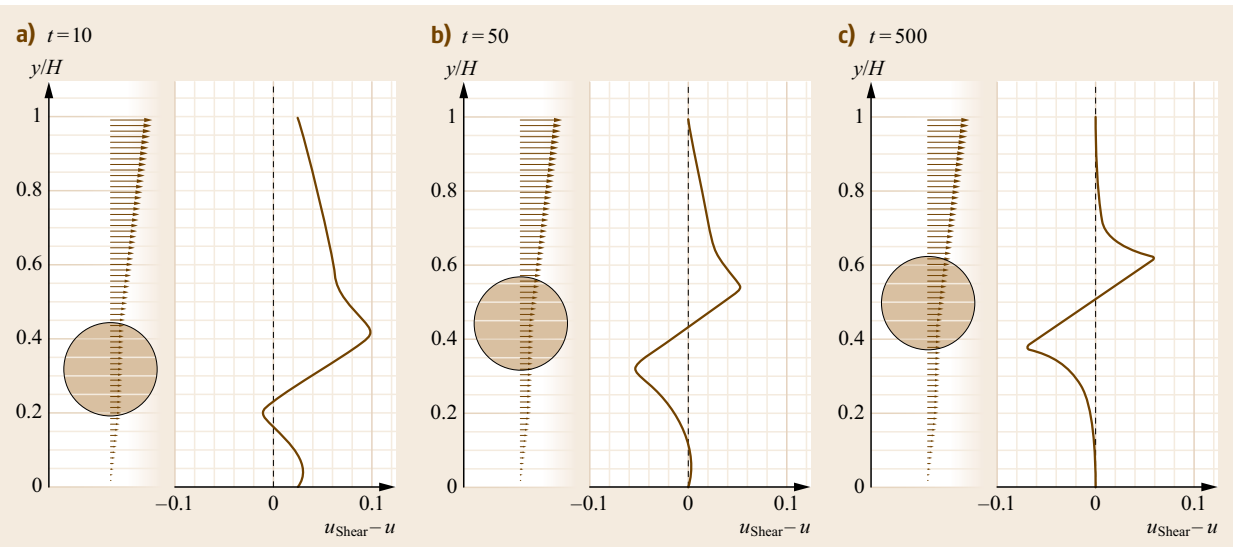


Fig. 21.3a–c Local slip velocity profile taken at the axis passing by the centroid and parallel to y at three different times for the particle released in $y/H = 0.25$ (after [21.68])

profile almost symmetrizes (never becoming symmetrical) and so the repulsive force from the wall becomes small (never becoming null).

In summary, the combination of immersed boundary methods based on continuum mechanics approaches, and lattice Boltzmann methods, allows one to define a computational framework for precisely elucidating the transport mechanics of nanoconstructs. Future efforts are needed to perform systematic analyses varying the Reynolds numbers, the shape and deformability of nanoconstructs, and including the adhesive interaction with vascular and other circulating cells.

21.3 Multifunctional Polymeric Nanoconstructs

The enhanced permeability and retention effect (EPR) has been extensively exploited for facilitating the accumulation of systemically injected nanoparticles within the tumor tissue [21.69]. This effect relies on the occurrence of openings – fenestrations – within the tumor vasculature, which can be passively crossed by long circulating nanoparticles with a diameter on the average of 100 nm. However, the size of fenestrations depends on tumor type, location and stage, and its high variability often limits the efficacy of EPR-based delivery strategies. Moreover, the occurrence of fenestrations in human patients seems to be far less common than in preclinical models of the same disease [21.69]. This has pushed scientists to identify novel and complementary strategies for enhancing nanoparticle accumulation within neoplastic masses. The rapid and uncontrolled development of cancerous tissues is generally associated with high tortuosity and permeability of the tumor vasculature and minimal activity of the lymphatic system [21.70]. This is often responsible for lower blood perfusion and mean blood velocity in tumor capillaries as well as higher interstitial fluid pressure compared to healthy tissues. Nanoconstructs can be designed to take advantage of such differences and specifically lodge within the tumor vasculature. In the following subsections, examples are given of nanoconstructs targeting tumors via the EPR effect (spherical polymeric nanoconstructs – SPNs) and by exploiting specific features of tumor vasculature (discoidal polymeric nanoconstructs – DPNs).

21.3.1 Spherical Polymeric Nanoconstructs (SPNs) for Combination Therapy and Biomedical Imaging

Two different polymeric nanoconstructs will be presented here for the systemic delivery of multiple drug molecules and contrast agents for imaging. The two nanoconstructs are spherical in shape with a diameter of about 100 nm and are composed of the same constituents: poly(lactic-co-glycolic acid) (PLGA), and DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine) and DSPE-PEG (1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[amino(polyethyleneglycol)-2000] (Carboxylic acid)) lipid chains. Due to these features, these nanoconstructs are referred to as spherical polymeric nanoconstructs (SPNs).

In one configuration, SPNs are synthesized by nanoprecipitation. PLGA polymeric chains and two therapeutic agents – docetaxel (DTXL), a potent anticancer molecule, and curcumin (CURC), an anti-inflammatory molecule – are dissolved in acetonitrile (oil phase) (Fig. 21.4a). The lipid chains are dissolved in 4% ethanol at 65 °C and added dropwise into the oil phase while heating. Water is added to reduce the temperature while the mixture is gently stirred. The SPNs' size and surface zeta potential are estimated by dynamic light scattering, whereas electron microscopy imaging analyses are employed to confirm sphericity and size homogeneity (Fig. 21.4b). The same nanoconstructs can be readily labeled with ^{64}Cu for the development of a positron emission tomography (PET) imaging nanoagent. Specifically, the radioactive ^{64}Cu can be entrapped within a DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) cage conjugated directly to the lipid chains constituting the nanoconstruct surface. For this, SPNs are resuspended in acetate buffer (pH 5.5) and incubated with ^{64}Cu for 1 h. Details methods can be found in previous works by the authors [21.71]. The efficacy of combination therapy derived by coloading DTXL and CURC within the same nanoconstruct is demonstrated in vitro and in vivo, using xenograft mouse models of a brain tumor. The systemic administration of SPNs – every two days for 20 days – is responsible for complete tumor regression within 40 days from the treatment initiation (Fig. 21.4c). The same SPNs are also labeled with ^{64}Cu(DOTA) to follow up on the progression of the disease and nanoconstruct accumulation within the neoplastic mass via PET imaging; Fig. 21.4d.

In a second configuration, SPNs are prepared via a sonication-emulsion method. In this case, PLGA, DPPC, and the two drug molecules – DTXL and diclofenac (DICL), a potent anti-inflammatory drug – are dissolved in chloroform (oil phase). DSPE-PEG is dissolved in 4% ethanol (aqueous phase). Then, the oil phase is added dropwise to the aqueous phase under ultrasonication. The resulting emulsion is placed in a low-pressure environment under magnetic stirring. A schematic representation and electron microscopy images of the resulting nanoconstructs are presented in Fig. 21.5a,b. The synergistic effect of the therapeutic combination is tested in vitro (Fig. 21.5c,d) and in vivo (Fig. 21.5e,f), in xenograft mouse models of brain tumor. Tumor response curves suggest that the combination of DTXL and DICL constitutes a potent strategy for inhibiting tumor growth (Fig. 21.5e,f) [21.72].

21.3.2 Discoidal Polymeric Nanoconstructs (DPNs) for Maximizing Accumulation Within Tumor Vasculature

Discoidal polymeric nanoconstructs (DPNs) are designed to firmly adhere to endothelial cells lin-

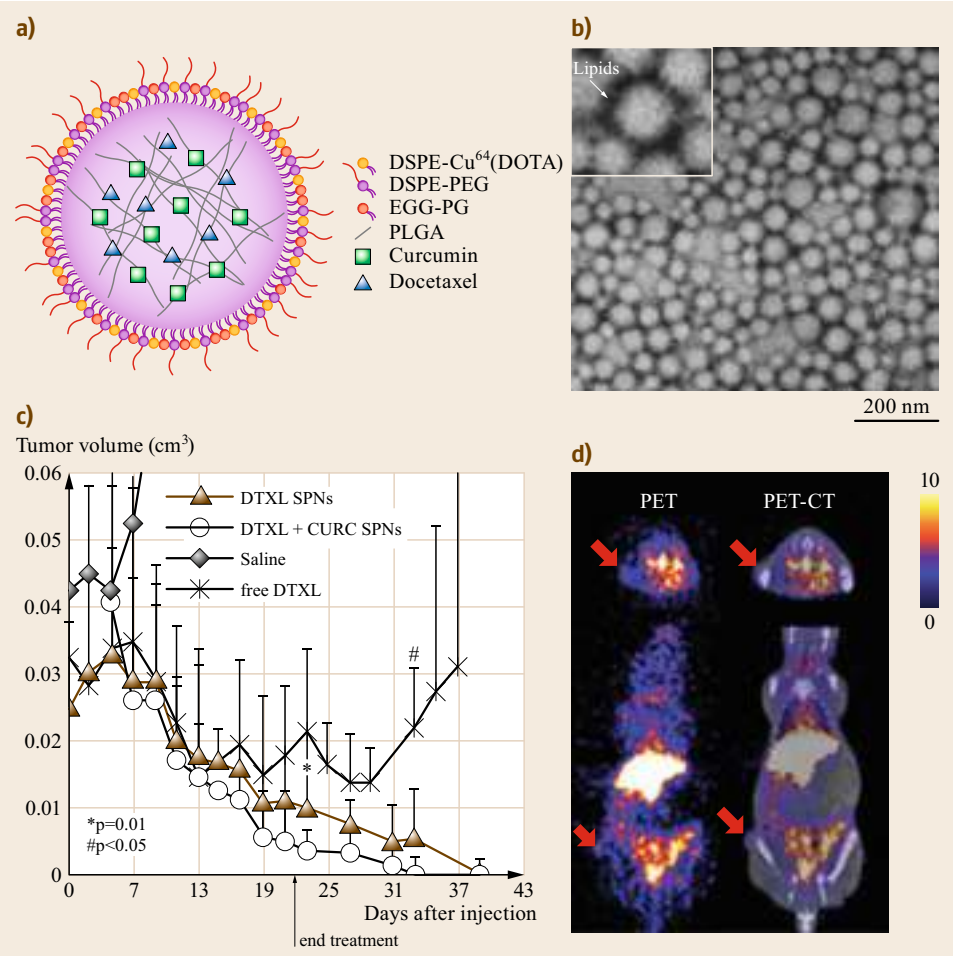


Fig. 21.4a–d *SPNs coloaded with DTXL and CURC for brain tumor therapy* (after [21.71]). **(a)** Schematic representation of SPNs. **(b)** Electron microscopy images of SPNs. **(c)** Brain tumor response to SPN combination therapy in xenograft mouse models. **(d)** Nuclear imaging of SPN accumulation in brain tumors over the course of treatment

ing the diseased vasculature in tumors [21.13]. It has already been shown that nonspherical submicron-sized nanoconstructs can efficiently navigate the circulatory system and sense vascular abnormalities by moving within the *cell-free layer*, behaving like platelets [21.73–76]. While nanoparticles smaller than 200 nm are mainly made through a bottom-up approaches based on emulsification and nanoprecipitation, DPNs are obtained via a modified hydrogel-template strategy [21.13]. The advantage in using a top-down approach rests in the ability to finely tune the size, shape, surface properties and mechanical stiffness (*4S parameters*) of nanoconstructs simultaneously and independently. DPNs' fabrication process starts with the preparation of a silicon master template via electron beam lithography (EBL). In one configuration, this master template comprises arrays of cylindrical holes with a 1 μm diameter and 400 nm depth. This structure is replicated into a polydimethylsiloxane (PDMS) template from which, eventually, several sacrificial poly(vinyl alcohol) (PVA) templates are fabricated. The PVA templates reproduce precisely the same structure as the original Si template. At this point, a polymeric paste is uniformly deposited on the sacrificial PVA template to fill up the wells. UV-light exposure polymerizes the polymers, which after dissolution in water of the sacrificial PVA template, returns the discoidal nanoconstructs [21.13].

DPNs present a highly porous and spongy matrix composed of the two main polymers entangled together: the hydrophobic PLGA and the hydrophilic PEG. Within this polymeric matrix, different therapeutic and imaging payloads can be dispersed depending on the biological applications. For instance, iron oxide nanocubes (NCs) with an edge size of ≈ 20 nm for magnetic resonance imaging, ablation therapy and magnetic guidance [21.13]; ^{64}Cu(DOTA) radioactive molecules for nuclear imaging [21.13]; and red fluorescent dyes (Rhodamine B or cy5) for optical imaging [21.13] can be directly incorporated in the polymeric matrix.

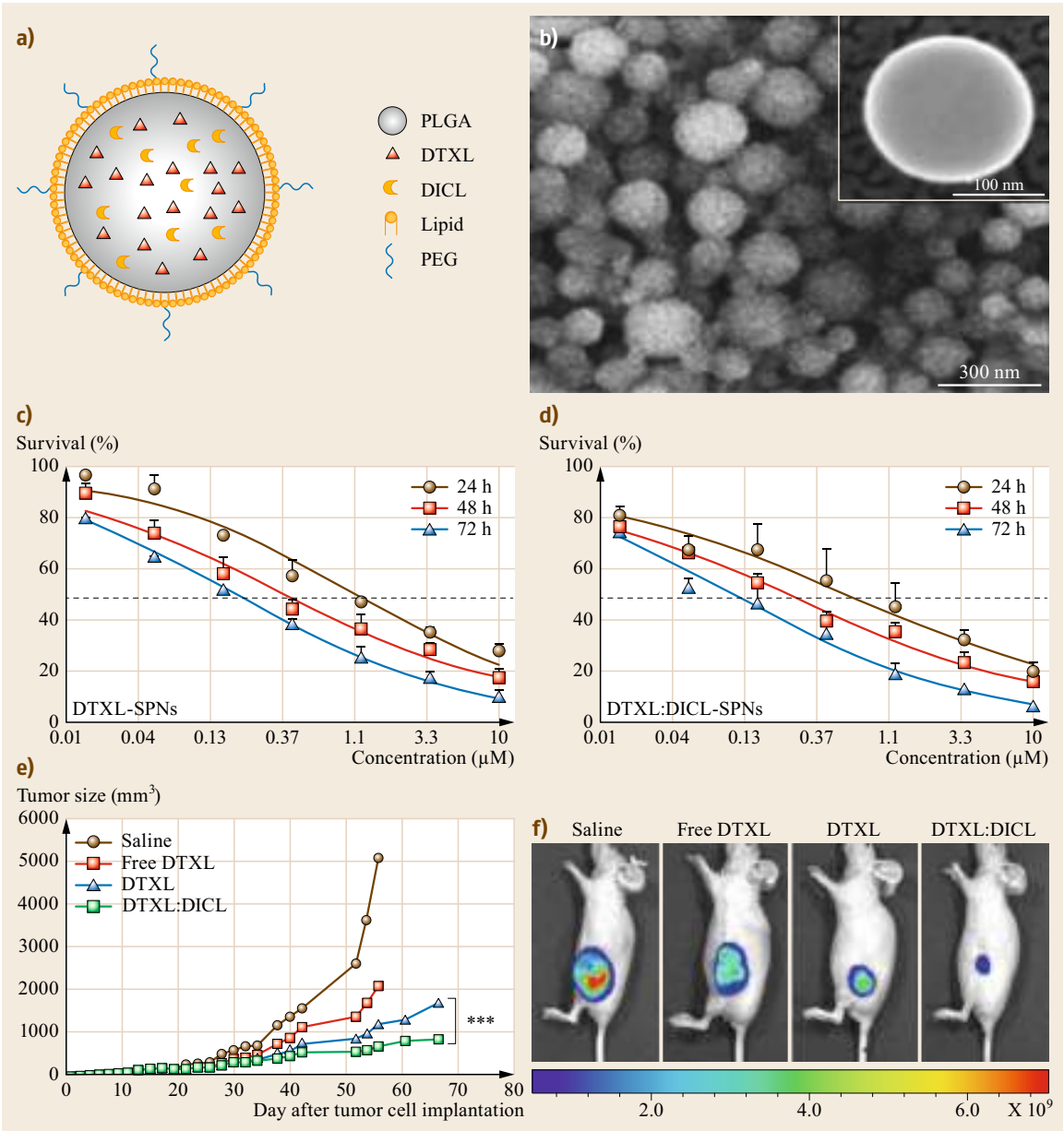


Fig. 21.5a–f *SPNs coloaded with DTXL and DICL for brain tumor therapy* (after [21.72]). **(a)** Schematic representation of SPNs. **(b)** Electron microscopy images of SPNs. **(c,d)** U87-MG brain tumor cell response to SPN combination therapy. **(e,f)** In vivo tumor response to SPN therapy and bioluminescence data for assessing tumor extension (after [21.72], with permission from Elsevier). *** indicates $P < 0.001$

As shown in Fig. 21.6a–d, the geometry of DPNs is characterized via scanning electron microscopy (SEM) and atomic force microscopy (AFM) before removal from the template, and optical microscopy in solution after dissolution of the PVA template. DPNs appear as disks with a circular base of ≈ 1000 nm and height of ≈ 400 nm. The Multisizer data presented in Fig. 21.6d show a size distribution with a peak at ≈ 800 nm, which results from averaging the two characteristic sizes (diameter and height). DPNs present a slightly negative surface electrostatic charge (−14 mV) due to the PLGA terminal carboxylic groups.

Besides controlling the size, shape and surface charge, the top-down fabrication process also allows

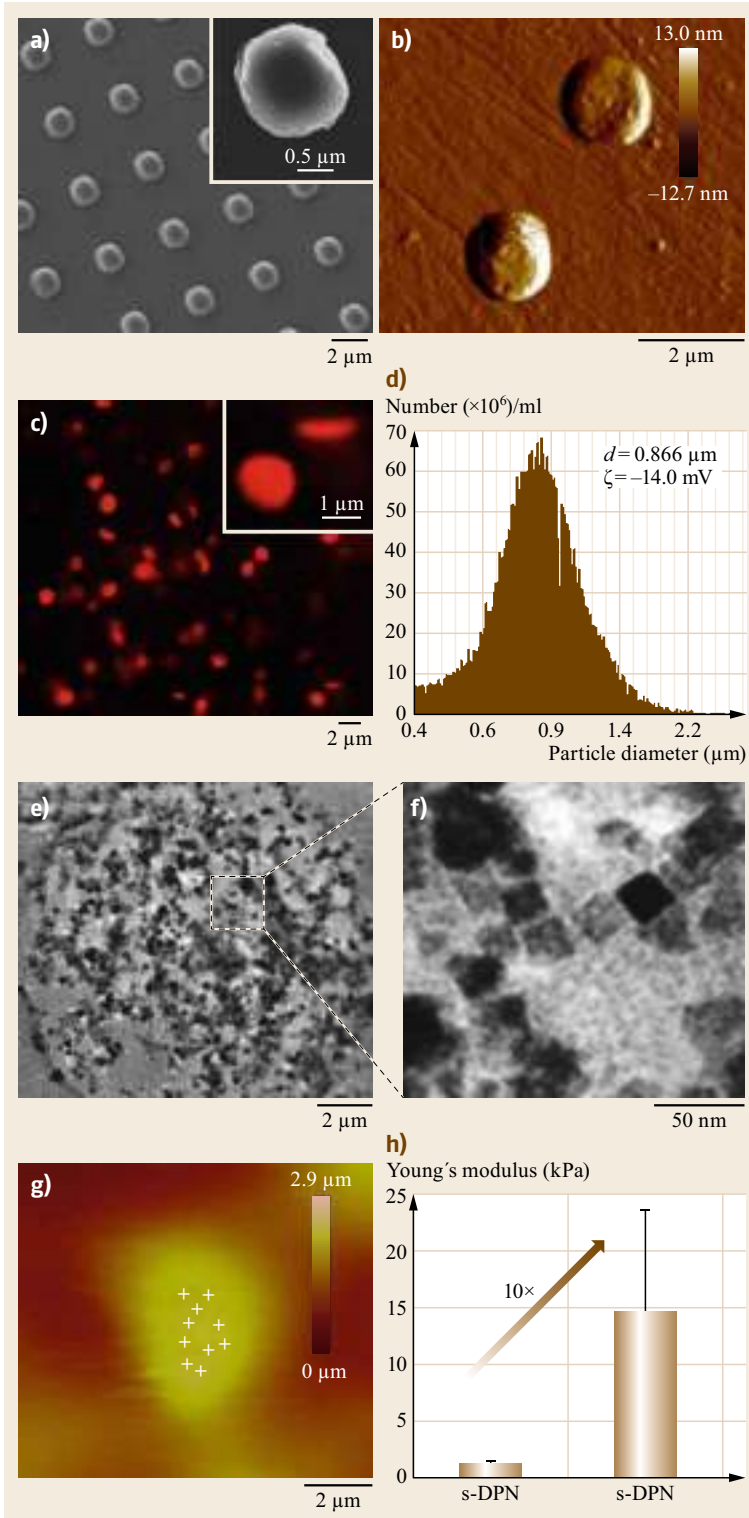


Fig. 21.6a–h *Physicochemical characterization of DPNs.* **(a–d)** Electron, atomic force and fluorescent microscopy characterization of 1000×400 nm DPNs. **(e,f)** Electron microscopy images of DPNs loaded with 20 nm NCs. **(g,h)** Assessment of DPN mechanical stiffness via atomic force microscopy. Adapted with permission from [21.13], copyright 2015 American Chemical Society

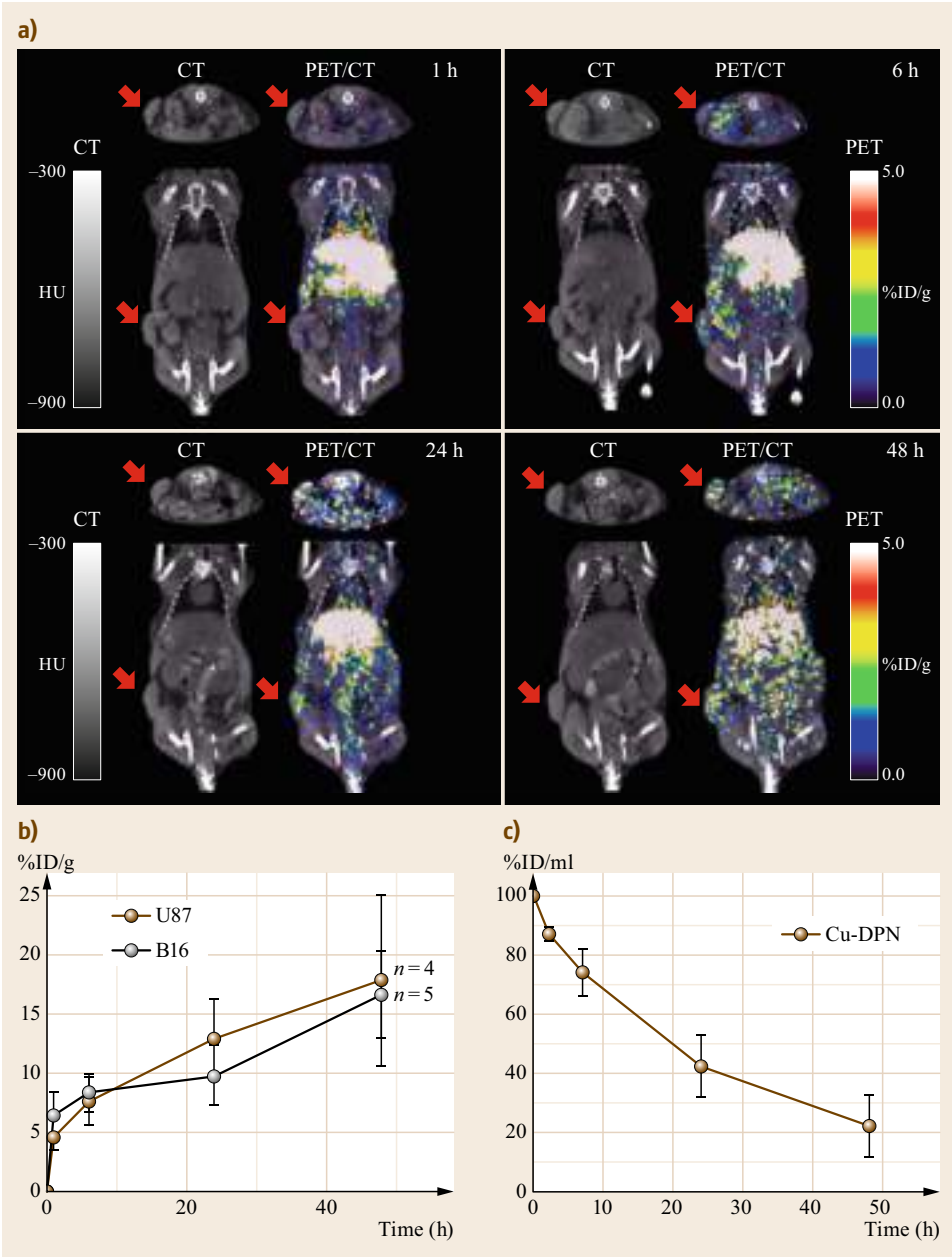


Fig. 21.7a–c *In vivo performance of soft DPNs.* **(a)** PET/CT imaging up to 48 h postintravenous injection of ^{64}Cu(DOTA)-DPNs. **(b)** DPN tumor accumulation. **(c)** DPN circulation half-life. Adapted with permission from [21.13], copyright 2015 American Chemical Society

the modulation of DPN mechanical properties. Specifically, DPNs with different Young's moduli (soft DPNs and rigid DPNs) are synthesized by changing the concentration of polymers within the original paste. This allows the production of nanoconstructs whose mechanical stiffness can vary from a cell-like Young's modulus (1.5 kPa) to more rigid structures ($\approx$ 20 kPa).

The effect of mechanical stiffness on the in vivo performance of DPNs – blood longevity, biodistribution and macrophage uptake – is assessed in mouse models of brain and melanoma tumors. Specifically, biodistribution studies are performed by injecting DPNs labeled with ^{64}Cu(DOTA) into mice and following the progressive accumulation of DPNs via PET/CT imaging. Figure 21.7a presents transversal and coronal sections of PET/CT images, converted into percentage of injected dose per gram tissue (%ID/g), taken at one, six, 24 and 48 h postinjection of DPNs. These images show, for the initial time points, a significant signal from the

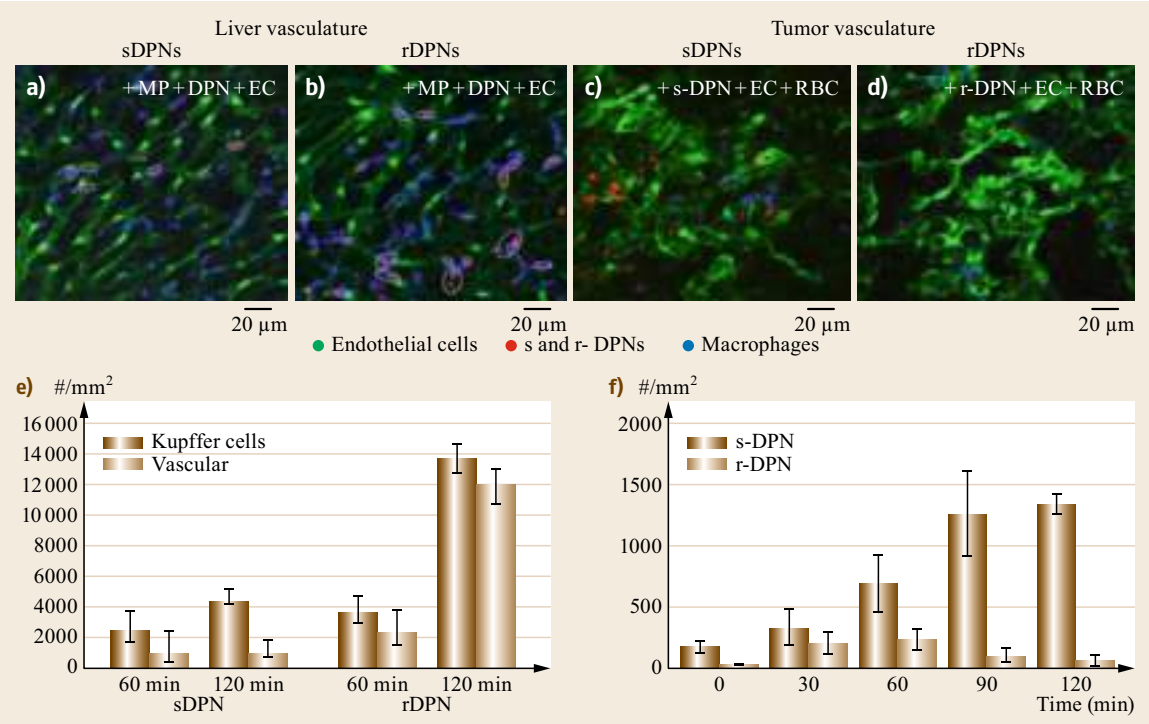


Fig. 21.8a–f *Intravital microscopy studies on soft and rigid DPNs.* **(a–d)** A set of representative liver and tumor images. Differential liver **(e)** and tumor **(f)** accumulation for the two different DPNs. Adapted with permission from [21.13], copyright 2015 American Chemical Society

liver and the abdominal cavity, which tends to reduce over time. Conversely, the signal within the tumor mass grows with time peaking at 24 and 48 h post-DPN injection via the tail vein. Tumor uptake of the soft DPNs reaches almost 20% ID/g tumor at 48 h postinjection (Fig. 21.7b). Note that this accumulation is much higher than what has been observed in similar tumors with SPNs. Figure 21.7c shows data on the blood concentration of DPNs demonstrating that a circulation half-life for the soft DPNs of ≈ 24 is estimated.

A direct comparison between soft and rigid DPNs is performed via intravital microscopy analysis (Fig. 21.8). This technique is used to quantify DPN differential accumulation in the liver and tumor mass. DPNs are systemically injected in tie-2 mice bearing a skin cancer model. These mice are genetically engineered to express green fluorescent proteins in the endothelial cells, so that the vasculature lights up in green under the microscope. As depicted by representative microscopy images (Fig. 21.8a–d), soft DPNs are less prone to sequestration by Kupffer cells in the liver while tending to more avidly deposit within the tumor vasculature as compared to rigid DPNs.

In summary, these examples continue to demonstrate that the size, shape, surface properties and mechanical stiffness of polymeric nanoconstructs should be properly designed and tailored during the fabrication process to maximize tumor accumulation and minimize sequestration by filtering organs, such as the liver, spleen and lungs.

21.4 Sensing and Drug Delivery with Porous Silicon Nanomaterials

Porous silicon (pSi) has been extensively researched and developed as a biomedical material since the first report of its biodegradability and bioactivity in the early 1990s [21.77]. Although pSi is a relatively new biomaterial, it benefits from advanced and established manufacturing, functionalization and characterization processes developed for silicon over several decades by the semiconductor, energy and electrochemistry industries, which accelerates its translation. In the last two-and-a-half decades of research, pSi has

proven simple and versatile to fabricate and functionalize, has displayed favorable interactions with biological systems, and was effective for drug delivery and biosensing.

21.4.1 Fabrication

Of the many top-down [21.82], bottom-up [21.83] and natural-source [21.84] approaches to obtain porous silicon, two main routes have taken center stage in biomedical applications: anodic etching [21.85] and metal-assisted chemical etching (MACE) [21.86]. In these approaches porous silicon forms by etching crystalline silicon wafers in hydrofluoric acid (HF) under conditions leading to local anisotropic oxidation of the silicon surface, which induces porosification. Anodic etching is the most widely documented route for generating porous silicon. During anodic etching, porous silicon forms by applying an anodic current through the wafer in an electrochemical cell with a HF electrolyte. Local injection of holes sustains oxidation at defects sites, followed by the dissolution of oxidized silicon in HF. Following an initial formation of pits at the site of surface defects, the pores propagate along crystalline axes [21.87] or current lines [21.88]. The porosification occurs mainly at the interface between the formed porous layer and the underlying pristine silicon [21.82]. In this way, porous silicon forms as a layer originating from the wafer surface and increasing in thickness as the etch progresses. It is possible to control pore diameter and porosity over wide ranges. The resistivity of the substrate, its doping type, the composition of the etching solution and the density of anodic current influence pore size [21.89]. Higher current density leads to larger pores and higher porosity, up to the limit where electropolishing of the surface occurs [21.82]. Since porosification occurs at the bottom of the porous layer, by modulating the current over time it is possible to vary pore morphology along the axis of etching. This variation can either generate pores with continuously variable size, or distinct discontinuous layers [21.90]. Ultimately, by repeating set cycles of current variation it is possible to generate periodic porous silicon structures (Fig. 21.9a). Due to the spatial variations in the dielectric constant associated with porosity, such periodic structures act as photonic crystals, which find extensive use in biosensing [21.91].

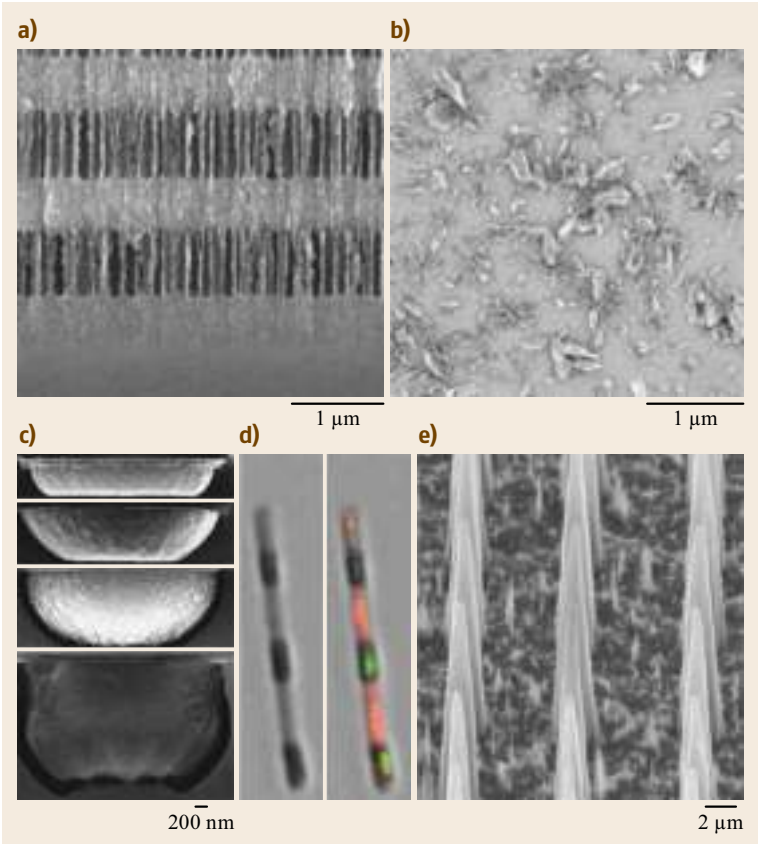


Fig. 21.9a–e *Porous silicon nanostructures.* **(a)** Photonic crystals of pSi obtained with periodic variation of anodic current (after [21.78], courtesy of Wiley). **(b)** Polydisperse pSi nanoparticles obtained from sonication of anodically generated freestanding porous silicon layers (after [21.79], copyright 2011 American Chemical Society). **(c)** Tailored particles obtained from photolithography combined with anodic etching (after [21.80], courtesy of Wiley). **(d)** Nanobarcodes obtained from multilayer metal-assisted chemical etching through variation of oxidant concentration (after [21.80], courtesy of Wiley). **(e)** Nanoneedles obtained from photolithography combined with metal-assisted chemical etching (after [21.81])

Crushing freestanding porous silicon layers forms particles. Forming an electropolishing layer using high anodic currents detaches a pSi layer from the substrate [21.92]. Ultrasonication [21.93] or ball milling [21.94] then crushes the layer into micro- or nanoparticles (Fig. 21.9b). Particles are then size-selected by differential sedimentation in a centrifuge [21.95] or by filtering [21.96]. This strategy is rapid and high yield but grants only limited control over particle size, and no control over particle shape. Controlling size and shape of a nanoparticle influences its behavior as a drug delivery system by determining its biodistribution [21.74, 97], degradation profile and biointerface [21.98]. Combining photolithographic patterning with anodic etching provides tight control over size and shape of the resulting particles (Fig. 21.9c). This strategy forms discoidal, quasihemispherical and tubular nanoparticles, with aspect ratios between two and six and base axes between 400 nm and 3.2 µm [21.99–101]. Modulating the etching forms multilayer nanoparticles. This approach can form a thin layer at the top of the particle to guarantee structural stability, followed by a layer with porosity over 80% and pores larger than 60 nm that would not otherwise be stable by itself [21.100]. Particles with such large pores and high porosity can improve drug loading efficiency and can load large *second-stage* nanoparticles in a *multistage* drug delivery approach [21.101]. Microfabrication combined with anodic etching increases the production yield for tailored porous silicon nanoparticles. This approach first forms porous silicon pillars by deep reactive ion etching, and then selectively porosifies such pillars into stacks of discoidal or quasihemispherical particles.

MACE is more versatile than anodic etching, since it combines anisotropic silicon etching and porosification in a single process [21.99, 102]. In MACE, a metal nanostructure (thin film, nanoparticle, dendrite) catalyzes charge injection into silicon from an oxidizing solution of hydrofluoric acid [21.103]. The charge injection is localized at the silicon–metal interface and the hydrofluoric acid dissolves the oxidized silicon. The speed and extent of silicon oxidation depends on the density and mobility of injected charges [21.104]. This in turn controls the etching rate, the profile and the porosification of the silicon [21.80]. Combining and controlling lithography, metal deposition and MACE yields a wide variety of silicon nanostructures, both porous and nonporous. By tuning silicon resistivity and the etchant composition it is possible to finely tailor pore size and porosity during MACE [21.80]. Etch duration controls the size of the nanostructure along the etching axis, with the etching rate depending on the resistivity of the silicon and the etching solution [21.80, 105]. MACE can modulate pore size and porosity along the etching axis by varying oxidant concentration over time [21.80]. Combining MACE with anodic etching varies the porosity of the nanostructures by controlling the anodic current applied over time [21.106]. These nanostructures are lifted off the substrate either by creating an electropolishing layer as the last process step [21.106] or by mechanical removal [21.80]. The resulting structures are freestanding high-aspect-ratio photonic crystals (nanobarcodes) (Fig. 21.9d). Tailoring metal deposition controls the geometry of the resulting nanostructure. Electroless deposition of noble metal nanoparticles from solution or the deposition of discontinuous metal films (subpercolation threshold) yields vertically aligned, randomly ordered nanowires [21.102]. Patterning the location of the metal deposition with photoresist [21.107], nanospheres [21.108] or a dielectric layer [21.80, 81] creates arbitrary base patterns for the nanostructures, for the formation of nanopillars [21.80, 108], nanofins [21.80] and nanoparticles [21.109]. The typical diameter of MACE nanowires obtained ranges from 10−100 nm; nanopillars and nanofins have typical dimensions in the 300−2000 nm range; and nanoparticles can be as small as 400 nm. Post-MACE plasma etching of the nanopillars shapes them into nanoneedles, with a tailored tip diameter down to less than 50 nm (Fig. 21.9e) [21.81]. Multiple cycles of electroless metal deposition and MACE on nanopillars yield a selective etch of the rim, shaping them into nanorods, nanocones or nanopencils depending on the details of the procedure [21.110]. MACE combined with deep silicon etching forms solid pillars with a thin porous silicon film at the top [21.111].

21.4.2 Surface Chemistry

The surface of a material is a primary determinant of its performance for drug delivery and sensing, as the interaction with biological systems occurs at the interface. Also, the overall properties of porous silicon strongly depend on its surface, due to the high surface-to-volume ratio. The native surface of porous silicon is hydrogen terminated [21.112]. In this native form, porous silicon is photoluminescent, electroluminescent and chemoluminescent [21.113]. The porous structure and the resistivity of the silicon controls the position of the emission band, ranging from the blue to the near-infrared. Although not extremely efficient, thanks to this luminescence pSi nanomaterials can combine imaging alongside their primary delivery and/or sensing task [21.96]. The hydrogen termination of silicon is thermodynamically unstable and in ambient conditions it progressively forms a nonstoichiometric *native* oxide layer [21.114]. The surface defects within this ox-

ide layer affect the luminescence and quench it to levels unsuitable for any application [21.115]. In pursuit of passivation strategies that preserve this valuable optoelectronic behavior, several surface derivatizations were developed [21.116]. Thermal, plasma or wet oxidation yield a thicker and more stoichiometric oxide [21.81, 117]. This silicon oxide surface provides simple and extremely versatile routes of functionalization through silane chemistry [21.118]. Organofunctional alkoxysilanes form covalent Si–O–Si bonds with the hydroxyl groups on the silicon oxide surface, creating a self-assembled mono- or multilayer depending on the details of the process. One of the four arms of the alkoxysilanes can carry a functional group that acts either as the starting point for bioconjugation strategies, or provides functionality at the biointerface. Typical silanization processes yield amine, sulfhydryl and aldehyde groups, which serve as the starting point for many bioconjugation strategies and biological interactions [21.119].

Thermal carbonization in the presence of acetylene yields a silicon carbide surface, either C or CH_2 terminated, depending on the process conditions [21.120]. The CH_2 surfaces can be oxidized to form hydroxyl groups for silane functionalization, or the surface hydrogens can mediate hydrosilylation with undecylenic acid. The pristine hydrogen surface can also be directly hydrosilylated through thermal, photochemical or lewis acid mediated reactions [21.85]. The undecylated surface provides a carboxylic acid for biological interaction and functionalization.

21.4.3 Biointerface

Porous silicon is biodegradable, with a degradation kinetic influenced by its porous structure, its surface chemistry and the environment. Oxidized porous silicon readily undergoes hydrolysis and dissolution by attack of OH^- ions (Fig. 21.10a) [21.121].

The kinetic of this dissolution depends on the surface-to-volume ratio (i. e., on the porosity and wall thickness size) and on the concentration of OH^- ions in solution [21.124]. Surface functionalization by silanization [21.80] or with polyethylene glycol [21.125] slows the degradation kinetic of porous silicon particles in a chain-length-dependent fashion, likely by hampering the diffusion of nucleophilic moieties and that of the dissolved silicon to and from the particle surface. Porous silicon surfaces that degrade very slowly due to their low porosity can catalyze the deposition of minerals from biological buffers and contribute to the mineral formation process necessary for osteogenesis [21.77, 126]. Surface carbonization improves the stability of porous silicon in aqueous solutions, significantly slowing the degradation kinetics and the ability of alkaline solutions to attack the porous structure [21.127]. There is strong evidence that intravenously injected porous silicon particles degrade in vivo [21.99, 128–130], and that their degradation kinetics strongly depends on their microenvironment [21.128]. The changes in their photonic properties that occur with degradation allow monitoring of the process in vivo [21.131].

Passivated porous silicon is a highly biocompatible material. Although the native hydrogen termination can generate reactive oxygen species [21.132], its oxidized form is safe at the concentrations relevant for biomedical applications [21.95, 123]. Porous silicon as a cell culture substrate [21.126, 133, 134] and in nanoparticle form [21.35, 95] interacting with a wide range of mammalian cells does not exhibit significant toxic effects. The carbonized surface is less cytocompatible than the oxidized one, in a dose- and size-dependent fashion [21.95]. Animal models prove the safety of multiple pSi administration routes, including oral [21.135], intravenous [21.129], intraocular [21.136] and intratumoral [21.137]. Intratumoral administration is safe in humans [21.137]. Orally administered porous silicon particles are not absorbed through the intestine [21.135], while their intravenous injection does not modulate cytokines, indicating a favorable immune response profile [21.129].

Phagocytic and nonphagocytic cells internalize porous silicon micro- and nanoparticles in a surface chemistry-dependent fashion [21.95, 138]. Positively and negatively charged particles are readily opsonized and internalized by cells, while particles functionalized with polyethylene glycol moieties are more resistant to opsonization and internalization [21.123]. Coating the particles with a leukocyte-derived membrane also delays their recognition and internalization by phagocytic cells [21.37]. These *leukolike* particles cross epithelial layers more efficiently than *naked* pSi ones and injected intravenously they display prolonged circulation times. The particles are internalized by endocytotic pathways and trafficked through the endolysosomal system to the perinuclear region of cells [21.56, 139]. A nanoparticle payload loaded within the pores can be sorted within the endolysosomal system and follow a different fate, which can ultimately lead to exocytosis (Fig. 21.10b) [21.123].

Flatter, discoidal nanoparticles marginate and interact with epithelial layers more efficiently than hemispherical ones [21.74]. In vivo, the size and aspect ratio of the particles affects their biodistribution (Fig. 21.10c) [21.74, 99]. Smaller particles more efficiently negotiate the organs of the mononuclear phagocytic system and the small capillaries of the alveoli, accumulating more efficiently at a tumor target site [21.74, 99]. Flat, discoidal particles accumulate

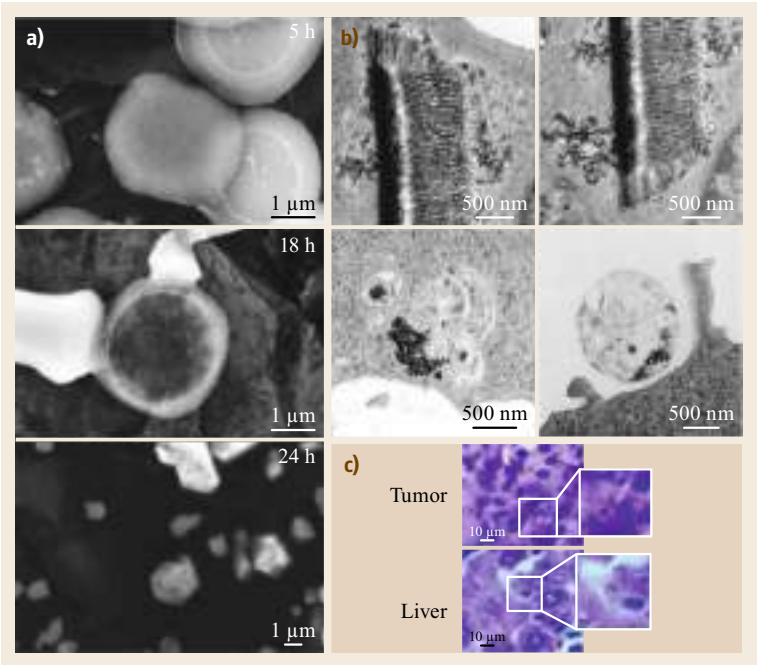


Fig. 21.10a–c *Porous silicon biointerface.* **(a)** Biodegradation of porous silicon particles in cell culture media over the course of 24 h (after [21.122], with permission from Elsevier). **(b)** Intracellular trafficking and sorting of a multistage porous silicon carrier and its iron oxide payload. The porous silicon carrier is trafficked to the perinuclear area while the payload is recycled through exocytosis (after [21.123], courtesy of Wiley). **(c)** Accumulation of porous silicon particles within tumor and liver of a murine breast cancer model (after [21.99], courtesy of Wiley)

preferentially in the lungs and heart, while cylindrical ones accumulate in the liver. Live imaging of injectable pSi particles tracks their biodistribution in real time [21.96, 130, 140]. Time-resolved fluorescence imaging enhances the signal originating from pSi over the fluorescence tissue background, improving the limit of detection [21.130].

21.4.4 Drug Delivery

Porous silicon particles efficiently load a variety of payloads. They load small-molecule drugs across a wide range of solubilities with differences in acid/base character and lipophilicity, and from multiple biopharmaceutical classifications [21.94]. They modulate the solubility of the payload as a function of pH [21.94]. In general pSi encapsulation enhances the solubility of poorly soluble drugs and reduces that of highly soluble ones [21.94]. Porous silicon particles efficiently load biological payloads, including proteins [21.141], peptides [21.142] and nucleic acids [21.81, 108], while mediating their sustained release over several hours and days. The mesoporous structure provides a harvesting effect thereby concentrating the biological payload with respect to the loading solution [21.81, 100]. Capping the porous structure with pH-responsive moieties, such as pH-responsive polymers [21.123] or nanovalves [21.143], enables triggered or smart drug delivery. A polymeric coating protects biological payloads from degradation [21.141]. An agarose coating prevents access of proteolytic enzymes within the porous structure, increasing the stability of a protein payload. This coating does not affect the degradation profile of the particle, nor the release profile of the payload.

Porous silicon particles also load and release nanoparticles in a dose-dependent fashion (Fig. 21.11a) [21.35]. The loading efficiency depends on the relative pore-to-particle size, with a pore size twice that of the particle providing ideal loading [21.100]. Particles larger than the pores do not load due to size exclusion [21.80]. Sequentially loading nanoparticles from the largest diameter to the smallest achieves compartmentalized loading into porous silicon structures (Fig. 21.11b) [21.35, 80].

Porous silicon nanoparticles mediate light absorption and photothermal therapy by conjugation with porphyrins [21.145]. The delivery of antibiotics from pSi particles sustains antibacterial activity over several weeks [21.146]. The particles release chemotherapeutic agents to kill cancer cells both in untargeted and targeted fashions. The *leukolike* system can effectively cross an endothelial barrier and deliver doxorubicin to the far side of it, causing the death of cancer cells while preserving the viability of the endothelium (Fig. 21.11c) [21.56]. Intraperitoneal delivery of a pSi multistage system delivers siRNA for sustained silencing of target genes, inducing ovarian tumor regression in mouse models [21.129]. Similarly, intravenous injection of polymeric doxorubicin

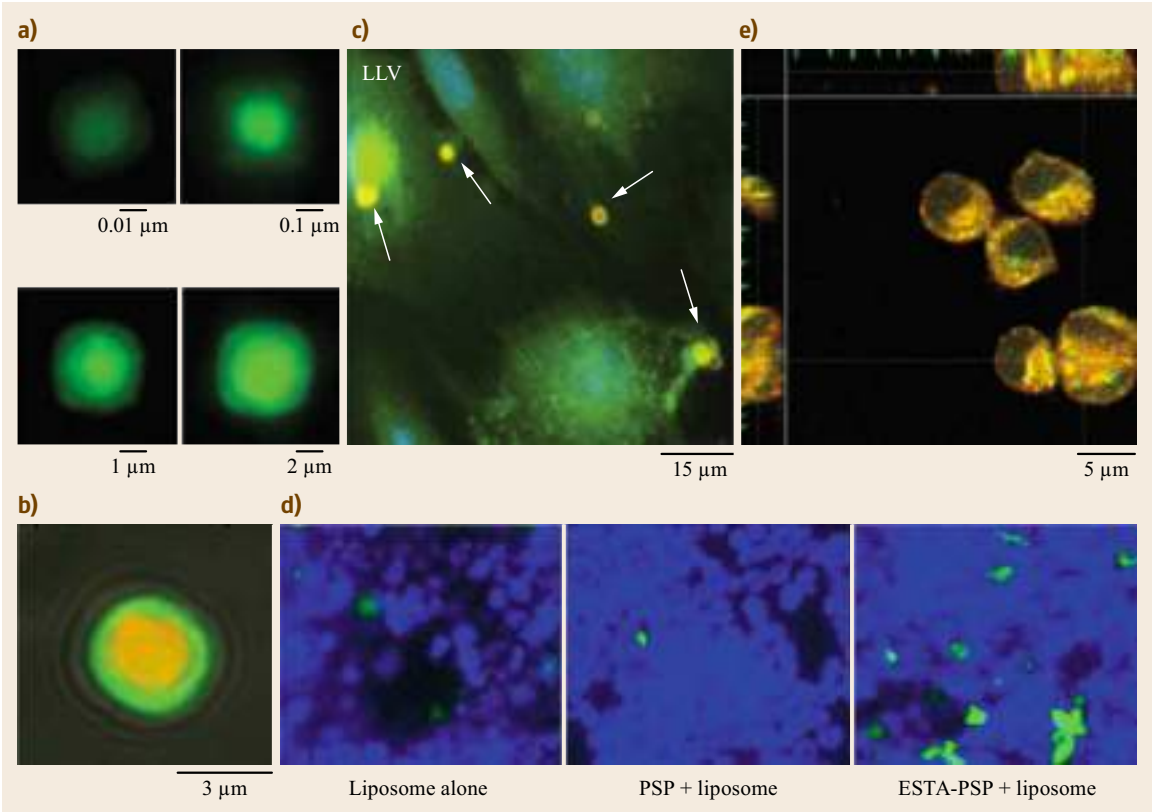


Fig. 21.11a–e *Porous silicon for drug delivery.* **(a)** Dose-dependent loading of quantum dots (*green*) into a porous silicon particle (after [21.35]). **(b)** Size-dependent compartmentalized loading of carbon nanotubes (*green*) and quantum dots (*red*) into a porous silicon particle (after [21.35]). **(c)** In vitro delivery of nucleic acids from porous silicon nanoneedles. Green fluorescent protein expressed from plasmid (*green*) and siRNA (*red*) (after [21.81]). **(d)** Association of leukolike particles to inflamed endothelium in vitro through LFA1–ICAM interaction (after [21.37]). **(e)** Association of porous silicon particles with bone marrow through e-selectin targeting in a murine model. Targeting efficiency is compared between delivery of payload alone (liposome alone), delivery with pSi particles (PSP + liposome) and delivery with targeted pSi particles (ESTA-PSP + liposome). Cell nuclei are in *blue*, liposomes are in *green* (after [21.144], courtesy of Wiley)

nanoparticles loaded within the porous silicon matrix accumulates doxorubicin at metastatic liver sites inducing their regression, and improving overall survival in a mouse model [21.147]. Porous silicon functionalized with bioactive molecules to actively recognize tumors effectively localizes at the tumor site and improves therapeutic efficacy over untargeted approaches (Fig. 21.11d) [21.54, 144, 148]. Porous silicon particles can deliver immunostimulators eliciting a host response in the context of immunotherapy [21.149].

Porous silicon nanoneedles are efficient delivery systems for nucleic acids [21.81] and nanoparticles [21.150]. They interface tightly with the cell and deliver their payload directly to the cytosol, bypassing the endolysosomal system (Fig. 21.11e) [21.150]. Nanoneedles interfacing with cells (nanoinjection) modulate expression of target genes following delivery of siRNA or plasmid DNA for up to 92% of the cells in culture [21.81]. Nanoinjection in vivo is safe and does not elicit detectable immune reactions [21.81]. In the first example of nanoneedle-mediated in vivo gene therapy, nanoinjection localized delivery to a defined area within the muscle of a mouse and delivered a plasmid encoding for the human vascular endothelial growth factor (VEGF) [21.81]. The nanoinjection induced localized expression of the target protein and the formation of new blood vessels connected to the vasculature of the host [21.81].

Human trials with pSi particles proved their safety, and the efficacy of locally delivered brachytherapy through intratumoral injection of neutron-activated porous silicon nanoparticles [21.137].

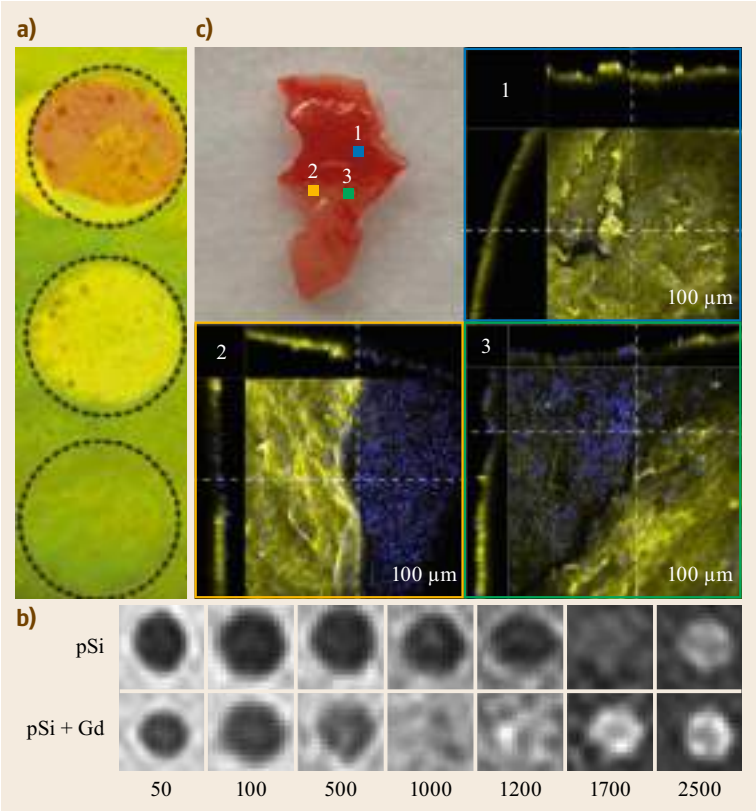


Fig. 21.12a–c *Sensing with porous silicon.* **(a)** Colorimetric detection of protease activity on a porous silicon photonic crystal. Higher protease activity changes the crystal color from *green* to *red* (after [21.151], courtesy of Wiley). **(b)** Enhanced r_1 relaxivity of gadolinium payload when loaded into porous silicon. Comparison between the relaxivity of the porous silicon particle alone (pSi), and the gadolinium loaded (pSi + Gd) (after [21.39]). **(c)** Mapping protease activity at the tumor margin with porous silicon nanoneedles. Protease activity is detected by intracellular fluorescent signals in three locations across the margin. Cell nuclei are in *blue*, fluorescent signal is in *yellow* (after [21.152], courtesy of Wiley)

21.4.5 Sensing

Biosensors can leverage the photonic properties of porous silicon to detect small molecules [21.153], proteins [21.151], nucleic acids [21.154] and pathogens [21.155]. The general principle of operation relies on the adsorption of biomolecules inducing a shift in the transmittance or reflectance spectrum of a porous silicon photonic crystal (Fig. 21.12a) [21.156]. The adsorption process can be made specific for a target analyte through a selective enzymatic substrate [21.157] or the bioconjugation of a complementary biomolecule that selectively recognizes it. Ligand–receptor interactions [21.158], base-pair complementarity [21.154] and antibody/aptamer–target interactions [21.159] are among the most common strategies developed to generate porous silicon biosensors. Most of these biosensors are thin films developed for use ex vivo, but can be formulated as particles [21.93]. While optical reporting is the most common strategy, electrochemical [21.159] reporting strategies are employed when advantageous.

Porous silicon nanostructures are efficient Raman substrates. The porous structure can host metal nanoparticles mediating surface-enhanced Raman scattering to achieve ultralow limits of analyte detection in biologically relevant buffers [21.111].

The natural photoluminescence of porous silicon makes it an ideal candidate for near-infrared imaging. The organ distribution and clearance of pSi particles can be imaged live [21.96, 130, 140]. Porous silicon particles absorb ultrasound, mediating both imaging and thermal therapy [21.160]. The confinement of gadolinium nanoparticles within the porous silicon structure enhances their longitudinal relaxivity, improving their performance as MRI (magnetic resonance imaging) contrast agents (Fig. 21.12b) [21.39, 161].

Porous silicon nanoneedles, thanks to their tight interaction with cells, can detect enzymatic activity within the cell cytosol. A fluorescently labeled peptide cleavable by Cathepsin B and bioconjugated to the nanoneedles can discriminate cancer cells from healthy ones in coculture, thanks to their cytosolic enzymatic activity [21.152]. The same sensor can discriminate tumor clinical samples of tissue from healthy ones, and image the intracellular activity of Cathepsin B across a tumor margin region (Fig. 21.12c).

21.5 Nanomedicine Synthesis by Microfluidics Technology

In drug development the poor water solubility of drugs is one of the major problems. Thus, for the last decades, the pharmacological properties of free drugs have been improved using different strategies [21.162–165]. For example, drugs can be integrated into carriers for protection from acidic pH or enzymes that may degrade them, with tunable release properties for increasing their bioavailability [21.166–168]. This is particularly important for chemotherapeutic agents, which usually have low solubility and stability, leading to deleterious off-target side effects and poor therapeutic efficacy. Targeted therapy in cancer is very important, and different nanocarriers have been extensively used as nanomedicine tools for the treatment of solid tumors [21.96, 169–171]. For example, porous silicon (pSi) materials [21.89, 165, 172] have been extensively used in drug delivery and biomedicine due to their properties, which can overcome some of the challenges in the field of drug delivery [21.142, 165, 173]. However, like in many drug delivery systems, these materials per se do not provide fully controlled drug delivery of the payloads. To overcome this hurdle different fabrication strategies of nanomaterials have been reported, including microfluidics technology [21.13, 174].

In general, this technique can be used to precisely manipulate fluids in the micrometric scale in a highly reproducible way [21.176], using droplet-based methods [21.176] for carriers in the micrometric range or mixing-based microfluidics based on the nanoprecipitation principle for preparing nanorange delivery systems [21.81, 178]. This technique allows the production of drug delivery systems with high encapsulation efficiency, and a precise control of the size and dispersity of the carrier. This is particularly important for the co-delivery of multiple therapeutics using the same carrier, allowing for the precise control over the drug loading and release for simultaneously encapsulated cargoes and facilitating combination therapy, especially if hydrophobic and hydrophilic payloads are codelivered, and minimizing the batch-to-batch variability.

In this section, the microfluidics technique for preparing nanomedicines and its applications in cancer are presented and discussed with practical examples of our most recent research on this topic, focusing on the fabrication of multiple therapeutics loaded into multicomponent platform systems with precisely controlled drug loading and release prepared by microfluidic-based platforms.

21.5.1 Why Microfluidics for the Synthesis of Nanomedicines?

During the last years, microfluidics has impacted the fields of biological analysis, chemical synthesis, tissue engineering, colloid generation, and cell encapsulation [21.179–182]. Microfluidics offers great possibilities, such as the capacity to use tiny amounts of samples and reagents, detection and separation of compounds with high resolution and sensitivity, and exploitation of the laminar flow occurring at the microchannels at a reduced cost [21.183]. The manipulation of multiphase flows in microfluidics can be used to produce lipidic and polymeric particles by single emulsion-templating [21.184, 185], double emulsion-

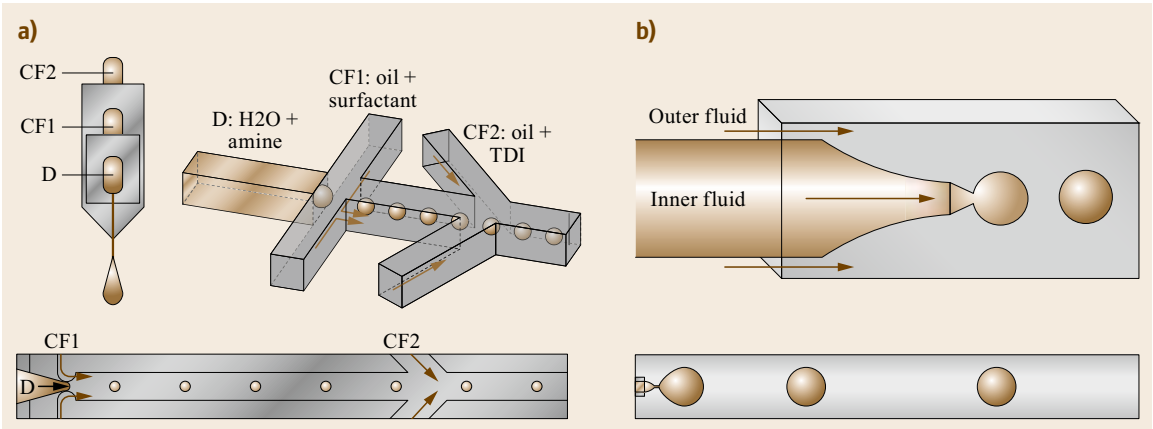


Fig. 21.13 **(a)** Schematics of the geometry of droplet formations in a PDMS microfluidic device; *D*: dispersed fluid, *CF*: continuous fluid for emulsification, *CF*2: continuous fluid for encapsulation (after [21.175], copyright 2015 American Chemical Society). **(b)** Glass capillary flow-focusing microfluidic device; a truly 3-D flow can be achieved, as compared to the PDMS device (after [21.176, 177], reproduced with permission)

templating [21.186], and rapid mixing [21.81]. The size of the droplets and nanoparticles can be finely tuned by controlling the flow rates of the fluids pumped inside the capillaries, where the viscosity and interfacial tension of the fluids, and the type and concentration of the polymers or lipids used, stabilizes the droplets or forms the nanoparticles, which are highly monodisperse and have high encapsulation efficiencies [21.81, 187–189].

The production of particles can be achieved by using microchannels etched primarily on a Si rubber of poly(dimethylsiloxane) (PDMS) using soft lithography [21.190], by chemical attachment to a glass slide, or by a circular capillary (or a series of them) inserted concentrically into another capillary and coaxially aligned and glued to a glass slide [21.47]. In the case of glass capillary, the production of truly three-dimensional (3-D) flows can be achieved [21.176], and they are very robust and compatible with a wide variety of solvents [21.59], as compared to the PDMS ones (Fig. 21.13). However, a high production efficiency in the assembly can be achieved with microchannels in PDMS platforms. The fluid flow manipulation offered by the microfluidics technique allows for the control of the material formation dynamics, thus facilitating the fabrication of many advanced drug delivery systems, e.g., double and higher-order emulsions, microcapsules, nanoparticles and other colloidal delivery systems [21.13, 174, 176, 177, 179, 191–193]. In addition, this technology further offers precise control over the shape, size, and geometry of the carriers, as well as the loading of different cargoes [21.174, 179, 194–196].

The fabrication of nanocarriers using the microfluidics technique [21.81, 198], based on the nanoprecipitation approach [21.199] and the formation of nanoparticles (Fig. 21.14), has been suggested to be triggered by the *Ouzo effect* [21.157, 200]. In this case, the mixing process in microfluidic channels can be laminar or turbulent. The former is achieved through hydrodynamic flow-focusing, in which an inner stream is narrowed and focused enabling the rapid diffusion of solvents and nonsolvents inside and outside the stream, and thus the nanoprecipitation of the solute [21.197]. Thus, the size, polydispersity, and the drug loading and release of nanoparticles can be controlled by varying the flow rates, the polymer composition, and the polymer concentration [21.197]. When increasing the viscosity of the liquids or the flow rate or when decreasing the density of the fluids, a higher Reynolds number (Re) can be achieved [21.81, 178]. The transition from low to high Re dictates changes in the mixing patterns of the fluids inside a microfluidic device from laminar flow to microvortices and unsteady jetting [21.81]. In this case, nanoparticles can be prepared with smaller sizes at higher Re [21.81, 178]. Furthermore, the size of the nanoparticle has been shown to be more sensitive to Re than to the solute composition, demonstrating the importance of the mixing speed in the formation of nanoparticles [21.178].

Recently, micro- and nanoparticles have been produced with microfluidics for excellent control of size, defined composition and morphology, as well as size distribution [21.81, 142, 201–205]. Santos' group and collaborators have used microfluidic devices consisting of an assembly of borosilicate glass capillaries on a glass slide to develop advanced micro- and nanocarriers to encapsulate and deliver different cargoes (e.g., small drug molecules and peptides) [21.81, 142, 202, 203]. For example, a microfluidic platform for synthesizing diverse types of homogeneous nanoparticles with mass production rates up to 242.8 g/day was reported [21.81]. Microfluidics was also successfully used to fabricate a pH-responsive multistage system based on polymeric–inorganic nanoparticles for simul-

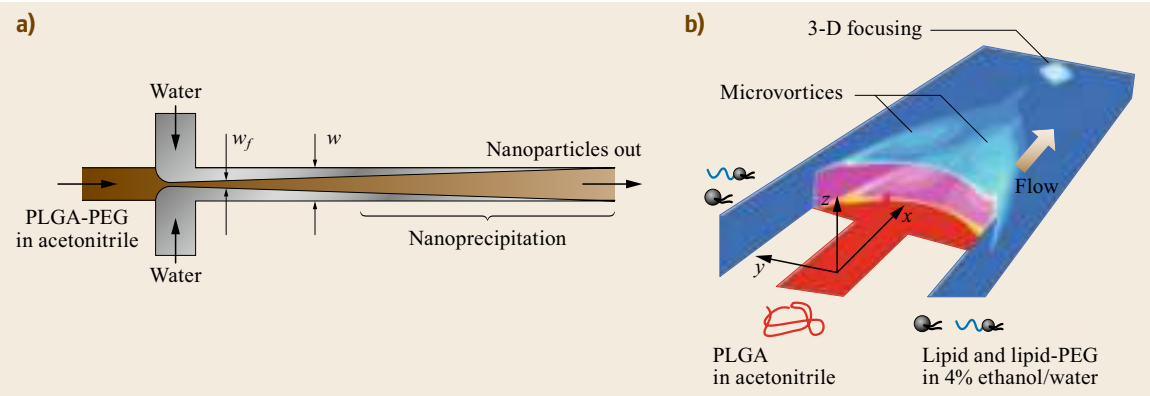


Fig. 21.14a,b Schematic representation of the nanoprecipitation process in a microfluidic device by **(a)** hydrodynamic flow focusing (after [21.197]) and **(b)** microvortices (after [21.178], copyright 2017 American Chemical Society)

taneous delivery of fluorouracil and celecoxib for colon cancer combined therapy [21.203]. In addition, pSi nanoparticles have been assembled in giant liposomes on a microfluidic chip for coloading and codelivery of multidrugs combined with synthesized DNA nanostructures, gold nanorods, and magnetic nanoparticles for overcoming the multidrug resistance in cancer cells [21.201]. Similarly, a multifunctional composite consisting of chitosan and cell-penetrating peptide-modified poly(lactide-co-glycolide) (PLGA) and pSi nanoparticles was prepared by microfluidics, specifically designed for simultaneous codelivery of peptide (glucagon-like peptide-1) and an enzymatic inhibitor (dipeptidyl peptidase 4) for therapy of diabetes [21.202].

21.5.2 Microfluidic Production of Nanocarriers Using PDMS Microfluidic Devices

As a result of the advantages of the microfluidic technology, diffusion, emulsification, and mixing approaches have been employed for the continuous preparation of a variety of nanocarriers, including liposomes [21.206], lipid–polymer hybrid nanoparticles [21.207], polymeric nanoparticles [21.208], and polymeric micelles [21.197]. The majority of these nanocarriers have been fabricated with PDMS microfluidic devices, and more recently also with glass capillaries by Santos' group. A few examples of these nanocarriers produced by PDMS are discussed below.

Nanoliposomes and Polymersomes

Using a microfluidic flow-focusing method, *Jahn* et al. fabricated nanoliposomes by employing a central flow of a phospholipid-containing ethanol solution focused at a microchannel cross-junction between two aqueous buffer streams [21.209]. As a result of the dilution of ethanol above a critical concentration, the lipids spontaneously self-assembled into nanoliposomes, which depending on the flow rates, were monodisperse carriers with sizes of 50−150 nm [21.206, 209]. Block copolymer vesicles or *polymersomes* have also been fabricated using a microflow-focusing device (MFFD) for the control of their size [21.210]. The polymersome size was tuned from 40 nm to 2 μm with a narrow size distributions and excellent reproducibility by changing the flow rate ratio in the water and ethanolic inlet channels, thus forming the polymersomes at the ethanol–water interface.

Polymeric Nanoparticles

A very limited number of studies have been reported for the preparation of polymeric nanoparticles using the microfluidic approach, and the majority of these studies rely on the diffusion-governed mixing processes to induce the self-assembly of polymers for cancer treatment. For example, *Kolishetti* et al. developed dual-drug-loaded nanoparticles with precise control over the ratio of payloads with varying physicochemical properties [21.208]. To achieve this, a water-soluble Pt(IV) prodrug was conjugated with the polylactic acid (PLA) polymer backbone, and then a blend of Pt(IV)-conjugated PLA and carboxyl-terminated PLGA-*block*-poly(ethylene glycol) copolymer (PLGA-PEG-COOH) together with docetaxel was pumped into the microfluidic channels (Fig. 21.15a–c). As a result, a great drug encapsulation efficiency and high loading of both hydrophilic and hydrophobic drugs was achieved, further enabling the ratio control over these two drugs. Taking advantage of the functional groups of the PLGA-PEG-COOH polymer, an A10 aptamer was further conjugated onto the nanoparticles' surface for specific targeting of the prostate cancer cells [21.211]. A sustained release for both drugs over a period of 48−72 h was also observed (Fig. 21.15d), independent of the drugs' solubility.

Majedi et al. have also used solvent-free microfluidics for the production of paclitaxel (PTX)-loaded chitosan nanoparticles with fine-tuned physicochemical properties [21.212]. With this method the formation of nanoparticles was driven by changing the pH conditions in an aqueous environment. The avoidance of the organic solvent contamination and the mixing time during the nanoprecipitation process fixed the size and surface potential of the nanoparticles. Consequently, the obtained nanoparticles showed pH-dependent drug release profiles (Fig. 21.15e), as a result of the different solubility of the chitosan at different pH conditions.

Hybrid Nanocarriers

Recently, polymer and lipid hybrid nanocarriers have also been fabricated using microfluidics [21.207, 213]. These complex lipid–polymeric hybrid nanoparticles combine the unique features of both liposomes and polymeric nanoparticles, possessing high drug encapsulation efficiency, high stability in physiological fluids and relatively slow drug release kinetics. The fabrication of the hybrid nanoparticles can be done in a single mixing step by controlling the microscale mixing processes (microvortex), thus achieving hybrid nanoparticles production rates as high as 3 g/h with excellent reproducibility [21.207]. By changing the flow rates in the microvortex platform, the particle size can be tuned (30−170 nm).

In another example, a PDMS-based microfluidic mixer consisting of Tesla structures was also demon-

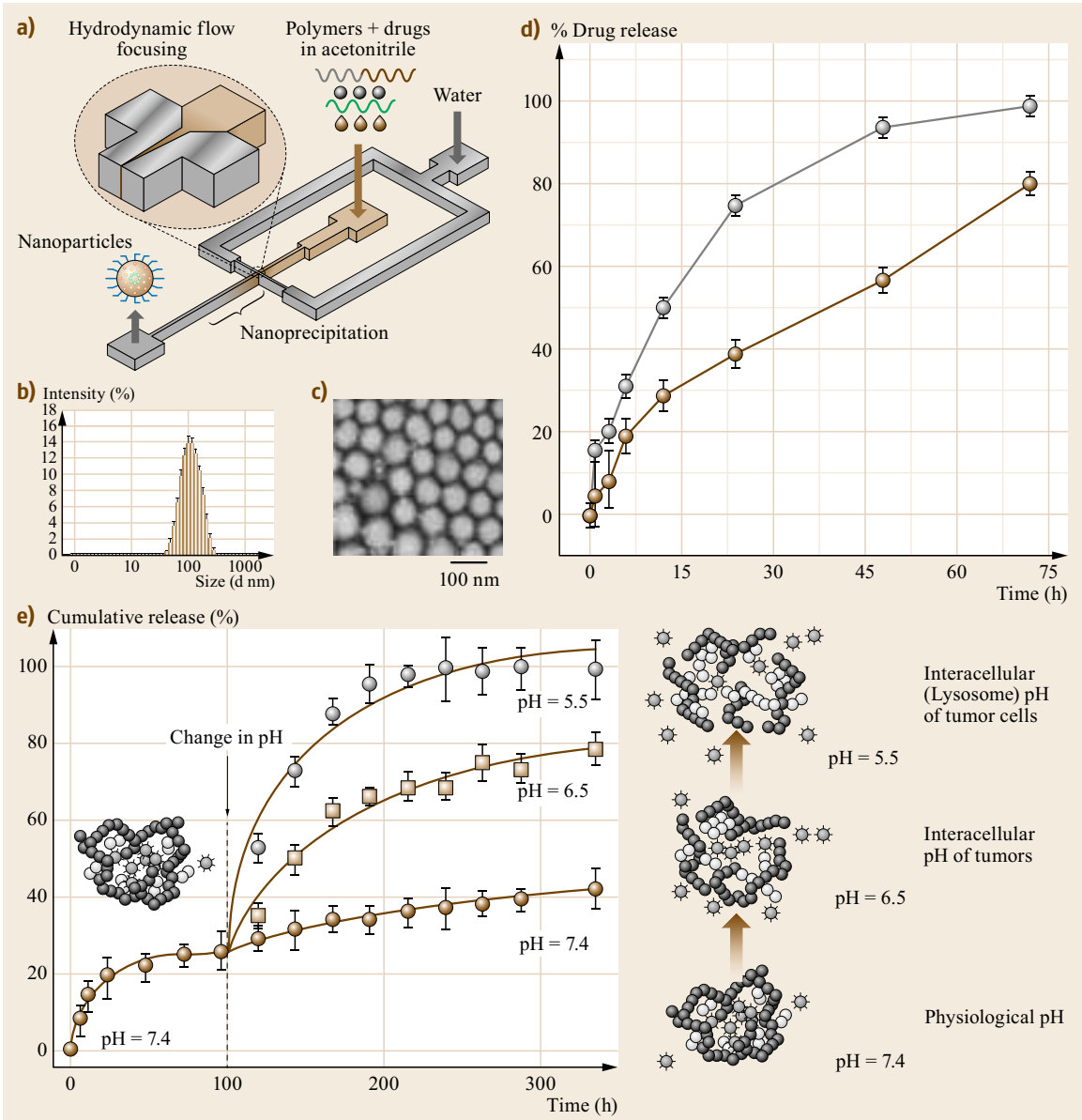


Fig. 21.15 **(a)** Synthesis and characterization of nanoparticles by dynamic light scattering **(b)**, and **(c)** transmission electron microscope (TEM). **(d)** In vitro release kinetics of the encapsulated Pt (*brown*) and docetaxel (*grey*) from the nanoparticles in pH 7.4 at 37 °C (after [21.208]). **(e)** Release of PTX from nanoparticles synthesized by microfluidics after pH change at 37 °C (after [21.212])

strated to be effective in the production of lipid–polymeric hybrid nanoparticles (monodisperse of sizes of 40 nm) by enhancing the mixing of fluids in transverse dispersion [21.214]. This technique has also been used to fabricate hybrid lipid–quantum dot (QD) nanocomposites with a homogeneous size distribution of 60 nm, where the nanocomposite contained on average four QDs encapsulated per nanocomposite.

21.5.3 Microfluidic Production of Nanocarriers Using Glass Capillary Microfluidic Devices

Microfluidic glass capillary devices are not yet as common as the PDMS ones. However, a few studies have been reported in the literature and two examples are discussed next.

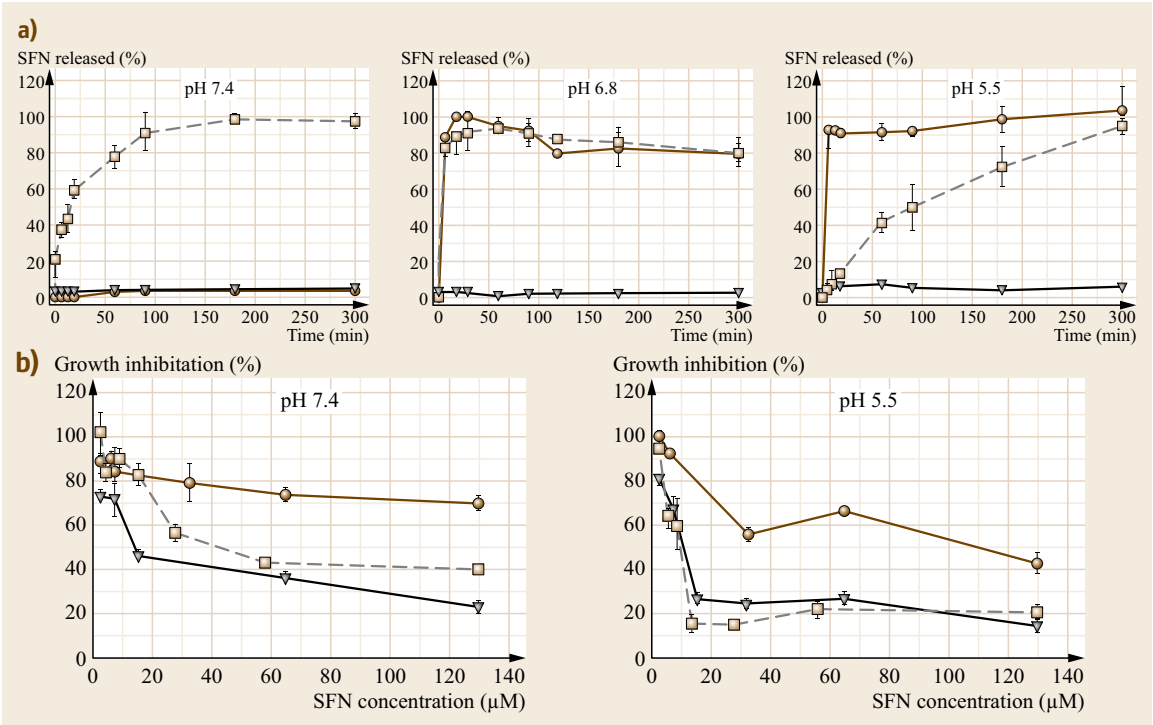


Fig. 21.16 **(a)** Dissolution and drug release profiles of SFN (*solid brown line*), SFN-loaded nanoparticles (*dashed black line*) and nanocomposite (*solid gray line*) in plasma pH 7.4 and PBS enriched with 10% serum at pH 6.8 and 5.5. **(b)** Growth inhibition of PC3MM2 cells incubated with SFN (*black solid line*), SFN-loaded nanoparticles (*dashed gray line*) and the nanocomposite (*solid line*) at pH 7.4 and 5.5 for 24 h (after [21.142])

In order to overcome some of the challenges of the drug nanocarriers for the treatment of tumors, such as the poor water solubility of certain anticancer drugs, fast clearance from the bloodstream and lack of controlled drug delivery at the tumor site [21.215, 216], a smart hybrid nanocomposite was developed consisting of a blend of PEG-*block*-poly(L-histidine) (PEG-PHIS) and PEG-PLA assembled on the surface of sorafenib (SFN)-loaded pSi nanoparticles (circa 130 nm), spontaneously formed by nanoprecipitation in a flow-focusing microfluidic chip [21.142]. The final nanocomposite had a size of 260 nm and a surface charge of −5 mV. This nanocomposite was designed for controlled drug delivery upon acidification of the environment triggered by the tumor extracellular microenvironment or the intracellular endosomes or lysosomes (typical pH values of 5–6.8 [21.217, 218]). PEG-PHIS possessed a pH-responsive behavior due to the presence of an imidazole ring in each histidine unit of the polypeptide bearing a protonable nitrogen, thus a pH-triggered drug release from the pores of the pSi nanoparticles upon the rupture of the polymeric structure assembled on its surface could be achieved. SFN, a cytotoxic drug that inhibits cell proliferation and angiogenesis, was loaded into the pores of the pSi nanoparticles and further encapsulated into the polymeric structure, with a final drug loading degree of 4.9% [21.142]. The SFN release profiles from pSi nanoparticles and the nanocomposite at different conditions, in order to mimic the physiological tumor microenvironment and the intracellular conditions, are shown in Fig. 21.16a. This same system was the used to assess how the pH could trigger the release of SFN from the nanocomposite in order to inhibit the growth of prostate cancer cells, PC3MM2, in different experimental conditions (Fig. 21.16b) [21.142]. It was observed that the concentration of SFN needed to inhibit the cell population to half (IC_{50}) at both pH values was lower than 15 μM, with an IC_{50} of ≈ 27 μM and 130 μM at pH 7.4 for the pure drug and the nanocomposite respectively. At pH 5.5, the IC_{50} of the nanocomposite was reduced to 32 μM, showing that the acidic conditions triggered the release, and thus the drug effectively inhibited the proliferation of PC3MM2 cells.

Increasingly sophisticated strategies are required for combination cancer therapy with co-administration of multiple drugs for synergistic benefits. Combination therapy can be accomplished by coloading multiple

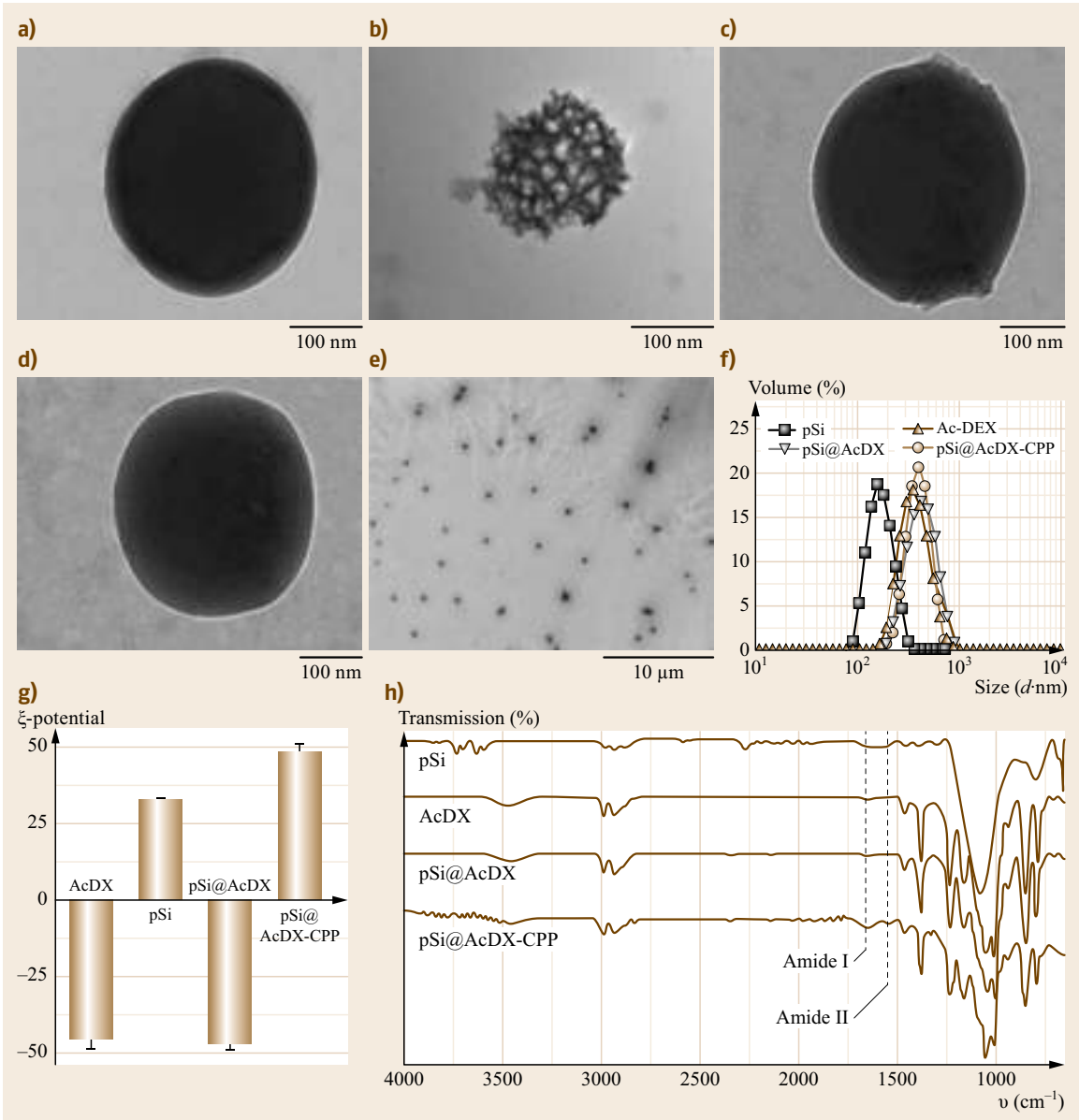


Fig. 21.17a–h TEM images of AcDX **(a)**, pSi **(b)**, pSi@AcDX **(c)** and **(d)** pSi@AcDX-CPP showing the encapsulation of a pSi nanoparticle within the AcDX nanoparticle. **(e)** Low-magnification TEM image and **(f)** volume-based size distribution of the highly homogeneous pSi@AcDX-CPP. **(g)** ζ-potential of the prepared nanosystems. **(h)** FTIR spectra of the nanocarriers confirming the encapsulation of pSi inside the AcDX matrix and the successful surface functionalization with CPP (after [21.81])

drugs into a single nanocarrier system, which can deliver a correct ratio of each drug to the target of interest, generating synergistic therapeutic effects, suppressing the drug resistance and severe adverse effects, as well as allowing controlled drug exposure over time. Therefore, by using a one-step microfluidic method, an advanced nanocomposite consisting of an encapsulated pSi nanoparticle and an acid-degradable polymeric matrix (acetalated dextran; AcDX), pSi@AcDX, was efficiently prepared [21.81] To further enhance the intracellular drug delivery, a nona-arginine cell-penetrating peptide (CPP) was conjugated onto the nanocomposite's surface to form the pSi@AcDX-CPP nanocomposite. The release of the payloads from the nanocomposites was

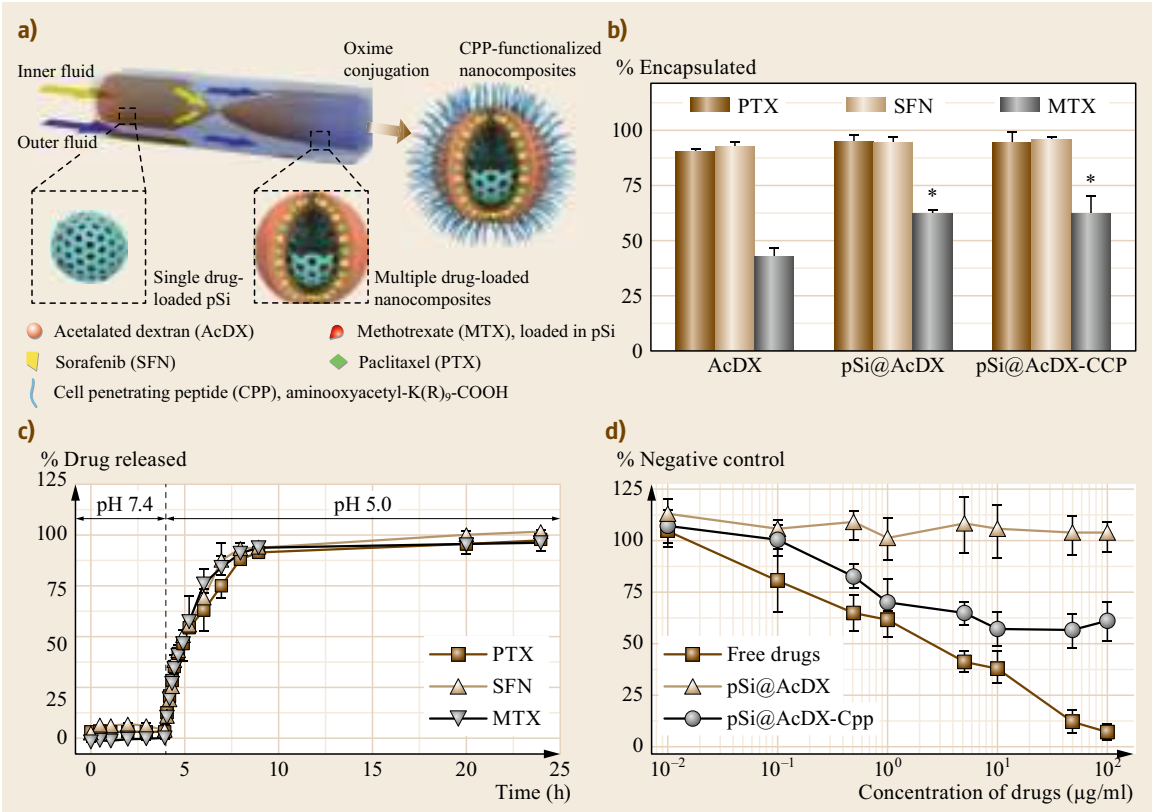


Fig. 21.18 **(a)** Illustration of the process for synthesizing CPP-functionalized multidrug-loaded pSi@AcDX-CPP (not to scale) using microfluidic coflow-focusing nanoprecipitation. **(b)** Encapsulation efficiency of PTX, SFN, and MTX. **(c)** Release profiles of PTX, SFN, and MTX from the pSi@AcDX-CPP at pH 7.4 and 5.0, and at 37 °C. **(d)** Growth inhibition of MDA-MB-231 cancer cells incubated with different nanocarriers and free drugs for 24 h at different concentration and at pH 7.4. The total concentration of PTX, SFN, and MTX (circa 1 : 1 : 1, w/w/w) ranged from 0.01 to 100 mg/ml. (after [21.81])

mainly controlled by the degradation rate of the polymeric matrix, which only took place when the pH was lower than 6.0 [21.219, 220]. Figure 21.17 shows an example of the transmission electron microscope (TEM) images, dynamic light scattering (DLS) for size, surface zeta (ζ)-potential for the charge, and Fourier transmission infrared (FTIR) of the prepared nanocomposites.

The above-mentioned nanocomposite was further loaded with different drug molecules for cancer treatment (Fig. 21.18a). Figure 21.18b shows the encapsulation efficiency for the simultaneously loaded PTX, SFN, and methotrexate (MTX) drugs inside the produced nanocomposites. The release studies of the three drugs were conducted at both single pH conditions (pH 7.4 and 5.0), which demonstrated that all the three drugs were trapped inside the pSi@AcDX-CPP (no drug release at pH 7.4), whereas regardless of the solubility of the loaded drugs, they all showed similar release profiles at pH 5.0 (Fig. 21.18c), precisely controlled by the hydrolysis of the AcDX matrix. Finally, the effect of the multidrug-loaded pSi@AcDX-CPP on the proliferation of both MCF-7 and MDA-MB-231 breast cancer cells was evaluated (Fig. 21.18d) [21.81]. Clearly, the combination of free PTX, SFN and MTX remarkably reduced the cell growth in a concentration-dependent manner. After CPP conjugation, the drug-loaded nanocomposite significantly controlled the cell proliferation inhibition, because of the pH-triggered decomposition of the AcDX matrix in the endolysosomal compartments, unlike the bare drug combinations that were indiscriminately more toxic to the cells than the functionalized nanocomposites.

21.5.4 Summary

In this section, a brief overview on the microfluidics techniques used to prepare different nanosystems was presented and discussed. Overall, promising and ad-

vanced multicomposite platforms for controlled drug delivery of multicargoes can be prepared by the microfluidics technique, which is already revolutionizing the drug delivery field. The great advantages of microfluidics are the simplicity and reproducibility of device fabrication, the capability to synthesize nanoparticles with narrower size distributions (< 1% size variation), improved batch-to-batch reproducibility, and higher drug loadings of different cargoes with a theoretical efficiency of 100%.

21.6 Electroless Formation of Metal Nanoparticle Aggregates

Metal nanoparticles are at the backbone of disciplines and techniques like optics, spectrometry and medicine because of their unique optical properties. These properties have been known since ancient times despite the physical/operational principle behind them being obscure. It is sufficient to evoke the art of decoration in Romans (forth century AD), of which a striking example is represented by the Lycurgus cup, as well as by numerous vessels and windows made of decorated glass. Roman glass workers discovered, in fact, that the addition of precious metals (such as gold or silver) during the processing of the molten glass allowed them to obtain a product that changed color depending on the light conditions. In particular, they used salts of gold whose ions, under the effect of the heat of the molten glass, were reduced to metallic nanoparticles. After cooling, the precious nanoparticles remained trapped in the matrix of silicon oxide (glass), creating a decorated glass, now known as *dichroic* glass, characterized by light-dependent colors both in reflection and in transmission. Moreover, in Italy, during 15th and 16th centuries, silver and gold salts were mixed in a paste of clay and vinegar and applied to vessels to obtain, after cooking, the reduction of metal ions to metallic nanoparticles and decorations with wonderful dichroic effects [21.221]. What is the *magic* process that transforms metal salt into particles with such color effects? The described examples are the first registered examples of nanotechnology applications: in particular, examples of a process to generate metal nanoparticles that in modern science is known as *electroless* deposition. This process consists of the chemical reduction of metal ions into metal nanoparticles in aqueous solution: it is the elected method for the metal plating of materials especially in electronic products. The reduction of the metal ions occurs by adding a reducing agent to the solution: metal ions attract the necessary electrons for their reduction from the reaction or oxidation of a reducing agent. The presence of a catalyzer accelerates the process, but sometimes it is the substrate itself that works as the catalyzer on which metal ions and the reducing agent are adsorbed: in this condition the process is an autocatalytic process. Moreover, the reaction does not need electrical current to proceed: it is a spontaneous process. The generic reaction between the metal ions Me^{n+} and the reducing agent Red is

$$m\,Me^{n+} + n\,Red = m\,Me^{0} + n\,Ox\,, \tag{21.1}$$

where n are the electrons necessary to produce the atomic metal Me^0. The driving force of the process is the potential differences of the two redox half-reactions. The Nernst equation gives the relation between the redox potential and the equilibrium constant K_{eq}

$$E_{eq} = E_0 + \frac{RT}{nF}\ln K_{eq}\,, \tag{21.2}$$

where an equilibrium/quasi-equilibrium state of the system is assumed and E_{eq} is the equilibrium electrode potential, E_0 is the standard reduction potential, $R = 8.314\,JK^{-1}\,mol^{-1}$ is the absolute gas constant, T is solution temperature, n is the number of electrons involved in the reaction, and $F = 96\,485\,C\,mol^{-1}$ is the Faraday constant. Moreover, it is known that K_{eq} can be expressed as

$$K_{eq} = \frac{k^+}{k^-} = \frac{\prod_{j=1}^{N}[a(C_j)]^{\nu_j^+}}{\prod_{j=1}^{N}[a(C_j)]^{\nu_j^-}}\,, \tag{21.3}$$

where k^+, k^- are the kinetics constant of the direct and reverse reactions, ν^+, ν^- are the stoichiometric coefficients of the direct and reverse reactions, $a(C_j)$ is the activity of the j-th chemical component in the electrolytic solution, and N is the total number of chemical components. In the specific case of dilute solutions, the equilibrium constant can be rewritten as

$$K_{eq} = \frac{\prod_{j=1}^{N}[Ox]_j^{\nu_j^+}}{\prod_{j=1}^{N}[Red]_j^{\nu_j^-}}\,, \tag{21.4}$$

where $[Ox]_j$ is the molar concentration of the j-oxidized species (the reagents of the redox reaction) and $[Red]_j$ is the molar concentration of the j-reduced species (the products of the redox reaction) [21.222, 223]. The equations above show how the thermodynamic and the kinetics of the electroless process are strongly connected

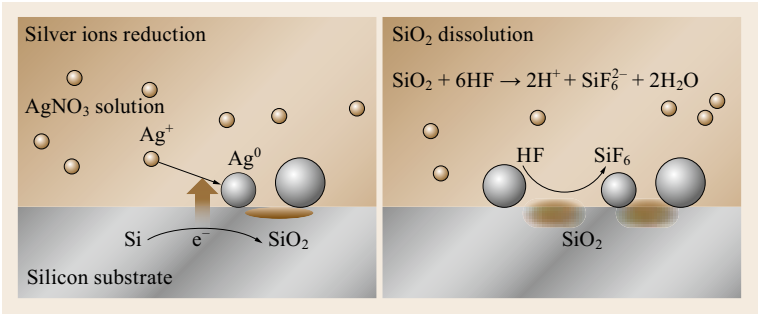


Fig. 21.19 Schematic of electroless deposition of silver on silicon

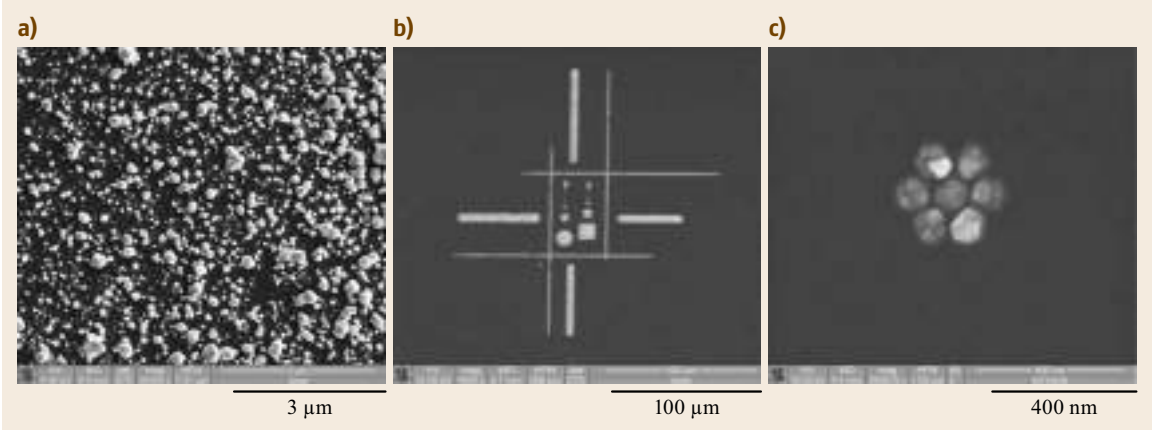


Fig. 21.20a–c Silver nanoparticles are deposited on silicon substrates. The final shape of the structures depends on the pattern obtained through electron beam lithography methods (**b,c**); in the absence of a lithographed pattern, nanoparticles are generated on the substrate with a random distribution (**a**)

and that the total reaction is influenced by concentrations in solution, temperature, time of reaction and type of metals. A particular case is the electroless deposition of metals on a substrate of a semiconductor like silicon. Metal ions exchange electrons with silicon itself which is, at the same time, the reducing agent and catalyzer. The reaction's solution consists of a 0.15 M HF aqueous solution and a metal salt (as auric chloride ($AuCl_3$) or silver nitrate ($AgNO_3$)) dissolved in an opportune concentration. The metal salt concentration and the process conditions (temperature and reaction time) must be adjusted according to the size of silicon samples in solution and the desired metal nanospheres' diameter. The function of hydrofluoric acid is to dissolve silicon oxide resulting from the reaction, making useful new silicon atoms from the substrate, as in Fig. 21.19.

For the particular case of silver nanoparticles, the reactions are:

At the anode (silicon oxidation)

$$\mathrm{Si} + 2\mathrm{H_2O} \rightarrow \mathrm{SiO_2} + 4\mathrm{H^+} + 4\mathrm{e^-}\,. \tag{21.5}$$

At the cathode (silver reduction)

$$\mathrm{Ag^+} + \mathrm{e^-} \rightarrow \mathrm{Ag^0}\,, \tag{21.6}$$

where redox potentials are $E_{01} = -0.86\,\mathrm{V}$, $E_{02} = 0.8\,\mathrm{V}$ [21.224]. Similarly, the reactions for gold nanoparticles are:

At the anode (silicon oxidation)

$$\mathrm{Si} + 2\mathrm{H_2O} \rightarrow \mathrm{SiO_2} + 4\mathrm{H^+} + 4\mathrm{e^-}\,. \tag{21.7}$$

At the cathode (gold reduction)

$$\mathrm{Au^{3+}} + 3\mathrm{e^-} \rightarrow \mathrm{Au^0}\,. \tag{21.8}$$

For this configuration, the standard redox potentials for the two half-reactions are $E_{03} = -0.86\,\mathrm{V}$, $E_{04} = 1.52\,\mathrm{V}$ [21.225, 226]. Both silver and gold react with a similar mechanism, which involves a nucleation phase with the formation of metallic nuclei and then a phase of grain growth. The kinetics of reaction are different for the two metals and in particular it is slower for gold with respect to silver. Moreover, gold forms a greater number of particles with respect to silver, obtaining smaller grains. The mechanism is not very well understood, but it is speculated that the reason behind a similar behavior might reside in the higher potential value of gold with respect to silver, which promotes the nucleation

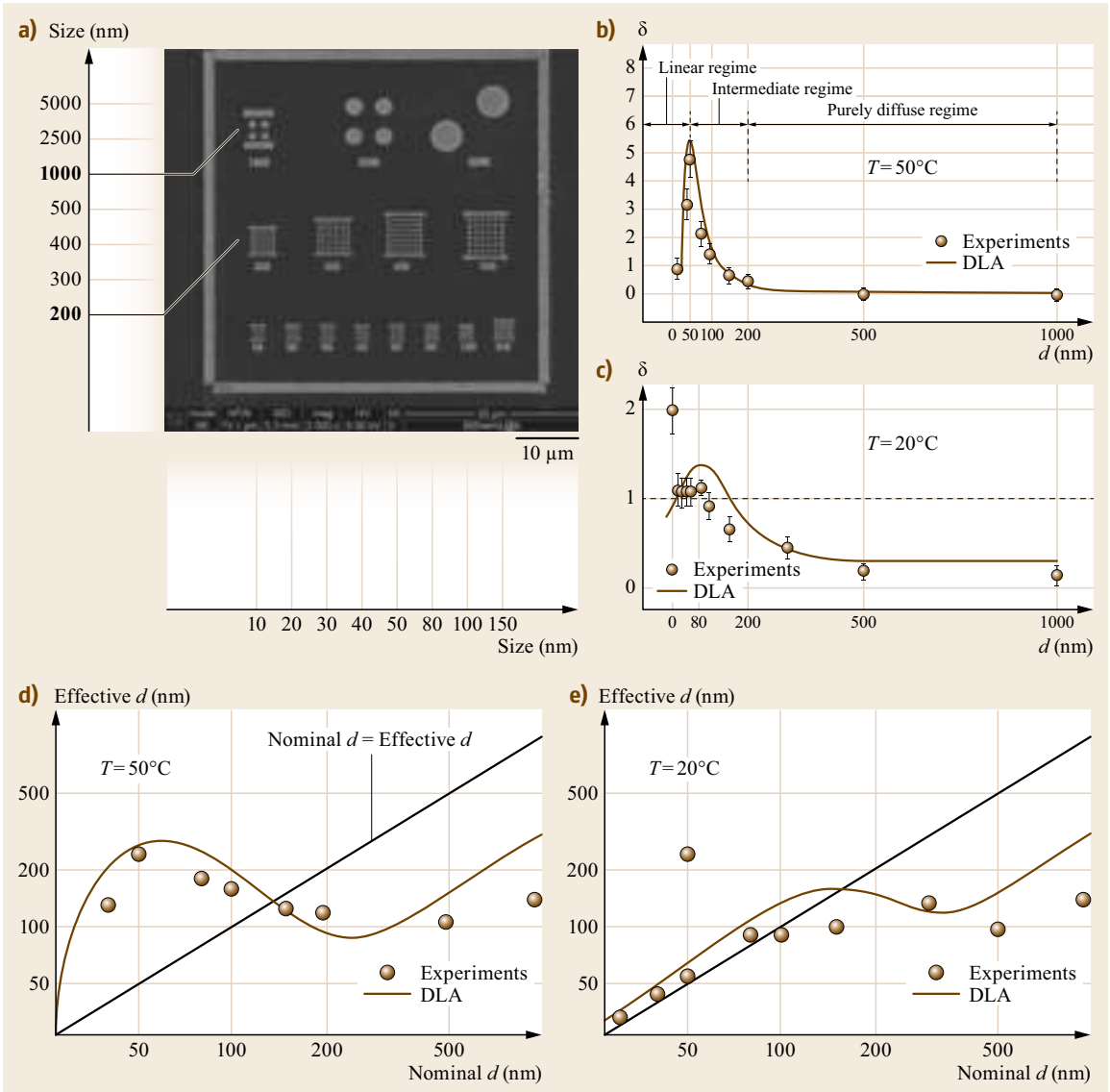


Fig. 21.21a–e In a number of experimental growths in which pattern size was varied over a significant range **(a)**, it was found that silver nanoparticle exhibit the maximum density for a 50 nm pattern size at $T = 50\,^{\circ}\mathrm{C}$ **(b)**, and for a 80 nm pattern size at $T = 20\,^{\circ}\mathrm{C}$ **(c)**. At each temperature, this meant the deviations from nominal and real particle size for different pattern sizes could be found **(d,e)**. A pattern is a template from which particles are obtained after electroless deposition (after [21.107])

of more particles and unbalances the equilibrium of the reactions.

Recent developments of nanofabrication techniques, such as electron beam lithography (EBL) or focused ion beam (FIB), have improved control over the size and shape of nanopatterns. The electroless technique is a simple method to fill those patterns and obtain metal nanoparticles with the desired characteristics. Consider a silicon substrate patterned with a microscale (in the case of optical lithography) or nanoscale (in the case of electron beam lithography) motif. On putting a similar sample in a solution of metal salt, a redox reaction is initialized that is typical of the electroless process: the reaction yields the formation of metal nanograins in the patterned portions of the silicon substrate (holes) exposed to the solution (Fig. 21.20). In Fig. 21.20, silver nanoparticles are deposited on silicon substrates. The final shape of the structures de-

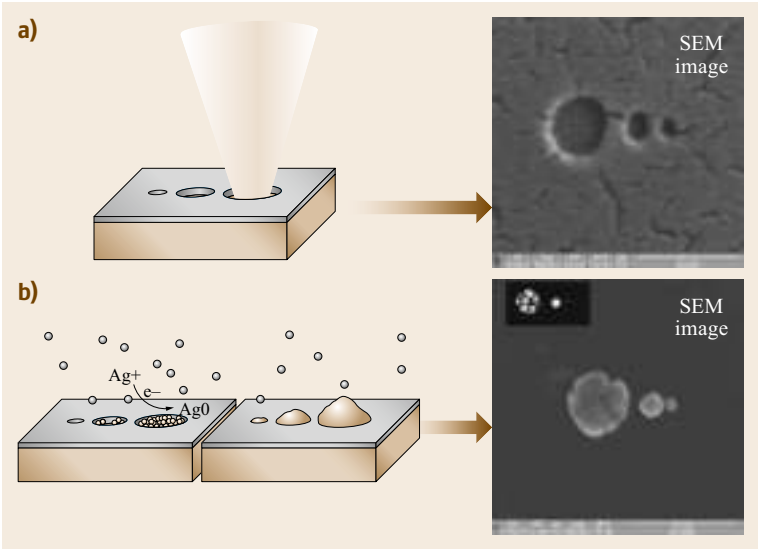


Fig. 21.22a–b Nanolenses are arrays of three silver spheres on a plane in which the size and distance between the spheres progressively reduces. Nanolenses can be obtained by integrating an electron beam lithography technique for the formation of a pattern on a silicon substrate with nanometric resolution **(a)** with a process of electroless deposition **(b)**, which loads the pattern to obtain the final nanoparticles

pends on the pattern obtained through electron beam lithography methods (Fig. 21.20b,c); in the absence of a lithographed pattern, nanoparticles are generated on the substrate with a random distribution (Fig. 21.20a).

Coluccio et al. [21.107] studied the combined effects of the geometry and the growth conditions on size and distribution of the silver nanoparticles. They demonstrated a very high sensitivity of the roughness and compactness of particle clusters to the variables of the process, including pattern size and pattern distance. They found a limiting hole diameter below which metal growth follows a linear regime for which isolated smooth spheres are obtained. Above this limit the structures are discontinuous (Fig. 21.21). This is an important observation because it introduces a criterion of optimal design for the fabrication of nanoparticles in nanophotonics, where tight control over the size and shape of the particles is required.

Thus, by combining electron beam lithography for pattern formation at the nanoscale (that is a top-down approach) with electroless deposition (that is a bottom-up approach), one may obtain nanoparticles with a precise control over the geometry of the particles. The major application of a similar technology is in Raman spectroscopy and surface-enhanced Raman spectroscopy [21.227]. On interacting with an external electromagnetic radiation field, metal nanoparticles sustain localized surface plasmon (LSP) resonance, which is collective oscillation of electrons at the metal–air interface. Similar oscillations are generated in the limit of vanishingly small particle diameters with respect to the EM wavelength, that is when $d \ll \lambda$. Over the particles' surface is thus observed the formation of polarization charges and a consequent dipolar field, which increases the electromagnetic field in close proximity to the particles. In the case of Raman spectroscopy, this phenomenon translates into a giant amplification of the Raman signal of the molecules in close proximity to the metal structures [21.18–20].

Using the described technique, nanostructures were realized to obtain giant SERS effects and probe biological samples with heretofore unattainable resolution. In [21.228], self-similar chains composed of three nanospheres were designed and fabricated, where the diameter of the spheres ranged from 20 to 180 nm and the gap between the smaller spheres was below 10 nm (Fig. 21.22). A similar design guarantees that the EM field in the gap is increased up to one thousand times, and the Raman signal, which goes with the power of four of the EM field, up to the incredible figure of 10^{12} times. A similar increment allows the detection of individual molecules in the gap; notice also that since the gap has a size of few nanometers, it functions like a zoom, or a filter, enhancing the signal of a few molecules against an interfering background. A similar nanolens was arrayed in a rectangular grid to obtain repeated units of the nanolens (Fig. 21.23). In contrast to an individual nanolens, an array of nanolenses may be exploited for the analysis of biological samples in a statistical approach. Notice that in this interpretation, a nanolens in the array is a pixel that carries single measurements. Using this strategy, *Coluccio* et al. [21.228] recognized single-point mutation in a subset of peptides extracted from the BRCA1 protein (a protein responsible for breast cancer) with a sensitivity in the picomolar (10^{-12} M) range. The same strategy was used to detect DNA oligonucleotides with six DNA bases (Fig. 21.24) [21.228]. Interestingly,

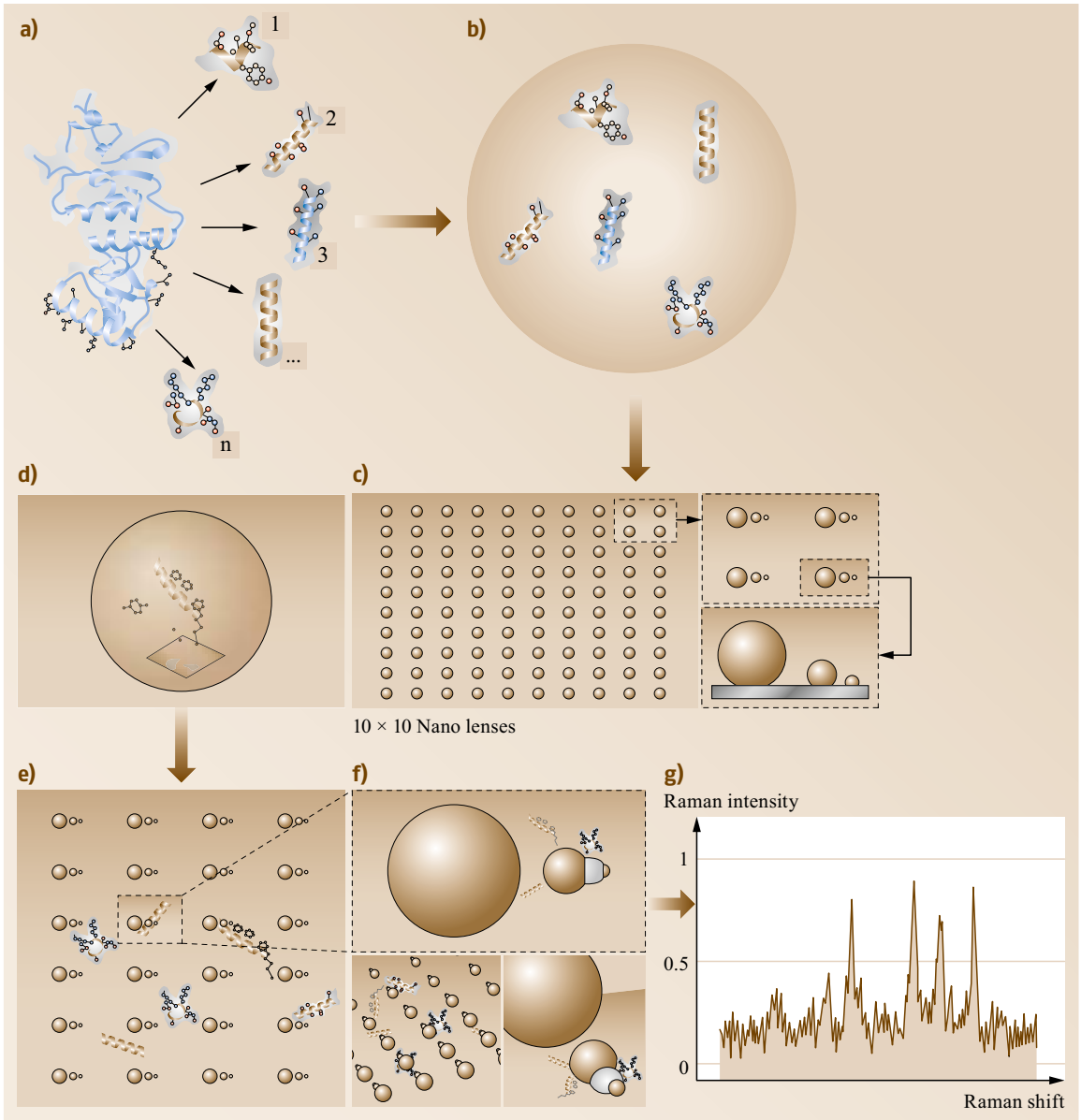


Fig. 21.23a–g Nanolenses were packed into an array to verify single mutations in peptides extracted from the BRCA1 protein **(a)**. The peptides mixture was extracted **(b)** and deposited on a nanolense matrix **(c)**. Due to the small size of each nanolense hot spot, only short fragments could fit into those spots **(e,f)** and be analyzed by SERS **(g)** (after [21.228])

oligonucleotides with nine bases could be detected only using a fitting procedure, in that nine bases did not entirely fit in the hot spots of the device.

21.6.1 DLA Methods Can Predict Electroless Formation of Metal Nanoparticles

In the described method, metal ions stick together to form the intended structures. Under certain experimental conditions, nanoparticles exhibit a dendritic structure (Fig. 21.25) with a recursive motif, which may be described using a fractal formalism.

Fractals are generally observed in nonequilibrium growth processes, where complex structures are obtained using short algorithms. This suggests that the electroless deposition of nanoparticles may be reproduced using mathematical modeling and computer simulations. In a recently proposed approach [21.229],

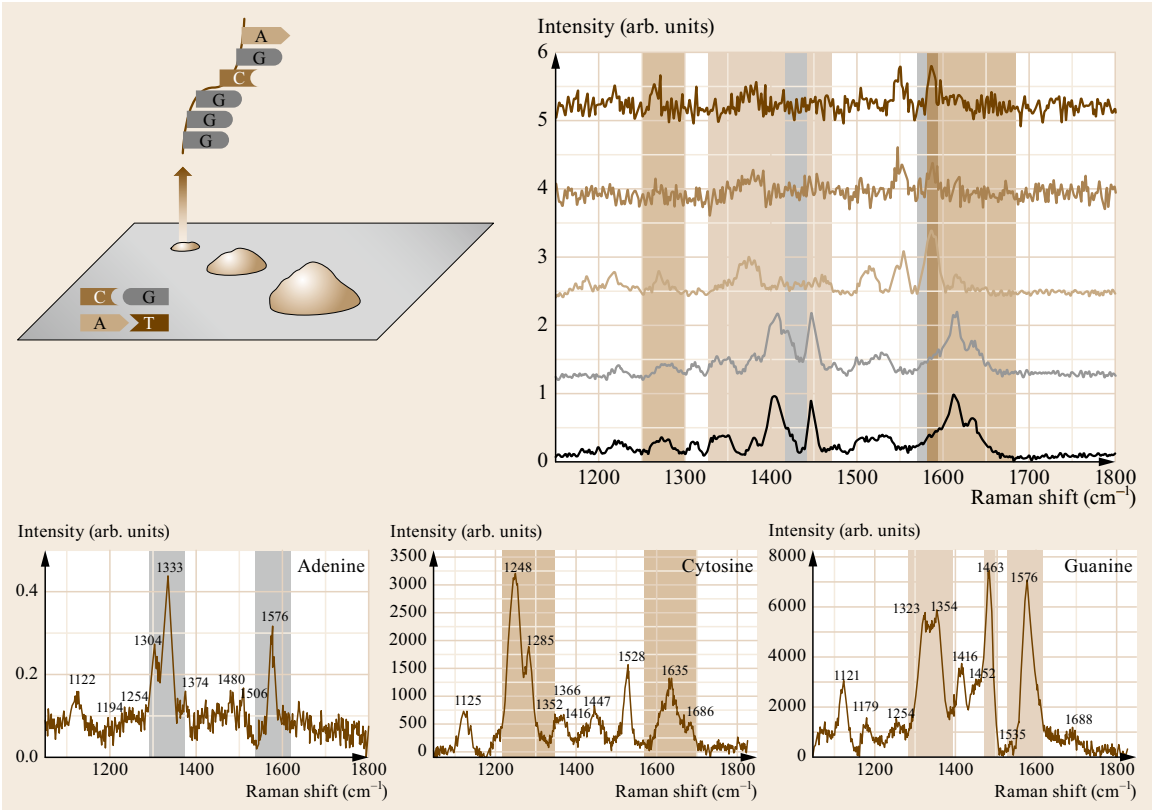


Fig. 21.24 Nanolenses arrays and Raman spectroscopy were used to detect DNA bases in short (six) and long (nine) oligonucleotide sequences

diffusion-limited aggregation (DLA) models were used to decipher the mechanisms of particle growth at the nanoscale. DLA is a simulation method based on simple algorithms in contrast to models that instead make use of analytical solution of equations [21.230]. In many situations it is possible to model a system by entities that diffuse and occasionally react upon encountering each other. These systems are governed by two time scales: the diffusion time, which is the characteristic time required for two particles to meet, and the reaction time, which is the characteristic time for particles to react when held in proximity (within the reaction range) to each other. When the diffusion time dominates over the reaction time, the process is limited by diffusion. Electroless aggregates can be correctly described by DLA models because in electroless deposition diffusion is the main driving force that regulates the dynamics of aggregation [21.229, 231, 232]. The model makes it possible to decipher, at an atomic level, the rules governing the evolution of the growth front, and to explore ways to tailor film morphology to obtain specific characteristics. DLA was formulated for the first time in the pioneering works of *Witten* and *Sander* in the early 1980s [21.233, 234], where the rules of aggregation of small building blocks into supramolecular structures was introduced, as are recapitulated below.

The Model

In what follows, we will describe the DLA model in the particular case of electroless nanoparticle growth on patterned silicon substrates, generated for example through electron beam lithography techniques, as in [21.107, 228]. In doing so, the domain where ions are represented is divided into two regions: the silicon substrate exposed to growth, and the inert resist surface, which defines the boundaries of deposition. In the model, the displacement of a metal ion at any time is arbitrary, and thus the trajectory of the particles can be correctly described by a random walk, resembling a Brownian motion (Fig. 21.26a).

Brownian motion is a continuous-time probabilistic process, and its use is motivated by mathematical convenience and the accuracy of the model to describe an otherwise complex phenomenon. At very short timescales, the motion of a particle undergoing Brownian motion is dominated by its inertia, its dis-

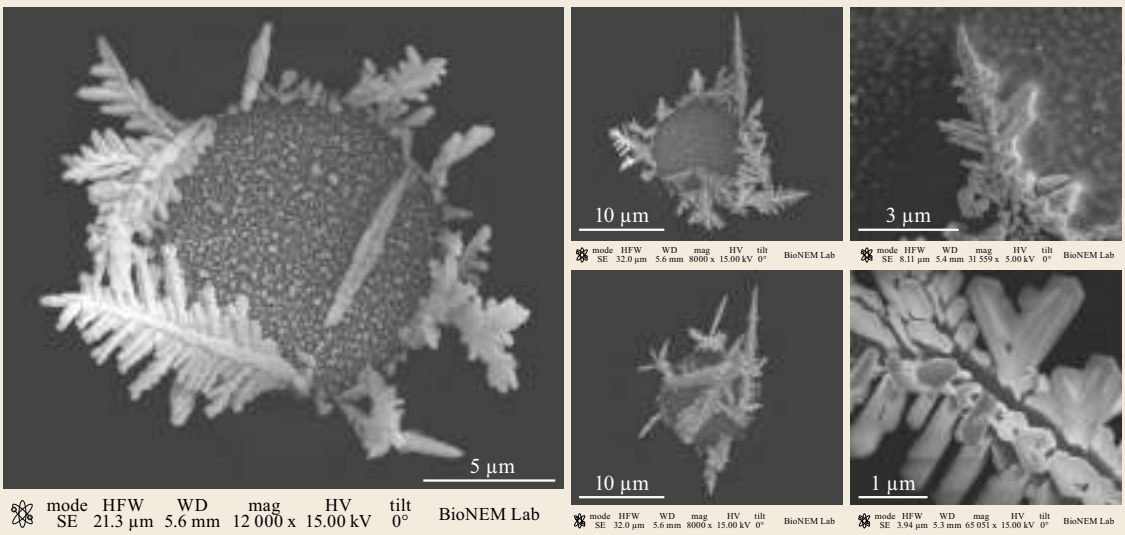


Fig. 21.25 Under certain conditions of electroless deposition, final structures reveal a dendritic fractal structure

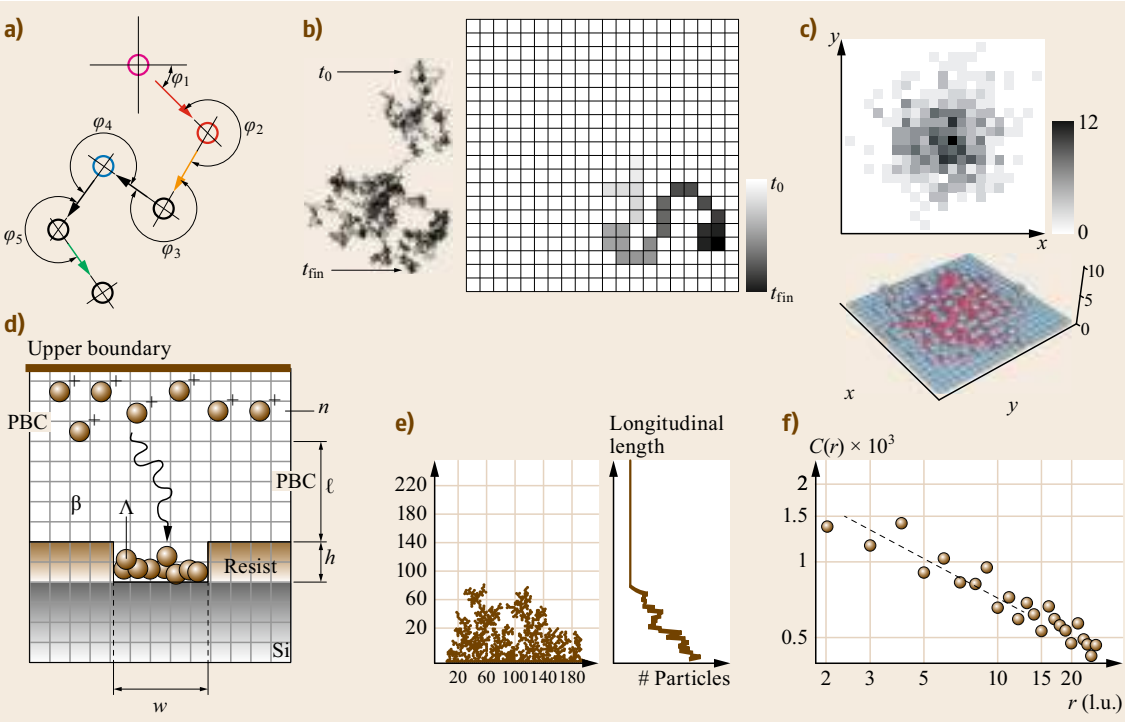


Fig. 21.26a–f Electroless deposition can be simulated using DLA methods. The displacement of individual ions in solution is described by a random Brownian motion **(a)**. The displacement can be reproduced in a grid where displacement takes discrete steps **(b)**. Brownian motion is described by statistics of the second order **(c)**. The entire system can be reproduced in a grid using this concept **(d)**. The simulated aggregate **(e)** has a fractal nature and its fractal dimension can be derived from power spectrum density methods **(f)**. For the present configuration, the parameters of the simulations were adjusted as follows. The simulation was halted after a number of iterations Ni=O(10^{10}). The total number of particles N, and the length of each seed were held fixed, being $N = 2000$, length $(\xi) = 25$ (l.u.)

placement will be linearly dependent on time, $\Delta x = v\tau$, and this can be reproduced or simulated in a regular grid (Fig. 21.26b) where particles are dislodged by the finite distance Δx in the time interval τ, that is the mean time between collisions. The instantaneous velocity of the particles, v, is maintained constant during τ, and it would depend solely upon the energy of the system. The distribution of displacement probability of the Brownian particle itself is best described using a Gaussian density function, centered around the origin of the walk, that means that the most probable position of a particle after a sufficiently large number of steps is its initial position (Fig. 21.26c). The root mean square distance of the walk gives a measure of the extent of spread of the particle ensemble, and this would be proportional to time [21.235]

$$\langle r^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle = 4Dt \,, \tag{21.9}$$

where x, y, and r are the Cartesian coordinates in the plane, while the celebrated Stokes–Einstein equation may be used to derive the diffusion coefficient D [21.236]

$$D = \frac{k_\mathrm{B}T}{6\pi\mu a} \,, \tag{21.10}$$

where k_B is the Boltzmann constant, T the absolute temperature of the system, μ the viscosity of the medium, and a is the diameter of a particle with mass m. D can be alternatively expressed in terms of Δx and τ as

$$D = \frac{\Delta x^2}{2\tau} = \frac{(v\tau)^2}{2\tau} = \frac{k_\mathrm{B}T}{2m} \,, \tag{21.11}$$

where the kinetic theory of gases has been used to obtain the right-hand term of (21.11) [21.235]. Notice that, on comparing (21.9) and (21.10), given the results of the simulations, the behavior of the real system would be determined.

This concept may be used to simulate the deposition of metal ions in a patterned silicon substrate. Consider, for ease of visualization, the scheme in Fig. 21.26d. The well defines the region of the domain exposed to growth, i. e., the patterned silicon substrate (for applications in which pattern distance cannot be neglected, the wells can be more than one). At a distance l from the well, N particles are released in the system, where N should be chosen with care to reproduce the initial concentration of silver ions. At any cycle the particles move within a regular square pattern of cells by one lattice unit (l.u.), and thus the mean path length is $\Delta x = 1$ l.u. At the sides (that is, the left and the right) of the domain, periodic boundary conditions (PBC) are imposed, and when an individual particle passes through one of those faces, it instantaneously reappears on the opposite face with the same velocity. In topological terms, the plane can be thought of as being mapped onto a torus; by doing so, an infinite perfect tiling of the system is simulated in the longitudinal direction, and this accounts for the fact that the area of the wells is small compared to the area above the wells. At the upper boundary of the system, and at the side walls of the well, a bouncing condition is imposed, whereby the particles that collide with those boundaries would rebound in an opposite direction. At the lower border of the system, the seeds ξ represent the regions of fresh silicon substrate exposed to growth. The length of the seed (that is, the length of the well) can be conveniently adjusted according to the problem at study. Therefore, ξ is a line of nucleation sites. The structure of the aggregate $\beta(t)$ will depend upon the length of the seed and the chemical conditions of growth. When a particle migrates to touch ξ it stops, is incorporated by ξ, and the size of the aggregate is augmented by one unit. At the same time, a new particle is instantaneously created and randomly positioned in the upper region of the domain to keep the total number of particles n constant during the process. After a certain number of repetitions, aggregates are formed as in the example of Fig. 21.25e. The multibranched arrangement of particles recalls the dendritic, fractal nature that electroless grown systems reveal under certain growth conditions.

The structure of clusters of occupied lattice sites exhibits geometric scaling relationships, which are characteristic of fractals and which can be used to estimate an effective fractal dimensionality. In similar aggregates, the fractal dimension D_f exhibits a limiting value of about $5/3$. The fractal dimension is a parameter that can be used to describe intimately the topography of a variety of systems, especially at the nanoscale. The importance of D_f resides in the fact that it is used in the definition of certain parameters or properties that describe the deposit.

For instance, as the thickness T of the aggregate scales with the total number of deposited particles n, as [21.237]

$$T \approx n^{1/(D_\mathrm{f}-1)} \tag{21.12}$$

the fractal dimension is also correlated to the distribution or number I of clusters with size S as a function of cluster size [21.237]

$$I(S) \approx S^{-(1+1/D_\mathrm{f})} \,. \tag{21.13}$$

Most importantly, the mean cluster size can be deduced and correlated to n via a very simple power law, being [21.238]

$$S \approx n^{D_\mathrm{f}/(D_\mathrm{f}-1)} \tag{21.14}$$

and thus, given n, which can be easily calculated, the mean cluster size can be readily derived. The packing or density factor κ can therefore be obtained as the ratio of S to the size of a single pattern, $\kappa = S/d$. Thus κ is a measure of the compactness of the aggregates. It is used to verify the influence of process parameters on the aggregate/particle quality. Larger κ indicates compact nanoparticles similar to their nominal shape.

The fractal dimension of numerical aggregates can be derived from their power spectrum density function $C(r)$; Fig. 21.26f. $C(r)$ delivers the information content of a cluster as a function of the characteristic length scale r. In a log–log diagram, $C(r)$ has a linear behavior. In such a diagram, if the slope of $C(r)$ is high, the information content decays rapidly with r, and conversely the information content ranges over wider length scales if the slope is low. From the slope, one may derive the fractal dimension as $D_f = (8-B)/2$. The mathematical procedure necessary to find $C(r)$ goes beyond the purpose of this section and can be found in [21.239].

Application of DLA Growth in Silver Nanoparticle Growth

In [21.229], a diffusion-limited aggregation approach and Fourier analysis were utilized to describe the nature of electroless deposits. In aggregates in which the size of the patterns was varied over a significant range, it was found the best pattern size for electroless silver particle growth is 50 nm at T=25 °C and 80 nm at T=50 °C. In [21.240], in simulations in which pattern distance was varied, it was found that the distance between patterns influences growth and that, when pattern distance is equal to or greater than three times the pattern size, particle quality is maximized. This suggests that in producing silver nanoparticles with electroless methods, one may use design maps akin to state diagrams in their graphical representation, in which on varying both pattern size and pattern distance in the diagram, for fixed chemical conditions and temperature, one may understand the effects of the process parameters on particle quality (κ), and choose the best combination of pattern size and pattern distance (input) for obtaining nanoparticles with desired characteristics (output).

21.7 2-D and 3-D Optical Nanostructures

In the majority of severe diseases, the detection of traces of biological species from the human body (blood and its components, saliva, biopsies, etc.) is of extreme importance for the possible impacts on human health. The early detection of several illnesses (e.g., cancer, neurodegenerative diseases or others) could anticipate the therapy of the same disease and arrest its development and progression. This would improve the life quality of the involved persons and additionally reduce the health care costs. That is why it is important to develop advanced screening and early diagnostic devices.

However, a great number of issues come into play before body fluids may be used in routine screening procedures and for early detection of severe illnesses [21.241]. In recent years, part of the scientific community has been moving towards the development of innovate bioanalytical tools that allow it to reach unprecedented resolution and to simplify the protocol of analysis, satisfying the aforementioned needs.

The handling and analysis of complex biological samples, and in particular individuating a single component among a cluster of molecules, is challenging and conventionally requires complicated protocols to pretreat complex samples [21.242]. Early diagnosis is often dependent on the detection of a few molecules in diluted solutions, which are invisible to the current sensors due to a limit on the resolution of analysis. Moreover, for certain screening procedures it is important not to affect the phenotype of the biological sample [21.243], especially for certain screening procedures that require analysis over long periods of a specific sample (e.g., a specific cell population). Thus it becomes important to introduce new methodologies that are not invasive with respect to the samples. Finally, the heterogeneity of behavior of the human body and its response to medical treatments requires the development of bioanalytical tools compatible with personalized medicine, and this is translated into the development of devices that are portable, fast, and provide high-throughput and high-content analysis with reduced costs. One way to overcome these issues is the development of microfluidic devices and innovative label-free plasmonic/spectroscopic nanosensors [21.244].

21.7.1 Integration of Optical Nanostructures in Microfluidic Devices

Microfluidics is an interdisciplinary discipline that focuses on the transport, manipulation and analysis of small amount of liquids, cells and particles [21.245]. Microfluidic devices have been developed and exploited by several research groups as bioanalytical tools [21.246]. This is due to several advantages over

conventional bioanalytical technologies. In fact, these devices guarantee high portability, accurate control for handling of samples, simplified sample pretreatment protocols, low consumption of samples and reagents, high resolution of analysis and low costs of production and analysis [21.247].

There are many examples in the literature of microfluidic devices used for analytical purposes. These have been developed for DNA detection [21.248], protein analysis [21.249], and detection of other biomarkers (e.g., discriminating healthy from tumor cells, phenotypes cell screening, etc.) [21.250].

A microfluidic device is capable of handling biological samples with very high accuracy and resolution, simplifying the protocol of analysis [21.251]. A particular and interesting case is the possibility to screen cell populations with single-cell resolution [21.252].

It is well known that individual cells, even those identical in appearance, differ in numerous characteristics. Due to this heterogeneity, traditional biochemical assays, which analyze cells in bulk, do not allow us to get rich information when single cells are studied. Single-cell analysis allows the investigation of cell activities and potentially new biomarkers. A sequential handling and manipulation of cells and cell suspensions then becomes important for scientific research and clinical diagnostic applications. Usually, cells are cultured and analyzed in large-scale environments such as Petri dishes. Therefore, it is difficult to analyze individual cells without the influence of other cells. These issues can be overcome by using microfluidics [21.253].

It is also crucial to analyze the composition of a single cell at a subcellular level or metabolites, and this would be a step forward in the conventional analysis of cells. Another critical issue is thus to individuate a sensing methodology that allows a resolution of analysis at the single-cell level [21.254]. Monitoring of cells over a longer period is most frequently performed by fluorescence microscopy. However, it is quite common that the fluorescent dyes used for specific subcellular staining interfere with the development of cells, changing their phenotype and metabolism. Another limitation is that fluorescent dyes bleach quite fast. Finally, this technique allows the investigation of known molecules only. These aspects make fluorescence staining strategies less desirable with respect to label-free sensing methodologies. This can be done, for instance, by integrating plasmonic nanostructured devices with Raman spectroscopy. Raman spectroscopy has been used to determine cellular status, such as living cells, dead cells, apoptotic cells, proliferating cells, differentiating cells, and tumor or healthy cells. Raman spectroscopy can be integrated into microfluidic devices, providing accurate control over biological samples (on single cells) and allowing samples to be left over long periods in physiological or conditioned environments.

The main advantages of such a technique are:

- The capability to perform analysis in a label-free manner, reducing the pretreatment steps for a biological sample
- The specificity of analysis due to a specific fingerprint for each biological substance
- The possibility to analyze biological substances in water solutions.

The main disadvantage is the small probability that a Raman scattering event can occur, and in some cases this could produce very weak signals, difficult to detect in complex biological microenvironments. Fortunately, plasmonic nanostructures can strongly improve the Raman effect, with enhancing factors up to 10^{12}, and producing the well-known SERS effect (surface enhanced Raman scattering). Then, the integration of plasmonic nanosensors into microfluidic devices will circumvent the issue of small Raman scattering cross-section, and will produce a valid tool for label-free biochemical analysis in microfluidic environments.

Plasmonic nanosensors are metal nanostructures in which surface plasmons are generated by an external exciting laser source, producing a large enhancement of the Raman signal when the plasmonic resonance conditions are matched. The microscopic origin of the SERS effect relies solely on a local enhancement due to surface metal (Au and Ag) nanoparticles. Obviously this effect is firmly determined by the presence of the metallic nanostructures or a roughened metal surface, and only when the probed molecules are close to them can the SERS effect arise. Finally, the high spatial resolution of the nanostructures allows the resolution of very complex mixtures, simplifying the pretreatment protocol of biological samples [21.228].

For example, a microfluidic device integrating microfluidic traps, able to handle single cells, coupled to a Raman spectrometer has high potential for the analysis of living cells. In this case the choice of the appropriate materials for the fabrication processes of each piece composing the microfluidic device becomes crucial for performing proper Raman measurements. In fact, these have to produce a negligible background signal during Raman analysis (e.g., calcium fluoride).

Single-cell analysis allows the study of drug stimulation and can distinguish the diversity of the individual cell activities, and potentially allows the investigation of new biomarkers. A sequential handling and manipulation of cells and cell suspensions then becomes important for scientific research and clinical diagnostic applications. The advent of microfabrication technology in the field of tissue engineering has enabled

new and innovative ways to analyze cellular interactions in bioreactors and cell behavior in a microfluidic environments that mimic in vivo body conditions. The microfluidic bioreactors or microbioreactors are microscale versions of conventional bioreactors with significant advantages in terms of cost-effectiveness, portability, performance, and materials used. Microfluidic devices can ensure accurate single-cell handling by simple microfluidic control with minimal consumption of reagents.

The exploitation of a plasmonic nanodevice coupled to a Raman spectrometer would allow (thanks to the localization of light in small structures and the resulting enhancement of the electromagnetic field) the study of single biomolecules with high spatial resolution, directly on the cellular membrane of a single cell.

The plasmonic nanodevices integrated into microfluidic devices can have different geometries and configurations. Two examples will be given in the next sections.

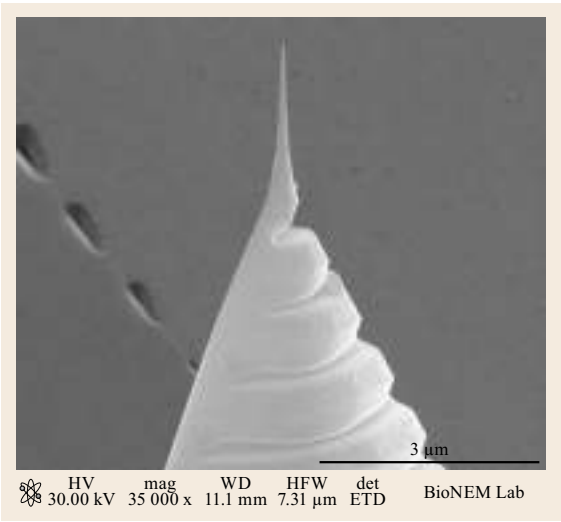


Fig. 21.27 SEM image of a nanoplasmonic device (after [21.249] with permission from Elsevier)

21.7.2 Three-Dimensional Optical Nanostructure

Plasmonic nanodevices can be made of a dielectric *core* with a conical geometry, covered by a metal layer as shown in Fig. 21.27.

These devices were grown using electron beam induced deposition from a gas precursor containing a platinum-carbon polymer ($(CH_3)_3Pt(CpCH3)$) and deposition of gold by sputtering. The devices had a base of about 400 nm, a height of 2.5 µm and a radius of curvature at the tip of about 5 nm

These devices are based on the amplification of the electromagnetic field due to the adiabatic compression of the wavelength of a plasmon polariton on the metal layer of a nanocone. In this way it is possible to obtain a resolution that depends on the radius of the tip of the cone (approximately 10 nm), which is a much lower resolution than the diffraction limit of light.

The outer interface between the gold layer and the surrounding dielectric medium constitutes a surface for the propagation of plasmon polaritons (PPs). The PPs are excited by the Raman laser source that is coupled to a grating realized at the base of the cone. Grating periodicity is adjusted in order to produce an optimal coupling with the 532 nm laser light of Raman microspectroscopy equipment. Once excited, PPs travel along the cone side and undergo adiabatic compression as described in [21.255]. Along with adiabatic compression, a large enhancement of the electromagnetic field is produced at the cone tip, thus producing an enhancement of the Raman signal of biomolecules close to the tip. During Raman measurements the objective focus is at the level of the cone tip in order to fully exploit the signal enhancement. Consequently the illumination of the grating is not optimal, but even in this condition a good coupling with the far-field light is achieved, as demonstrated in [21.255]. The sharpness of the tip cone, with a tip radius in the nm range, ensures high enhancement factors and high spatial resolution for Raman measurements.

These nanodevices can be integrated in a microfluidic system close to microfluidic traps where individual cells can be trapped and analyzed. The traps consist of microfluidic cavities placed within a microfluidic channel of dimensions a bit larger than a cell (about 20–25 µm). At the bottom of the cavity was fabricated a narrow channel (about 3 µm wide), which allowed the sucking in of the cells passing in the proximity of the cavity. When the cell reaches the bottom of the cavity, it plugs the narrow channel avoiding further suction. In this way, it is possible to trap cells in series. By alternating the direction of the motion, it is possible to displace the cell from the cavity and suck it again in the same cavity but with a different orientation, so that it is possible to analyze different parts of the membrane of the same cell (Fig. 21.28).

21.7.3 Two-Dimensional Optical Nanostructure

Alternative plasmonic nanodevices can be made by using a 2-D configuration such as nanodimers integrated in microfluidic traps, as shown in Fig. 21.29.

In this case, the trap is realized by fabricating a constriction in width along the channel. It has been

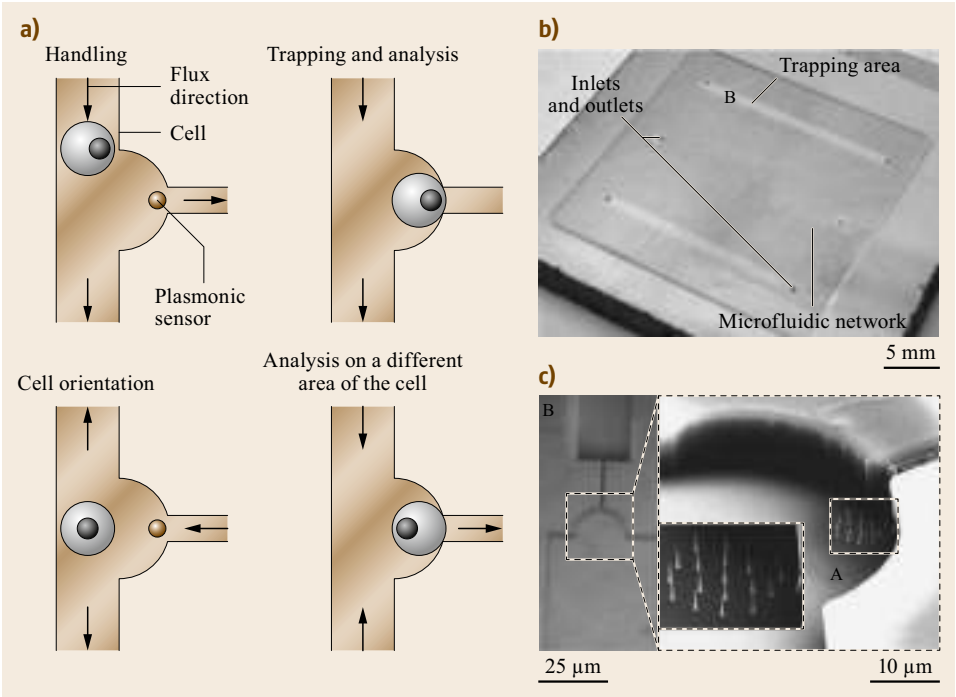


Fig. 21.28 **(a)** Schematic representation of the cell trapping and analysis protocol. The buffer movement allows trapping or releasing the cell. **(b)** Isometric view of the microfluidic device. **(c)** Optical microscope image of the microfluidic trap and SEM image of the microfluidic trap integrating the nanodevices (after [21.249] with permission from Elsevier)

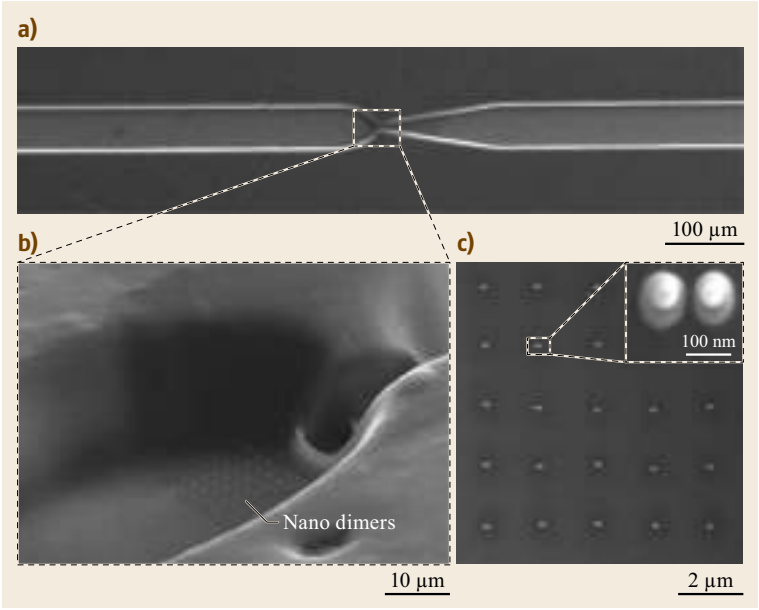


Fig. 21.29a–c Scanning electron image of the microfluidic trap and the integrated array of nanodimers. **(a)** Top view; **(b)** isometric view; **(c)** zoom in the area where the nanodimers are integrated (after [21.244], Optics Express)

designed in such a way that the cell is only temporarily trapped. A dynamic change of the flow conditions is responsible for cell entrapment. When a constant flow rate is applied a cell is moved toward the trap. When the cell is stopped at the constriction, the fluidic resistance increases accordingly and since the flow rate is constant, the pressure drop increases locally. The increase in pressure increases the forces acting on the cell surface and this provokes a cell deformation that allows the same cell to move through the trap, continuing its motion in the channel. After that, another cell approaches the trap. The role of the trap is to stop the cell for a well-defined amount of time necessary to record the Raman spectrum, which can be executed in about

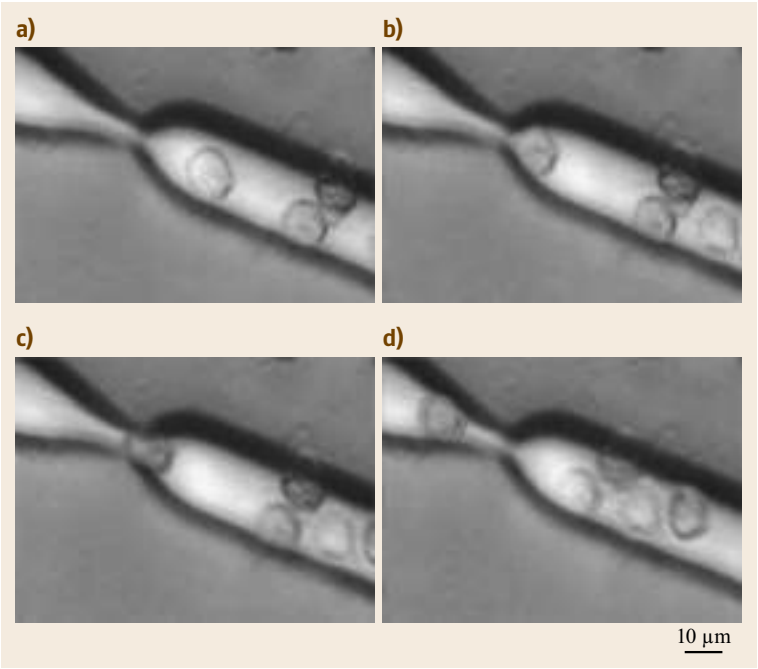


Fig. 21.30a–d Sequence of the cell analysis mechanism: **(a)** the cell approaches the trap followed by other cells ($t = 0$ s); **(b)** the cell is trapped and the flow is stopped ($t = 2$ s) – in this phase Raman measurements can be performed on the cell ($t = 4$ s); **(c)** the increasing pressure due to the clogged channel deforms the cell forcing it through the trap ($t = 6$ s); **(d)** the cell is released and continues to move over the trap while other cells approach the trap, aligning themselves to it ($t = 8$ s) (after [21.244] Optics Express 2016)

3 s. In such a way, it is possible to perform a continuous analysis of cells by Raman spectroscopy in a passive way (Fig. 21.30).

In this case the nanodimers can be fabricated by electron lithography and electroless deposition as already described in the other sections.

An advantage of such a configuration is that it is more compatible with mass-production fabrication protocols. Three-dimensional nanodevices are more complicated and expensive to fabricate but the freedom to work in three dimensions allows the exploitation of better nonlinear optical properties, obtaining more efficient plasmonic nanodevices.

Acknowledgments. Dr. Hélder A. Santos acknowledges the Academy of Finland (project numbers 252215 and 281300), the University of Helsinki and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement number 310892 for financial support. Dr. Dongfei Liu and Dr. Bárbara Herranz Blanco (Division of Pharmaceutical Chemistry and Technology) are also acknowledged for their discussions and for providing some of the material in the microfluidics section. Professor Francesco Gentile acknowledges the Italian Minister of Health (Project number GR-2010-2320665).

References

21.1 W.R.S. Sanhai, J.H. Sakamoto, R. Canady, M. Ferrari: Seven challenges for nanomedicine, Nat. Nanotechnol. **3**, 242–244 (2008)

21.2 O. Salata: Applications of nanoparticles in biology and medicine, J. Nanobiotechnol. (2004) doi:10.1186/1477-3155-2-3

21.3 K. Riehemann, S.W. Schneider, T.A. Luger, B. Godin, M. Ferrari, H. Fuchs: Nanomedicine–challenge and perspectives, Angewandte Chemie Int. Edn. **48**(5), 872–897 (2009)

21.4 R. Petros, J. DeSimone: Strategies in the design of nanoparticles for therapeutic applications, Nat. Rev. Drug Discov. **9**(8), 615–627 (2010)

21.5 M. Ferrari: Cancer nanotechnology: Opportunities and challenges, Nat. Rev. Cancer **5**, 161–171 (2005)

21.6 W.M. Haynes (Ed.): *CRC Handbook of Chemistry and Physics* (CRC, Boulder 1998)

21.7 Z. Zhang, M.A. Horsch, M.H. Lamm, S.C. Glotzer: Tethered nano building blocks: Toward a conceptual framework for nanoparticle self-assembly, Nano Lett. **3**(10), 1341–1346 (2003)

21.8 P.F. Damasceno, M. Engel, S.C. Glotzer: Predictive self-assembly of polyhedra into complex structures, Science **337**, 453 (2012)

21.9 X. Lu, M. Rycenga, S. Skrabalak, B. Wiley, Y. Xia: Chemical synthesis of novel plasmonic nanoparticles, Ann. Rev. Phys. Chem. **60**, 167–192 (2009)

21.10 M.G. Guzmán, J. Dille, S. Godet: Synthesis of silver nanoparticles by chemical reduction method and their antibacterial activity, Int. J. Chem. Biomol.

Eng. **2**(3), 104–111 (2009)
21.11 U. Mohanty: Electrodeposition: A versatile and inexpensive tool for the synthesis of nanoparticles, nanorods, nanowires, and nanoclusters of metals, J. Appl. Electrochem. **41**, 257–270 (2011)
21.12 N. German, A. Ramanavicius, A. Ramanavicien: Electrochemical deposition of gold nanoparticles on graphite rod for glucose biosensing, Sens. Actuators B **203**, 25–34 (2014)
21.13 J. Key, A.L. Palange, F. Gentile, S. Aryal, C. Stigliano, D. Di Mascolo, E. De Rosa, M. Cho, Y. Lee, J. Singh, P. Decuzzi: Soft discoidal polymeric nanoconstructs resist macrophage uptake and enhance vascular targeting in tumors, ACS Nano **9**(12), 11628–11641 (2015)
21.14 C. Daraio, S. Jin: Synthesis and patterning methods for nanostructures useful for biological applications. In: *Nanotechnology for Biology and Medicine at the Building Block Level*, ed. by G.A. Silva, V. Parpura (Springer, New York 2012)
21.15 T.D.H. Le, W. Bonani, G. Speranza, V. Sglavo, R. Ceccato, D. Maniglio, A. Motta, C. Migliaresi: Processing and characterization of diatom nanoparticles and microparticles as potential source of silicon for bone tissue engineering, Mater. Sci. Eng. C **59**, 471–479 (2016)
21.16 A. Jantschke, C. Fischer, R. Hensel, H.G. Braun, E. Brunner: Directed assembly of nanoparticles to isolated diatom valves using the non-wetting characteristics after pyrolysis, Nanoscale **6**, 11637–11645 (2014)
21.17 S. Schlücker: Surface-enhanced Raman spectroscopy: Concepts and chemical applications, Angew. Chem. Int. Edn. **53**(19), 4756–4795 (2014)
21.18 K. Kneipp, Y. Wang, H. Kneipp, L.T. Perelman, I. Itzkan, R.R. Dasari, M.S. Feld: Single molecule detection using surface-enhanced Raman scattering (SERS), Phys. Rev. Lett. **78**, 1667–1670 (1997)
21.19 K. Kneipp, H. Kneipp, I. Itzkan, R.R. Dasari, M.S. Feld: Surface-enhanced Raman scattering and biophysics, J. Phys. Condens. Matter **14**, R597–R624 (2002)
21.20 J. Kneipp, H. Kneipp, K. Kneipp: SERS-a single-molecule and nanoscale tool for bioanalytics, Chem. Soc. Rev. **37**(5), 1052–1060 (2008)
21.21 S.T. Hunt, M. Milina, A.C. Alba-Rubio, C.H. Hendon, J.A. Dumesic, Y. Román-Leshkov: Self-assembly of noble metal monolayers on transition metal carbide nanoparticle catalysts, Science **352**(6288), 974–978 (2016)
21.22 O. Govorov, H.H. Richardson: Generating heat with metal nanoparticles, Nano Today **2**, 30–38 (2007)
21.23 G. Baffou, R. Quidan: Thermo-plasmonics: Using metallic nanostructures as nano-sources of heat, Laser Photonics Rev. **7**(2), 171–187 (2013)
21.24 D. Maharaj, B. Bhushan: Scale effects of nanomechanical properties and deformation behavior of Au nanoparticle and thin film using depth sensing nanoindentation, J. Nanotechnol. **5**, 822–836 (2014)
21.25 D. Maharaj, B. Bhushan: Nanomechanical behavior of MoS_2 and WS_2 multi-walled nanotubes and Carbon nanohorns, Sci. Rep. **5**(8539), 1–9 (2015)
21.26 D. Chen, X. Qiao, X. Qiu, J. Chen: Synthesis and electrical properties of uniform silver nanoparticles for electronic applications, J. Mater. Sci. **44**(4), 1076 (2009)
21.27 A.H. Alshehri, M. Jakubowska, A. Młożniak, M. Horaczek, D. Rudka, C. Free, J.D. Carey: Enhanced electrical conductivity of silver nanoparticles for high frequency electronic applications, ACS Appl. Mater. Interfaces **4**(12), 7007–7010 (2012)
21.28 A. Akbarzadeh, M. Samiei, S. Davaran: Magnetic nanoparticles: Preparation, physical properties, and applications in biomedicine, Nanoscale Res. Lett. **7**(1), 144 (2012)
21.29 B. Nawrot, E. Gaggelli: Understanding the chemical mechanisms of life, Nat. Chem. Biol. **3**, 745–749 (2007)
21.30 G.M. Whitesides: The 'right' size in nanobiotechnology, Nat. Biotechnol. **21**, 1161–1165 (2003)
21.31 W.F. Marshall, K.D. Young, M. Swaffer, E. Wood, P. Nurse, A. Kimura, J. Frankel, J. Wallingford, V. Walbot, X. Qu, A.H. Roeder: What determines cell size?, BMC Biology **10**(101), 1–22 (2012) doi:10.1186/1741-7007-10-101
21.32 E. Hafen, H. Stocker: How are the sizes of cells, organs, and bodies controlled?, PLoS. Biol. **1**(3), e86 (2003)
21.33 Y. Fung: *Biomechanics, Mechanical Properties of Living Tissues* (Springer, New York 1993)
21.34 M. Elsabahy, K.L. Wooley: Design of polymeric nanoparticles for biomedical delivery applications, Chem. Soc. Rev. **41**, 2545–2561 (2012)
21.35 E. Tasciotti, X. Liu, R. Bhavane, K. Plant, A.D. Leonard, B.K. Price, M. Ming-Cheng Cheng, P. Decuzzi, J.M. Tour, F. Robertson, M. Ferrari: Mesoporous silicon particles as a multistage delivery system for imaging and therapeutic applications, Nat. Nanotechnol. **3**, 151–157 (2008)
21.36 E. Blanco, H. Shen, M. Ferrari: Principles of nanoparticle design for overcoming biological barriers to drug delivery, Nat. Biotechnol. **33**, 941–951 (2015)
21.37 A. Parodi, N. Quattrocchi, A.L. van de Ven, C. Chiappini, M. Evangelopoulos, J.O. Martinez, B.S. Brown, S.Z. Khaled, I.K. Yazdi, M.V. Enzo, L. Isenhart, M. Ferrari, E. Tasciotti: Synthetic nanoparticles functionalized with biomimetic leukocyte membranes possess cell-like functions, Nat. Nanotechnol. **8**, 61–68 (2013)
21.38 Z. Li, J.C. Barnes, A. Bosoy, J.F. Stoddart, J.I. Zink: Mesoporous silica nanoparticles in biomedical applications, Chem. Soc. Rev. **41**, 2590–2605 (2012)
21.39 J.S. Ananta, B. Godin, R. Sethi, L. Moriggi, X. Liu, R.E. Serda, R. Krishnamurthy, R. Muthupillai, R.D. Bolskar, L. Helm, M. Ferrari, L.J. Wilson, P. Decuzzi: Geometrical confinement of gadolinium-based contrast agents in nanoporous particles enhances T1 contrast, Nat. Nanotechnol. **5**, 815–821 (2010)

21.40 L. Dykmana, N. Khlebtsov: Gold nanoparticles in biomedical applications: Recent advances and perspectives, Chem. Soc. Rev. **41**, 2256–2282 (2012)

21.41 E.C. Dreaden, A.M. Alkilany, X. Huang, C.J. Murphy, M.A. El-Sayed: The golden age: Gold nanoparticles for biomedicine, Chem. Soc. Rev. **41**, 2740–2779 (2012)

21.42 Q.H. Tran, V.Q.A.-T. Le Nguyen: Silver nanoparticles: Synthesis, properties, toxicology, applications and perspectives, Adv. Nat. Sci. Nanosci. Nanotechnol. **4**, 033001 (2013)

21.43 S. Prabhu, E.K. Poulose: Silver nanoparticles: Mechanism of antimicrobial action, synthesis, medical applications, and toxicity effects, Int. Nano Lett. **2**(32), 1–10 (2012)

21.44 H.T. Nasrabadi, E. Abbasi, S. Davaran, M. Kouhi, A. Akbarzadeh: Bimetallic nanoparticles: Preparation, properties, and biomedical applications, Artif. Cells Nanomed. Biotechnol. **44**, 376–380 (2016)

21.45 C. Xu, S. Sun: New forms of superparamagnetic nanoparticles for biomedical applications, Adv. Drug Deliv. Rev. **65**, 732–743 (2013)

21.46 A.K. Gupta, M. Gupta: Synthesis and surface engineering of iron oxide nanoparticles for biomedical applications, Biomaterials **26**, 3995–4021 (2005)

21.47 J.H. Lee, J.-T. Jang, J.-S. Choi, S.H. Moon, S.-H. Noh, J.-W. Kim, J.-G. Kim, I.-S. Kim, K.I. Park, J. Cheon: Exchange-coupled magnetic nanoparticles for efficient heat induction, Nat. Nanotechnol. **6**, 418–422 (2011)

21.48 H.L. Ferrand, S. Bolisetty, A.F. Demirors, R. Libanori, A.R. Studart, R. Mezzenga: Magnetic assembly of transparent and conducting graphene-based functional composites, Nat. Commun. **7**, 12078 (2016)

21.49 L.Y. Chou, K. Zagorovsky, W.C. Chan: DNA assembly of nanoparticle superstructures for controlled biological delivery and elimination, Nat. Nanotechnol. **9**, 148–155 (2014)

21.50 F. Danhier, E. Ansorena, J.M. Silva, R. Coco, A.L. Breton, V. Préat: PLGA-based nanoparticles: An overview of biomedical applications, J. Control. Release **161**, 505–522 (2012)

21.51 D. Peer, J.M. Karp, S. Hong, O.C. Farokhzad, R. Margalit, R. Langer: Nanocarriers as an emerging platform for cancer therapy, Nat. Nanotechnol. **2**(12), 751–760 (2007) doi:10.1038/nnano.2007.387

21.52 C. Antoniades, C. Psarros, D. Tousoulis, C. Bakogiannis, C. Shirodaria, C. Stefanadis: Nanoparticles: A promising therapeutic approach in atherosclerosis, Curr. Drug Deliv. **7**(4), 303–311 (2010)

21.53 A.C. Anselmo, S. Mitragotri: An overview of clinical and commercial impact of drug delivery systems, J. Control. Release **190**, 15–28 (2014) doi:10.1016/j.jconrel.2014.03.053

21.54 A.L. van de Ven, P. Kim, O.A.H. Haley, J.R. Fakhoury, G. Adriani, J. Schmulen, P. Moloney, F. Hussain, M. Ferrari, X. Liu, S.-H. Yun, P. Decuzzi: Rapid tumoritropic accumulation of systemically injected plateloid particles and their biodistribution, J. Control. Release **158**(1), 148–155 (2012) doi:10.1016/j.jconrel.2011.10.021

21.55 M. Nabil, P. Decuzzi, P. Zunino: Modelling mass and heat transfer in nano-based cancer hyperthermia, R. Soc. Open Sci. **2**(10), 150447 (2015) doi:10.1098/rsos.150447

21.56 R.E. Serda, S. Ferrati, B. Godin, E. Tasciotti, X. Liu, M. Ferrari: Mitotic trafficking of silicon microparticles, Nanoscale **1**(2), 250–259 (2009) doi:10.1039/B9NR00138G

21.57 A.C. Anselmo, M. Zhang, S. Kumar, D.R. Vogus, S. Menegatti, M.E. Helgeson, S. Mitragotri: Elasticity of nanoparticles influences their blood circulation, phagocytosis, endocytosis, and targeting, ACS Nano **9**(3), 3169–3177 (2015) doi:10.1021/acsnano.5b00147

21.58 D.D. Joseph, D. Ocando: Slip velocity and lift, J. Fluid Mech. **454**, 263–286 (2002) doi:10.1017/S0022112001007145

21.59 T.R. Lee, M. Choi, A.M. Kopacz, S.H. Yun, W.K. Liu, P. Decuzzi: On the near-wall accumulation of injectable particles in the microcirculation: Smaller is not better, Sci. Rep. **3**, 2079 (2013) doi:10.1038/srep02079

21.60 T.R. Lee, M.S. Greene, Z. Jiang, A.M. Kopacz, P. Decuzzi, W. Chen, W.K. Liu: Quantifying uncertainties in the microvascular transport of nanoparticles, Biomech. Model. Mechanobiol. **13**(3), 515–526 (2014) doi:10.1007/s10237-013-0513-0

21.61 M.D. de Tullio, P. De Palma, G. Iaccarino, G. Pascazio, M. Napolitano: An immersed boundary method for compressible flows using local grid refinement, J. Comput. Phys. **225**(2), 2098–2117 (2007)

21.62 S.S. Hossain, T.J. Hughes, P. Decuzzi: Vascular deposition patterns for nanoparticles in an inflamed patient-specific arterial tree, Biomech. Model. Mechanobiol. **13**(3), 585–597 (2014) doi:10.1007/s10237-013-0520-1

21.63 S.S. Hossain, Y. Zhang, X. Liang, F. Hussain, M. Ferrari, T.J. Hughes, P. Decuzzi: In silico vascular modeling for personalized nanoparticle delivery, Nanomedicine **8**(3), 343–357 (2013) doi:10.2217/nnm.12.124

21.64 A.J.C. Ladd: Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 1. Theoretical foundation, J. Fluid Mech. **271**, 285–309 (1994)

21.65 A.J.C. Ladd: Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 2. Numerical results, J. Fluid Mech. **271**, 311–339 (1994)

21.66 S. Succi: *The Lattice Boltzmann Equation: For Fluid Dynamics and Beyond* (Oxford University Press, Oxford 2001)

21.67 C. Sun, C. Migliorini, L.L. Munn: Red blood cells initiate leukocyte rolling in postcapillary expansions: A lattice Boltzmann analysis, Biophys. J. **85**(1), 208–222 (2003) doi:10.1016/S0006-3495(03)74467-1

21.68 A. Coclite, M.D. de Tullio, G. Pascazio, P. Decuzzi: A combined lattice Boltzmann and immersed boundary approach for predicting the vascular transport of differently shaped particles, Comput. Fluids **136**, 260–271 (2016)

21.69 H. Maeda: Macromolecular therapeutics in cancer treatment: The EPR effect and beyond, J. Control. Release **164**(2), 138–144 (2012) doi:10.1016/j.jconrel.2012.04.038

21.70 R.K. Jain, T. Stylianopoulos: Delivering nanomedicine to solid tumors, Nat. Rev. Clin. Oncol. **7**(11), 653–664 (2010) doi:10.1038/nrclinonc.2010.139

21.71 C. Stigliano, J. Key, M. Ramirez, S. Aryal, P. Decuzzi: Radiolabeled polymeric nanoconstructs loaded with docetaxel and curcumin for cancer combinatorial therapy and nuclear imaging, Adv. Funct. Mater. **25**(22), 3371–3379 (2015) doi:10.1002/adfm.201500627

21.72 A. Lee, D. Di Mascolo, M. Francardi, F. Piccardi, T. Bandiera, P. Decuzzi: Spherical polymeric nanoconstructs for combined chemotherapeutic and anti-inflammatory therapies, Nanomedicine **12**(7), 2139–2147 (2016) doi:10.1016/j.nano.2016.05.012

21.73 S.Y. Lee, M. Ferrari, P. Decuzzi: Shaping nano-/micro-particles for enhanced vascular interaction in laminar flows, Nanotechnology **20**(49), 495101 (2009) doi:10.1088/0957-4484/20/49/495101

21.74 P. Decuzzi, B. Godin, T. Tanaka, S.Y. Lee, C. Chiappini, X. Liu, M. Ferrari: Size and shape effects in the biodistribution of intravascularly injected particles, J. Control. Release **141**(3), 320–327 (2010) doi:10.1016/j.jconrel.2009.10.014

21.75 P. Decuzzi, M. Ferrari: The adhesive strength of non-spherical particles mediated by specific interactions, Biomaterials **27**(30), 5307–5314 (2006) doi:10.1016/j.biomaterials.2006.05.024

21.76 P. Decuzzi, M. Ferrari: Design maps for nanoparticles targeting the diseased microvasculature, Biomaterials **29**(3), 377–384 (2008) doi:10.1016/j.biomaterials.2007.09.025

21.77 L.T. Canham: Bioactive silicon structure fabrication through nanoetching techniques, Adv. Mater. **7**(12), 1033–1037 (1995) doi:10.1002/adma.19950071215

21.78 H. Ouyang, M. Christophersen, P.M. Fauchet: Enhanced control of porous silicon morphology from macropore to mesopore formation, Phys. Stat. Sol. A **202**(8), 1396–1401 (2005) doi:10.1002/pssa.200461112

21.79 L. Xiao, L. Gu, S.B. Howell, M.J. Sailor: Porous silicon nanoparticle photosensitizers for singlet oxygen and their phototoxicity against cancer cells, ACS Nano **5**, 3651–3659 (2011)

21.80 C. Chiappini, X. Liu, J.R. Fakhoury, M. Ferrari: Biodegradable porous silicon barcode nanowires with defined geometry, Adv. Funct. Mater. **20**(14), 2231–2239 (2010) doi:10.1002/adfm.201000360

21.81 C. Chiappini, E. De Rosa, J.O. Martinez, X. Liu, J. Steele, M.M. Stevens, E. Tasciotti: Biodegradable silicon nanoneedles delivering nucleic acids intracellularly induce localized in vivo neovascularization, Nat. Mater. **14**(5), 532–539 (2015) doi:10.1038/nmat4249

21.82 X.G. Zhang: Morphology and formation mechanisms of porous silicon, J. Electrochem. Soc. **151**(1), C69–C80 (2004) doi:10.1149/1.1632477

21.83 F. Dai, J. Zai, R. Yi, M.L. Gordin, H. Sohn, S. Chen, D. Wang: Bottom-up synthesis of high surface area mesoporous crystalline silicon and evaluation of its hydrogen evolution performance, Nat. Comm. **5**, 1–11 (2014) doi:10.1038/ncomms4605

21.84 L. Batchelor, A. Loni, L.T. Canham, M. Hasan, J.L. Coffer: Manufacture of mesoporous silicon from living plants and agricultural waste: An environmentally friendly and scalable process, Silicon **4**(4), 259–266 (2012) doi:10.1007/s12633-012-9129-8

21.85 M.J. Sailor: Fundamentals of porous silicon preparation. In: *Porous Silicon in Practice* (Wiley-VCH, Weinheim 2012) pp. 1–42

21.86 X. Li, P.W. Bohn: Metal-assisted chemical etching in HF/H_2O_2 produces porous silicon, Appl. Phys. Lett. **77**(16), 2572–2574 (2000) doi:10.1063/1.1319191

21.87 S.F. Chuang, S.D. Collins, R.L. Smith: Preferential propagation of pores during the formation of porous silicon: A transmission electron microscopy study, Appl. Phys. Lett. **55**(7), 675 (1989) doi:10.1063/1.101819

21.88 H. Foll, M. Christophersen, J. Carstensen, G. Hasse: Formation and application of porous silicon, Mater. Sci. Eng. **39**, 93–141 (2002)

21.89 L.T. Canham, L. Canham (Eds.): *Handbook of Porous Silicon* (Springer, Cham 2014)

21.90 M.G. Berger, R. Arens-Fischer, M. Thönissen, M. Krüger, S. Billat, H. Lüth, S. Hilbrich, W. Theiß, P. Grosse: Dielectric filters made of PS: Advanced performance by oxidation and new layer structures, Thin Sol. Films **297**(1/2), 237–240 (1997) doi:10.1016/S0040-6090(96)09361-3

21.91 F. Cunin, T.A. Schmedake, J.R. Link, Y.Y. Li, J. Koh, S.N. Bhatia, M.J. Sailor: Biomolecular screening with encoded porous-silicon photonic crystals, Nat. Mater. **1**(1), 39–41 (2002) doi:10.1038/nmat702

21.92 J. Salonen, V.P. Lehto, E. Laine: Thermal oxidation of free-standing porous silicon films, Appl. Phys. Lett. **70**(5), 637–639 (1997) doi:10.1063/1.118294

21.93 T.A. Schmedake, F. Cunin, J.R. Link, M.J. Sailor: Standoff detection of chemicals using porous silicon "smart dust" particles, Adv. Mater. **14**(18), 1270–1272 (2002)

21.94 J. Salonen, L. Laitinen, A. Kaukonen, J. Tuura, M. Bjorkqvist, T. Heikkila, K. Vahaheikkila, J. Hirvonen, V. Lehto: Mesoporous silicon microparticles for oral drug delivery: Loading and release of five model drugs, J. Control. Release **108**(2/3), 362–374 (2005) doi:10.1016/j.jconrel.2005.08.017

21.95 H.A. Santos, J. Riikonen, J. Salonen, E. Mäkilä, T. Heikkilä, T. Laaksonen, L. Peltonen, V.P. Lehto, J. Hirvonen: In vitro cytotoxicity of porous silicon microparticles: Effect of the particle concentration, surface chemistry and size,

Acta. Biomater. **6**(7), 2721–2731 (2010) doi:10.1016/j.actbio.2009.12.043
21.96 J.-H. Park, L. Gu, G. von Maltzahn, E. Ruoslahti, S.N. Bhatia, M.J. Sailor: Biodegradable luminescent porous silicon nanoparticles for in vivo applications, Nat. Mater. **8**(4), 331–336 (2009) doi:10.1038/nmat2398
21.97 R.A. Petros, J.M. DeSimone: Strategies in the design of nanoparticles for therapeutic applications, Nat. Rev. Drug Discov. **9**(8), 615–627 (2010) doi:10.1038/nrd2591
21.98 S.E.A. Gratton, P.A. Ropp, P.D. Pohlhaus, J.C. Luft, V.J. Madden, M.E. Napier, J.M. DeSimone: The effect of particle design on cellular internalization pathways, Proc. Natl. Acad. Sci. **105**(33), 11613–11618 (2008) doi:10.1073/pnas.0801763105
21.99 B. Godin, C. Chiappini, S. Srinivasan, J.F. Alexander, K. Yokoi, M. Ferrari, P. Decuzzi, X. Liu: Discoidal porous silicon particles: Fabrication and biodistribution in breast cancer bearing mice, Adv. Funct. Mater. **22**(20), 4225–4235 (2012) doi:10.1002/adfm.201200869
21.100 C. Chiappini, E. Tasciotti, J.R. Fakhoury, D. Fine, L. Pullan, Y.C. Wang, L. Fu, X. Liu, M. Ferrari: Tailored porous silicon microparticles: Fabrication and properties, Chem. Phys. Chem. **11**(5), 1029–1035 (2010) doi:10.1002/cphc.200900914
21.101 C. Chiappini, E. Tasciotti, R.E. Serda, L. Brousseau, X. Liu, M. Ferrari: Mesoporous silicon particles as intravascular drug delivery vectors: Fabrication, in-vitro, and in-vivo assessments, Phys. Stat. Sol. C **8**(6), 1826–1832 (2011) doi:10.1002/pssc.201000344
21.102 C. Zhang, C. Li, Z. Liu, J. Zheng, C. Xue, Y. Zuo, B. Cheng, Q. Wang: Electrically conductive and optically active porous silicon nanowires, Nano. Lett. **9**(12), 4539–4543 (2009) doi:10.1021/nl903030h
21.103 C. Chartier, S. Bastide, C. Lévy-Clément: Metal-assisted chemical etching of silicon in HF–H_2O_2, Electrochim. Acta. **53**(17), 5509–5516 (2008) doi:10.1016/j.electacta.2008.03.009
21.104 N. Geyer, B. Fuhrmann, Z. Huang, J. de Boor, H.S. Leipner, P. Werner: Model for the mass transport during metal-assisted chemical etching with contiguous metal films as catalysts, J. Phys. Chem. C **116**(24), 13446–13451 (2012) doi:10.1021/jp3034227
21.105 X. Zhong, Y. Qu, Y.-C. Lin, L. Liao, X. Duan: Unveiling the formation pathway of single crystalline porous silicon nanowires, ACS Appl. Mater. Interf. **3**(2), 261–270 (2011) doi:10.1021/am1009056
21.106 J. Kim, H. Rhu, W. Lee: A continuous process for Si nanowires with prescribed lengths, J. Mater. Chem. **21**(40), 15889–15894 (2011) doi:10.1039/C1JM13831F
21.107 M. Coluccio, F. Gentile, M. Francardi, G. Perozziello, N. Malara, P. Candeloro, E. Di Fabrizio: Electroless deposition and nanolithography can control the formation of materials at the nanoscale for plasmonic applications, Sensors **14**(4), 6056–6083 (2014) doi:10.3390/s140406056
21.108 R. Elnathan, L. Isa, D. Brodoceanu, A. Nelson, F.J. Harding, B. Delalat, T. Kraus, N.H. Voelcker: Versatile particle-based route to engineer vertically aligned silicon nanowire arrays and nanoscale pores, ACS Appl. Mater. Interf. **7**(42), 23717–23724 (2015) doi:10.1021/acsami.5b07777
21.109 H. Alhmoud, B. Delalat, R. Elnathan, A.C. Rius, A. Chaix, M.L. Rogers, J.-O. Durand, N.H. Voelcker: Porous silicon nanodiscs for targeted drug delivery, Adv. Funct. Mater. **25**(7), 1137–1145 (2015) doi:10.1002/adfm.201403414
21.110 H. Lin, H.-Y. Cheung, F. Xiu, F. Wang, S. Yip, N. Han, T. Hung, J. Zhou, J.C. Ho, C.-Y. Wong: Developing controllable anisotropic wet etching to achieve silicon nanorods, nanopencils and nanocones for efficient photon trapping, J. Mater. Chem. A **1**(34), 9942–9946 (2013) doi:10.1039/C3TA11889D
21.111 F. Gentile, M.L. Coluccio, R.P. Zaccaria, M. Francardi, G. Cojoc, G. Perozziello, R. Raimondo, P. Candeloro, E. Di Fabrizio: Selective on site separation and detection of molecules in diluted solutions with super-hydrophobic clusters of plasmonic nanoparticles, Nanoscale **6**(14), 8208–8225 (2014) doi:10.1039/C4NR00796D
21.112 G.S. Higashi, Y.J. Chabal, G.W. Trucks, K. Raghavachari: Ideal hydrogen termination of the Si (111) surface, Appl. Phys. Lett. **56**(7), 656–658 (1990) doi:10.1063/1.102728
21.113 A.G. Cullis, L.T. Canham, P.D.J. Calcott: The structural and luminescence properties of porous silicon, J. Appl. Phys. **82**(3), 909 (1997) doi:10.1063/1.366536
21.114 M. Morita, T. Ohmi, E. Hasegawa, M. Kawakami, M. Ohwada: Growth of native oxide on a silicon surface, J. Appl. Phys. **68**(3), 1272 (1990) doi:10.1063/1.347181
21.115 M.A. Tischler, R.T. Collins, J.H. Stathis, J.C. Tsang: Luminescence degradation in porous silicon, Appl. Phys. Lett. **60**(5), 639 (1992) doi:10.1063/1.106578
21.116 M.V. Wolkin, J. Jorne, P.M. Fauchet, G. Allan, C. Delerue: Electronic states and luminescence in porous silicon quantum dots: The role of oxygen, Phys. Rev. Lett. **82**(1), 197–200 (1999) doi:10.1103/PhysRevLett.82.197
21.117 J. Riikonen, M. Salomäki, J. van Wonderen, M. Kemell, W. Xu, O. Korhonen, M. Ritala, F. MacMillan, J. Salonen, V.P. Lehto: Surface chemistry, reactivity, and pore structure of porous silicon oxidized by various methods, Langmuir **28**(28), 10573–10583 (2012) doi:10.1021/la301642w
21.118 J.B. Brzoska, I.B. Azouz, F. Rondelez: Silanization of solid substrates: A step toward reproducibility, Langmuir **10**(11), 4367–4373 (1994), doi:10.1021/la00023a072
21.119 G.T. Hermanson: *Bioconjugate Techniques*, 3rd edn. (Elsevier, Amsterdam 2013)
21.120 J. Salonen, E. Laine, L. Niinistö: Thermal carbonization of porous silicon surface by acetylene, J. Appl. Phys. **91**(1), 456 (2002) doi:10.1063/1.1421221
21.121 P. Allongue, V.C. Kieling, H. Gerischer: Etching of silicon in NaOH solutions, J. Electrochem. Soc.

140(4), 1009–1018 (1993) doi:10.1149/1.2056189
21.122 J.O. Martinez, C. Chiappini, A. Ziemys, A.M. Faust, M. Kojic, X. Liu, M. Ferrari, E. Tasciotti: Engineering multi-stage nanovectors for controlled degradation and tunable release kinetics, Biomaterials **34**, 8469–8477 (2013) doi:10.1016/j.biomaterials.2013.07.049
21.123 R.E. Serda, A. Mack, M. Pulikkathara, A.M. Zaske, C. Chiappini, J.R. Fakhoury, D. Webb, B. Godin, J.L. Conyers, X.W. Liu, J.A. Bankson, M. Ferrari: Cellular association and assembly of a multistage delivery system, Small **6**(12), 1329–1340 (2010) doi:10.1002/smll.201000126
21.124 S.H.C. Anderson, H. Elliott, D.J. Wallis, L.T. Canham, J.J. Powell: Dissolution of different forms of partially porous silicon wafers under simulated physiological conditions, Phys. Stat. Sol A **197**(2), 331–335 (2003) doi:10.1002/pssa.200306519
21.125 B. Godin, J. Gu, R.E. Serda, R. Bhavane, E. Tasciotti, C. Chiappini, X. Liu, T. Tanaka, P. Decuzzi, M. Ferrari: Tailoring the degradation kinetics of mesoporous silicon structures through PEGylation, J. Biomed. Mater. Res. A **94A**(4), 1236–1243 (2010) doi:10.1002/jbm.a.32807
21.126 W. Sun, J.E. Puzas, T.J. Sheu, X. Liu, P.M. Fauchet: Nano-to microscale porous silicon as a cell interface for bone-tissue engineering, Adv. Mater. **19**(7), 921–924 (2007) doi:10.1002/adma.200600319
21.127 T. Jalkanen, E. Mäkilä, Y.I. Suzuki, T. Urata, K. Fukami, T. Sakka, J. Salonen, Y.H. Ogata: Studies on chemical modification of porous silicon-based graded-index optical microcavities for improved stability under alkaline conditions, Adv. Funct. Mater. **22**(18), 3890–3898 (2012) doi:10.1002/adfm.201200386
21.128 A.T. Balter, Z. Shatsberg, M. Beckerman, E. Segal, N. Artzi: Mechanism of erosion of nanostructured porous silicon drug carriers in neoplastic tissues, Nat. Comm. **6**, 6208 (2015) doi:10.1038/ncomms7208
21.129 T. Tanaka, B. Godin, R. Bhavane, R.N. Alicea, J. Gu, X. Liu, C. Chiappini, J.R. Fakhoury, S. Amra, A. Ewing: In vivo evaluation of safety of nanoporous silicon carriers following single and multiple dose intravenous administrations in mice, Int. J. Pharm. **402**(1/2), 190–197 (2010) doi:10.1016/j.ijpharm.2010.09.015
21.130 L. Gu, D.J. Hall, Z. Qin, E. Anglin, J. Joo, D.J. Mooney, S.B. Howell, M.J. Sailor: In vivo time-gated fluorescence imaging with biodegradable luminescent porous silicon nanoparticles, Nat. Comm. **4**, 1–7 (2013) doi:10.1038/ncomms3326
21.131 M. Ariza-Avidad, A. Nieto, A.S. alinas-Castillo, L.F. Capitan-Vallvey, G.M. Miskelly, M.J. Sailor: Monitoring of degradation of porous silicon photonic crystals using digital photography, Nanoscale Res. Lett. **9**(1), 410 (2014) doi:10.1186/1556-276X-9-410
21.132 S.P. Low, K.A. Williams, L.T. Canham, N.H. Voelcker: Generation of reactive oxygen species from porous silicon microparticles in cell culture medium, J. Biomed. Mater. Res. A **93A**(3), 1124–1131 (2010) doi:10.1002/jbm.a.32610
21.133 S. Low, K. Williams, L. Canham, N. Voelcker: Evaluation of mammalian cell adhesion on surface-modified porous silicon, Biomaterials **27**(26), 4538–4546 (2006) doi:10.1016/j.biomaterials.2006.04.015
21.134 S.D. Alvarez, A.M. Derfus, M.P. Schwartz, S.N. Bhatia, M.J. Sailor: The compatibility of hepatocytes with chemically modified porous silicon with reference to in vitro biosensors, Biomaterials **30**(1), 26–34 (2009) doi:10.1016/j.biomaterials.2008.09.005
21.135 L.M. Bimbo, M. Sarparanta, H.A. Santos, A.J. Airaksinen, E. Mäkilä, T. Laaksonen, L. Peltonen, V.P. Lehto, J. Hirvonen, J. Salonen: Biocompatibility of thermally hydrocarbonized porous silicon nanoparticles and their biodistribution in rats, ACS Nano **4**(6), 3023–3032 (2010) doi:10.1021/nn901657w
21.136 S.P. Low, N.H. Voelcker, L.T. Canham, K.A. Williams: The biocompatibility of porous silicon in tissues of the eye, Biomaterials **30**(15), 2873–2880 (2009) doi:10.1016/j.biomaterials.2009.02.008
21.137 A.S.-W. Goh, A.Y.-F. Chung, R.H.-G. Lo, T.-N. Lau, S.W.-K. Yu, M. Chng, S. Satchithanantham, S.L.-E. Loong, D.C.-E. Ng, B.-C. Lim, S. Connor, P.K.-H. Chow: A novel approach to brachytherapy in hepatocellular carcinoma using a phosphorous32 (32P) brachytherapy delivery device – A first-in-man study, Int. J. Radiat. Oncol. Biol. Phys. **67**(3), 786–792 (2007) doi:10.1016/j.ijrobp.2006.09.011
21.138 R.E. Serda, J. Gu, R.C. Bhavane, X. Liu, C. Chiappini, P. Decuzzi, M. Ferrari: The association of silicon microparticles with endothelial cells in drug delivery to the vasculature, Biomaterials **30**(13), 2440–2448 (2009) doi:10.1016/j.biomaterials.2009.01.019
21.139 S. Ferrati, A. Mack, C. Chiappini, X. Liu, A.J. Bean, M. Ferrari, R.E. Serda: Intracellular trafficking of silicon particles and logic-embedded vectors, Nanoscale **2**(8), 1512–1520 (2010) doi:10.1039/C0NR00227E
21.140 E. Tasciotti, B. Godin, J.O. Martinez, C. Chiappini, R.C. Bhavane, X. Liu, M. Ferrari: Near-infrared imaging method for the in vivo assessment of the biodistribution of nanoporous silicon particles, Mol. Imag. **10**(1), 56 (2011)
21.141 E. De Rosa, C. Chiappini, D. Fan, X. Liu, M. Ferrari, E. Tasciotti: Agarose surface coating influences intracellular accumulation and enhances payload stability of a nano-delivery system, Pharm. Res. **28**(7), 1520–1530 (2011) doi:10.1007/s11095-011-0453-2
21.142 D. Liu, L.M. Bimbo, E. Mäkilä, F. Villanova, M. Kaasalainen, B. Herranz-Blanco, C.M. Caramella, V.P. Lehto, J. Salonen, K.-H. Herzig, J. Hirvonen, H.A. Santos: Co-delivery of a hydrophobic small molecule and a hydrophilic peptide by porous silicon nanoparticles, J. Control. Release **170**(2), 268–278 (2013) doi:10.1016/j.jconrel.2013.05.036

21.143 M. Xue, X. Zhong, Z. Shaposhnik, Y. Qu, F. Tamanoi, X. Duan, J.I. Zink: pH-operated mechanized porous silicon nanoparticles, J. Am. Chem. Soc. **133**(23), 8798–8801 (2011) doi:10.1021/ja201252e

21.144 A.P. Mann, T. Tanaka, A. Somasunderam, X. Liu, D.G. Gorenstein, M. Ferrari: E-selectin-targeted porous silicon particle for nanoparticle delivery to the bone marrow, Adv. Mater. **23**(36), H278–H282 (2011) doi:10.1002/adma.201101541

21.145 E. Secret, M. Maynadier, A. Gallud, M.G. Bobo, A. Chaix, E. Belamie, P. Maillard, M.J. Sailor, M. Garcia, J.O. Durand, F. Cunin: Anionic porphyrin -grafted porous silicon nanoparticles for photodynamic therapy, Chem. Comm. **49**(39), 4202–4204 (2013) doi:10.1039/C3CC38837A

21.146 M. Wang, J.L. Coffer, K. Dorraj, P.S. Hartman, A. Loni, L.T. Canham: Sustained antibacterial activity from triclosan-loaded nanostructured mesoporous silicon, Mol. Pharm. **7**(6), 2232–2239 (2010) doi:10.1021/mp100227m

21.147 R. Xu, G. Zhang, J. Mai, X. Deng, V. Segura-Ibarra, S. Wu, J. Shen, H. Liu, Z. Hu, L. Chen, Y. Huang, E. Koay, Y. Huang, J. Liu, J.E. Ensor, E. Blanco, X. Liu, M. Ferrari, H. Shen: An injectable nanoparticle generator enhances delivery of cancer therapeutics, Nat. Biotechnol. **34**(4), 414–418 (2016) doi:10.1038/nbt.3506

21.148 J.O. Martinez, M. Evangelopoulos, V. Karun, E. Shegog, J.A. Wang, C. Boada, X. Liu, M. Ferrari, E. Tasciotti: The effect of multistage nanovector targeting of VEGFR2 positive tumor endothelia on cell adhesion and local payload accumulation, Biomaterials **35**(37), 9824–9832 (2014) doi:10.1016/j.biomaterials.2014.08.024

21.149 L. Gu, L.E. Ruff, Z. Qin, M. Corr, S.M. Hedrick, M.J. Sailor: Multivalent porous silicon nanoparticles enhance the immune activation potency of agonistic CD40 antibody, Adv. Mater. **24**(29), 3981–3987 (2012) doi:10.1002/adma.201200776

21.150 C. Chiappini, J.O. Martinez, E. De Rosa, C.S. Almeida, E. Tasciotti, M.M. Stevens: Biodegradable nanoneedles for localized delivery of nanoparticles in vivo: Exploring the biointerface, ACS Nano **9**, 5500–5509 (2015) doi:10.1021/acsnano.5b01490

21.151 M.M. Orosco, C. Pacholski, G.M. Miskelly, M.J. Sailor: Protein-coated porous-silicon photonic crystals for amplified optical detection of protease activity, Adv. Mater. **18**(11), 1393–1396 (2006) doi:10.1002/adma.200502420

21.152 C. Chiappini, P. Campagnolo, C.S. Almeida, N. Annassi-Ghadi, L.W. Chow, G.B. Hanna, M.M. Stevens: Mapping Local cytosolic enzymatic activity in human esophageal mucosa with porous silicon nanoneedles, Adv. Mater. **27**(35), 5147–5152 (2015) doi:10.1002/adma.201501304

21.153 A. Caliò, I. Rea, J. Politi, P. Giardina, S. Longobardi, L.D. Stefano: Hybrid bio/non-bio interfaces for protein-glucose interaction monitoring, J. Appl. Phys. **114**(13), 134904 (2013) doi:10.1063/1.4824379

21.154 G. Rong, J.D. Ryckman, R.L. Mernaugh, S.M. Weiss: Label-free porous silicon membrane waveguide for DNA sensing, Appl. Phys. Lett. **93**(16), 161109 (2008) doi:10.1063/1.3005620

21.155 N. Massad-Ivanir, G. Shtenberg, A. Tzur, M.A. Krepker, E. Segal: Engineering nanostructured porous SiO_2 surfaces for bacteria detection via "direct cell capture", Anal. Chem. **83**(9), 3282–3289 (2011) doi:10.1021/ac200407w

21.156 L. De Stefano, I. Rea, I. Rendina, L. Rotiroti, M. Rossi, S. D'Auria: Resonant cavity enhanced optical microsensor for molecular interactions based on porous silicon, Phys. Stat. Sol. A **203**(5), 886–891 (2006) doi:10.1002/pssa.200521350

21.157 S. D'Auria, M. de Champdoré, V. Aurilia, A. Parracino, M. Staiano, A. Vitale, M. Rossi, I. Rea, L. Rotiroti, A.M. Rossi, S. Borini, I. Rendina, L. De Stefano: Nanostructured silicon-based biosensors for the selective identification of analytes of social interest, J. Phys. Condens. Matter **18**(33), S2019–S2028 (2006) doi:10.1088/0953-8984/18/33/S17

21.158 V.S.Y. Lin, K. Motesharei, K.-P.S. Dancil, M.J. Sailor, M.R. Ghadiri: A porous silicon-based optical interferometric biosensor, Science **278**(5339), 840–843 (1997) doi:10.1126/science.278.5339.840

21.159 J. Zhang, Y. Wu, B. Zhang, M. Li, S. Jia, S. Jiang, H. Zhou, Y. Zhang, C. Zhang, A.P.F. Turner: Label-free electrochemical detection of tetracycline by an aptamer nano-biosensor, Anal. Lett. **45**(9), 986–992 (2012) doi:10.1080/00032719.2012.670784

21.160 L.A. Osminkina, V.A. Sivakov, G.A. Mysov, V.A. Georgobiani, U.A. Natashina, F. Talkenberg, V.V. Solovyev, A.A. Kudryavtsev, V.Y. Timoshenko: Nanoparticles prepared from porous silicon nanowires for bio-imaging and sonodynamic therapy, Nanoscale Res. Lett. **9**(1), 1 (2014) doi:10.1186/1556-276X-9-463

21.161 A. Gizzatov, C. Stigliano, J.S. Ananta, R. Sethi, R. Xu, A. Guven, M. Ramirez, H. Shen, A. Sood, M. Ferrari, L.J. Wilson, X. Liu, P. Decuzzi: Geometrical confinement of Gd(DOTA) molecules within mesoporous silicon nanoconstructs for MR imaging of cancer, Cancer Lett. **352**(1), 97–101 (2014) doi:10.1016/j.canlet.2014.06.001

21.162 R. Duncan, R. Gaspar: Nanomedicine (s) under the microscope, Mol. Pharm. **8**(6), 2101–2141 (2011)

21.163 W.B. Liechty, D.R. Kryscio, B.V. Slaughter, N.A. Peppas: Polymers for drug delivery systems, Annu. Rev. Chem. Biomol. Eng. **1**, 149 (2010)

21.164 Y.H. Yun, B.K. Lee, K. Park: Controlled drug delivery: Historical perspective for the next generation, J. Control. Release **219**, 2–7 (2011)

21.165 H.A. Santos: *Porous Silicon for Biomedical Applications* (Elsevier, Amsterdam 2014)

21.166 N. Shrestha, M.-A. Shahbazi, F. Araújo, E. Mäkilä, J. Raula, E.I. Kauppinen, J. Salonen, B. Sarmento, J. Hirvonen, H.A. Santos: Multistage pH-responsive mucoadhesive nanocarriers prepared by aerosol flow reactor technology: A controlled dual protein-drug delivery system, Biomaterials

68, 9–20 (2015)
21.167 M.K. Marschütz, A. Bernkop-Schnürch: Oral peptide drug delivery: polymer–inhibitor conjugates protecting insulin from enzymatic degradation in vitro, Biomaterials **21**(14), 1499–1507 (2000)
21.168 S. Munier, I. Messai, T. Delair, B. Verrier, Y. Ataman-Önal: Cationic PLA nanoparticles for DNA delivery: Comparison of three surface polycations for DNA binding, protection and transfection properties, Colloids Surf. B **43**(3), 163–173 (2005)
21.169 H.S. Choi, W. Liu, P. Misra, E. Tanaka, J.P. Zimmer, B.I. Ipe, M.G. Bawendi, J.V. Frangioni: Renal clearance of quantum dots, Nat. Biotechnol. **25**(10), 1165–1170 (2007)
21.170 M.I. Stockman, D.J. Bergman, T. Kobayashi: Coherent control of nanoscale localization of ultrafast optical excitation in nanosystems, Phys. Rev. B **69**, 054202 (2004)
21.171 S. Stolnik, L. Illum, S. Davis: Long circulating microparticulate drug carriers, Adv. Drug Deliv. Rev. **64**, 290–301 (2012)
21.172 H.A. Santos, E. Mäkilä, A.J. Airaksinen, L.M. Bimbo, J. Hirvonen: Porous silicon nanoparticles for nanomedicine: Preparation and biomedical applications, Nanomedicine **9**(4), 535–554 (2014)
21.173 H.A. Santos, J. Hirvonen: Nanostructured porous silicon materials: Potential candidates for improving drug delivery, Nanomedicine **7**(9), 1281–1284 (2012)
21.174 F. Fontana, M.P. Ferreira, A. Correia, J. Hirvonen, H.A. Santos: Microfluidics as a cutting-edge technique for drug delivery applications, J. Drug Deliv. Sci. Technol. **34**, 76–87 (2016)
21.175 I. Polenz, D.A. Weitz, J.-C. Baret: Polyurea microcapsules in microfluidics: Surfactant control of soft membranes, Langmuir **31**(3), 1127–1134 (2015)
21.176 A. Utada, L.-Y. Chu, A. Fernandez-Nieves, D. Link, C. Holtze, D. Weitz: Dripping, jetting, drops, and wetting: The magic of microfluidics, MRS Bull. **32**(09), 702–708 (2007)
21.177 A.S. Utada, A. Fernandez-Nieves, H.A. Stone, D.A. Weitz: Dripping to jetting transitions in coflowing liquid streams, Phys. Rev. Lett. **99**(9), 094502 (2007)
21.178 Y. Kim, B.L. Chung, M. Ma, W.J. Mulder, Z.A. Fayad, O.C. Farokhzad, R. Langer: Mass production and size control of lipid–polymer hybrid nanoparticles through controlled microvortices, Nano Lett. **12**(7), 3587–3591 (2012)
21.179 P.M. Valencia, O.C. Farokhzad, R. Karnik, R. Langer: Microfluidic technologies for accelerating the clinical translation of nanoparticles, Nat. Nanotechnol. **7**(10), 623–629 (2012) doi:10.1038/nnano.2012.168
21.180 E.F. Petricoin, D.K. Ornstein, C.P. Paweletz, A. Ardekani, P.S. Hackett, B.A. Hitt, A. Velasco, C. Trucco, L. Wiegand, K. Wood, C.B. Simone, P.J. Levine, W.M. Linehan, M.R.E. Buck, S.M. Steinberg, E.C. Kohn, L.A. Liotta: Serum proteomic patterns for detection of prostate cancer, J. Natl. Cancer Inst. **94**(20), 1576–1578 (2002)
21.181 T. Rossow, J.A. Heyman, A.J. Ehrlicher, A. Langhoff, D.A. Weitz, R. Haag, S. Seiffert: Controlled synthesis of cell-laden microgels by radical-free gelation in droplet microfluidics, J. Am. Chem. Soc. **134**(10), 4983–4989 (2012) doi:10.1021/Ja300460p
21.182 G.M. Whitesides: The origins and the future of microfluidics, Nature **442**(7101), 368–373 (2006) doi:10.1038/nature05058
21.183 A. Manz, D.J. Harrison, E.M. Verpoorte, J.C. Fettinger, A. Paulus, H. Lüdi, H.M. Widmer: Planar chips technology for miniaturization and integration of separation techniques into monitoring systems: Capillary electrophoresis on a chip, J. Chromatogr. A **593**(1), 253–258 (1992)
21.184 Y. Hu, A. Bouamrani, E. Tasciotti, L. Li, X. Liu, M. Ferrari: Tailoring of the nanotexture of mesoporous silica films and their functionalized derivatives for selectively harvesting low molecular weight protein, ACS Nano **4**(1), 439–451 (2010)
21.185 J.S. Sander, L. Isa, P.A. Rühs, P. Fischer, A.R. Studart: Stabilization mechanism of double emulsions made by microfluidics, Soft Matter **8**(45), 11471–11477 (2012)
21.186 L.R. Arriaga, S.S. Datta, S.H. Kim, E. Amstad, T.E. Kodger, F. Monroy, D.A. Weitz: Ultrathin shell double emulsion templated giant unilamellar lipid vesicles with controlled microdomain formation, Small **10**(5), 950–956 (2014)
21.187 S. Nie, S.R. Emory: Probing single molecules and single nanoparticles by surface-enhanced Raman scattering, Science **275**(5303), 1102–1106 (1997)
21.188 Y.-C. Tan, V. Cristini, A.P. Lee: Monodispersed microfluidic droplet generation by shear focusing microfluidic device, Sens. Actuators B **114**(1), 350–356 (2006)
21.189 J.C. Stachowiak, D.L. Richmond, T.H. Li, A.P. Liu, S.H. Parekh, D.A. Fletcher: Unilamellar vesicle formation and encapsulation by microfluidic jetting, Proc. Natl. Acad. Sci. **105**(12), 4697–4702 (2008)
21.190 G. Whitesides, A. Stroock:: Flexible methods for microfluidics, J. Phys. Today **54**(6), 42–48 (2001)
21.191 A.S. Utada, E. Lorenceau, D.R. Link, P.D. Kaplan, H.A. Stone, D.A. Weitz: Monodisperse double emulsions generated from a microcapillary device, Science **308**(5721), 537–541 (2005) doi:10.1126/science.1109164
21.192 S. Yang, F. Guo, B. Kiraly, X. Mao, M. Lu, K.W. Leong, T.J. Huang: Microfluidic synthesis of multifunctional Janus particles for biomedical applications, Lab Chip **12**(12), 2097–2102 (2012) doi:10.1039/c2lc90046g
21.193 S. Mazzitelli, L. Capretto, F. Quinci, R. Piva, C. Nastruzzi: Preparation of cell-encapsulation devices in confined microenvironment, Adv. Drug Deliv. Rev. **65**(11/12), 1533–1555 (2013)
21.194 Y. Kim, F. Fay, D.P. Cormode, B.L. Sanchez-Gaytan, J. Tang, E.J. Hennessy, M.M. Ma, K. Moore, O.C. Farokhzad, E.A. Fisher, W.J.M. Mulder, R. Langer, Z.A. Fayad: Single step reconstitution of multifunctional high-density lipoprotein-derived nanomaterials using microfluidics,

ACS Nano **7**(11), 9975–9983 (2013) doi:10.1021/Nn4039063
21.195 H.A. Santos: Opinion paper: Microfluidics technique to revolutionize the drug delivery field: Current developments and applications, Curr. Drug Deliv. **12**(6), 642–644 (2015)
21.196 S. Marre, K.F. Jensen: Synthesis of micro and nanostructures in microfluidic systems, Chem. Soc. Rev. **39**(3), 1183–1202 (2010) doi:10.1039/b821324k
21.197 R. Karnik, F. Gu, P. Basto, C. Cannizzaro, L. Dean, W. Kyei-Manu, R. Langer, O.C. Farokhzad: Microfluidic platform for controlled synthesis of polymeric nanoparticles, Nano Lett. **8**(9), 2906–2912 (2008)
21.198 S.H. Kim, J.W. Kim, D.H. Kim, S.H. Han, D.A. Weitz: Polymersomes containing a hydrogel network for high stability and controlled release, Small **9**(1), 124–131 (2013)
21.199 S. Stainmesse, H. Fessi, J.P. Devissaguet, F. Puisieux, C. Theis: Process for the preparation of dispersible colloidal systems of a substance in the form of nanoparticles, US Patent 5133908 (US005133908A) (1992)
21.200 F. Ganachaud, J.L. Katz: Nanoparticles and nanocapsules created using the Ouzo effect: Spontaneous emulsification as an alternative to ultrasonic and high-shear devices, Chem. Phys. Chem. **6**(2), 209–216 (2005)
21.201 F. Kong, X. Zhang, H. Zhang, X. Qu, D. Chen, M. Servos, E. Mäkilä, J. Salonen, H.A. Santos, M. Hai, D.A. Weitz: Inhibition of multidrug resistance of cancer cells by co-delivery of DNA nanostructures and drugs using porous silicon nanoparticles@giant liposomes, Adv. Funct. Mater. **25**(22), 3330–3340 (2015) doi:10.1002/adfm.201500594
21.202 F. Araujo, N. Shrestha, M.A. Shahbazi, D. Liu, B. Herranz-Blanco, E.M. Makila, J.J. Salonen, J.T. Hirvonen, P.L. Granja, B. Sarmento, H.A. Santos: Microfluidic assembly of a multifunctional tailorable composite system designed for site specific combined oral delivery of peptide drugs, ACS Nano **9**(8), 8291–8302 (2015) doi:10.1021/acsnano.5b02762
21.203 H. Zhang, D. Liu, M.A. Shahbazi, E. Makila, B. Herranz-Blanco, J. Salonen, J. Hirvonen, H.A. Santos: Fabrication of a multifunctional nano-in-micro drug delivery platform by microfluidic templated encapsulation of porous silicon in polymer matrix, Adv. Mater. **26**(26), 4497–4503 (2014) doi:10.1002/adma.201400953
21.204 D. Liu, H. Zhang, B.-H. Blanco, E. Makila, V.P. Lehto, J. Salonen, J. Hirvonen, H.A. Santos: Microfluidic assembly of monodisperse multistage pH-responsive polymer/porous silicon composites for precisely controlled multidrug delivery, Small **10**(10), 2029–2038 (2014) doi:10.1002/smll.201303740
21.205 D. Liu, B. Herranz-Blanco, E. Makila, L.R. Arriaga, S. Mirza, D.A. Weitz, N. Sandler, J. Salonen, J. Hirvonen, H.A. Santos: Microfluidic templated mesoporous silicon-solid lipid microcomposites for sustained drug delivery, ACS Appl. Mater. Interfaces **5**(22), 12127–12134 (2013) doi:10.1021/am403999q
21.206 A. Jahn, S.M. Stavis, J.S. Hong, W.N. Vreeland, D.L. Devoe, M. Gaitan: Microfluidic mixing and the formation of nanoscale lipid vesicles, ACS Nano **4**(4), 2077–2087 (2010) doi:10.1021/Nn901676x
21.207 Y.T. Kim, B.L. Chung, M.M. Ma, W.J.M. Mulder, Z.A. Fayad, O.C. Farokhzad, R. Langer: Mass production and size control of lipid-polymer hybrid nanoparticles through controlled microvortices, Nano Lett. **12**(7), 3587–3591 (2012) doi:10.1021/Nl301253v
21.208 N. Kolishetti, S. Dhar, P.M. Valencia, L.Q. Lin, R. Karnik, S.J. Lippard, R. Langer, O.C. Farokhzad: Engineering of self-assembled nanoparticle platform for precisely controlled combination drug therapy, Proc. Natl. Acad. Sci. **107**(42), 17939–17944 (2010) doi:10.1073/pnas.1011368107
21.209 A. Jahn, W.N. Vreeland, M. Gaitan, L.E. Locascio: Controlled vesicle self-assembly in microfluidic channels with hydrodynamic focusing, J. Am. Chem. Soc. **126**(9), 2674–2675 (2004) doi:10.1021/Ja0318030
21.210 C.J. Thiele: Neuroblastoma. In: *Human Cell Culture*, Vol. 1, ed. by J.R.W. Masters, B. Palsson (Springer, Dordrecht 1998) pp. 21–53
21.211 J.O. McNamara, E.R. Andrechek, Y. Wang, K.D. Viles, R.E. Rempel, E. Gilboa, B.A. Sullenger, P.H. Giangrande: Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras, Nat. Biotechnol. **24**(8), 1005–1015 (2006) doi:10.1038/Nbt1223
21.212 F.S. Majedi, M.M. Hasani-Sadrabadi, J.J. VanDersarl, N. Mokarram, S. Hojjati-Emami, E. Dashtimoghadam, S. Bonakdar, M.A. Shokrgozar, A. Bertsch, P. Renaud: On-chip fabrication of paclitaxel-loaded chitosan nanoparticles for cancer therapeutics, Adv. Funct. Mater. **24**(4), 432–441 (2014)
21.213 A.J. Mieszawska, Y. Kim, A. Gianella, I. van Rooy, B. Priem, M.P. Labarre, C. Ozcan, D.P. Cormode, A. Petrov, R. Langer, O.C. Farokhzad, Z.A. Fayad, W.J.M. Mulder: Synthesis of polymer-lipid nanoparticles for image-guided delivery of dual modality therapy, Bioconj. Chem. **24**(9), 1429–1434 (2013) doi:10.1021/Bc400166j
21.214 P.M. Valencia, P.A. Basto, L.F. Zhang, M. Rhee, R. Langer, O.C. Farokhzad, R. Karnik: Single-step assembly of homogenous lipid-polymeric and lipid-quantum dot nanoparticles enabled by microfluidic rapid mixing, ACS Nano **4**(3), 1671–1679 (2010)
21.215 M. Wang, M. Thanou: Targeting nanoparticles to cancer, Pharmacol. Res. **62**(2), 90–99 (2010)
21.216 L. Zhang, F. Gu, J. Chan, A. Wang, R. Langer, O. Farokhzad: Nanoparticles in medicine: Therapeutic applications and developments, Clin. Pharmacol. Ther. **83**(5), 761–769 (2008)
21.217 K.W. Spitzer, R.D. Vaughan-Jones: Regulation of intracellular pH in mammalian cells. In: *The Sodium-Hydrogen Exchanger*, ed. by M. Kar-

mazyn, M. Arkiran, L. Fliegal (Springer, New York 2003) pp. 1–15
21.218 M. Stubbs, P.M. McSheehy, J.R. Griffiths, C.L. Bashford: Causes and consequences of tumour acidity and implications for treatment, Mol. Med. Today **6**(1), 15–19 (2000)
21.219 K.E. Broaders, J.A. Cohen, T.T. Beaudette, E.M. Bachelder, J.M.J. Frechet: Acetalated dextran is a chemically and biologically tunable material for particulate immunotherapy, Proc. Natl. Acad. Sci. USA **106**(14), 5497–5502 (2009)
21.220 E.M. Bachelder, T.T. Beaudette, K.E. Broaders, J.M.J. Frechet, M.T. Albrecht, A.J. Mateczun, K.M. Ainslie, J.T. Pesce, A.M. Keane-Myers: In vitro analysis of acetalated dextran microparticles as a potent delivery platform for vaccine adjuvants, Mol. Pharm. **7**(3), 826–835 (2010)
21.221 M. Stockman: Nanoplasmonics: The physics behind the applications, Phys. Today **64**(2), 39 (2011)
21.222 H. Wendt: Ullmann's encyclopedia of industrial chemistry. In: *Ullmann's Encyclopedia of Industrial Chemistry, "Electrochemistry"*, ed. by H. Wendt (Wiley-VCH, Winheim 2004)
21.223 V. Dimitrova, L. Gorker: Modified Nernst equation for electroless metal deposition, Prog. React. Kinet. Mech. **31**(1), 45–58 (2006)
21.224 D. Goia, E. Matijevic: Preparation of monodispersed metal particles, New J. Chem. **22**, 1203–1215 (1998)
21.225 S. Yae, N. Nasu, K. Matsumoto, T. Hagihara, N. Fukumuro, H. Matsuda: Catalytic activity of noble metals for metal-assisted chemical etching of silicon, Electrochimica. Acta. **53**, 35–41 (2007)
21.226 S. Yae, N. Nasu, K. Matsumoto, T. Hagihara, N. Fukumuro, H. Matsuda: Nucleation behavior in electroless displacement deposition of metals on silicon from hydrofluoric acid solutions, Electrochim. Acta. **53**(1), 35–41 (2007)
21.227 K. Kneipp: Surface-enhanced Raman scattering, Phys. Today **60**(11), 40–47 (2007)
21.228 M.L. Coluccio, F. Gentile, G. Das, A. Nicastri, A.M. Perri, P. Candeloro, G. Perozziello, R.P. Zaccaria, J.S.T. Gongora, S. Alrasheed, A. Fratalocchi, T. Limongi, G. Cuda, E. Di Fabrizio: Detection of single amino acid mutation in human breast cancer by disordered plasmonic self-similar chain, Sci. Adv. **1**(8), e1500487 (2015)
21.229 F. Gentile, M. Coluccio, A. Toma, E. Rondanina, M. Leoncini, F. De Angelis, G. Das, C. Dorigoni, P. Candeloro, E. Di Fabrizio: Electroless deposition dynamics of silver nanoparticles clusters: A diffusion limited aggregation (DLA) approach, Microelectron. Eng. **98**, 359–362 (2012)
21.230 R. Dawkins, D. ben-Avraham: Computer simulations of diffusion-limited reactions, Comput. Sci. Eng. **3**(1), 72–76 (2001)
21.231 T. Qiu, P. Chu: Self-selective electroless plating: an approach for fabrication of functional 1-D nanomaterials, Mater. Sci. Eng. R **61**, 59–77 (2008)
21.232 T. Qiu, X. Wu, Y. Mei, P. Chu, G. Siu: Self-organized synthesis of silver dendritic nanostructures via an electroless metal deposition method, Appl. Phys. A **81**, 669–671 (2005)
21.233 T. Witten, L. Sander: Diffusion-limited aggregation, a kinetic critical phenomenon, Phys. Rev. Lett. **47**(19), 1400–1403 (1981)
21.234 T. Witten, L. Sander: Diffusion-limited aggregation, Phys. Rev. B **27**(9), 5686–5697 (1983)
21.235 M. Saltzmann: *Drug Delivery* (Oxford University Press, Oxford 2001)
21.236 P. Decuzzi, F. Gentile, A. Granaldi, A. Curcio, F. Causa, C. Indolfi, P. Netti, M. Ferrari: Flow chamber analysis of size effects in the adhesion of spherical particles, Int. J. Nanomed. **2**(4), 689–696 (2007)
21.237 P. Meakin: Diffusion controlled deposition on surfaces: Cluster size distribution, interface exponents, and other properties, Phys. Rev. B **30**(8), 4207–4214 (1984)
21.238 Z. Racz, T. Vicsek: Diffusion controlled deposition: Cluster statistics and scaling, Phys. Rev. Lett. **51**(26), 2382–2385 (1983)
21.239 F. Gentile, E. Battista, A. Accardo, M. Coluccio, M. Asande, G. Perozziello, G. Das, C. Liberale, F. De Angelis, P. Candeloro, P. Decuzzi, E. Di Fabrizio: Fractal structure can explain the increased hydrophobicity of nanoporous silicon films, Microelectron. Eng. **88**, 2537–2540 (2011)
21.240 F. Gentile, M.L. Coluccio, P. Candeloro, M. Barberio, G. Perozziello, M. Francardi, E. Di Fabrizio: Electroless deposition of metal nanoparticle clusters: Effect of pattern distance, J. Vacuum Sci. Technol. B **32**(2), 031804 (2014)
21.241 G. Simone, N. Malara, V. Trunzo, M. Renne, G. Perozziello, E. Di Fabrizio, A. Manz: Galectin-3 coats the membrane of breast cells and makes a signature of tumours, Mol. Biosys. **10**(2), 258–265 (2014)
21.242 G. Simone: An alternative approach to the phase change of proteins in an aqueous mixture with ethanol, Chem. Eng. Res. Design **105**, 130–136 (2016)
21.243 G. Perozziello, R. La Rocca, G. Cojoc, C. Liberale, N. Malara, G. Simone, P. Candeloro, A. Anichini, L. Tirinato, F. Gentile, M.L. Coluccio, E. Carbone, E. Di Fabrizio: Microfluidic devices module tumor cell line susceptibility to NK cell recognition, Small **8**(18), 2886–2894 (2012)
21.244 G. Perozziello, P. Candeloro, A. De Grazia, F. Esposito, M. Allione, M.L. Coluccio, R. Tallerico, I. Valpapuram, L. Tirinato, G. Das, A. Giugni, B. Torre, P. Veltri, U. Kruhne, G.D. Valle, E. Di Fabrizio: Microfluidic device for continuous single cells analysis via Raman spectroscopy enhanced by integrated plasmonic nanodimers, Opt. Express **24**(2), A180–A190 (2016)
21.245 G. Simone: Demonstrating microdroplet coalescence for tailored and biodegradable microgel fabrication, RSC Adv. **5**(70), 56848–56854 (2015)
21.246 G. Simone: Micro analysis to map the glycome code, Proteomics **14**(9), 994–1000 (2014)
21.247 R. Catalano, G. Perozziello, G. Simone, P. Candeloro, F. Gentile, M.L. Coluccio, F. Pardeo, M. Burghammer, G. Cuda, C. Riekel, E. Di Fabrizio: Optimized fabrication protocols of microflu-

idic devices for x-ray analysis, Microelectron. Eng. **124**, 13–16 (2014)

21.248 G. Keramas, G. Perozziello, O. Geschke, C.B.V. Christensen: Development of a multiplex microarray microsystem, Lab Chip **4**(2), 152–158 (2004)

21.249 G. Perozziello, P. Candeloro, F. Gentile, A. Nicastri, A.M. Perri, M.L. Coluccio, E. Parrotta, A. De Grazia, M. Tallerico, F. Pardeo, R. Catalano, G. Cuda, E. Di Fabrizio: A microfluidic dialysis device for complex biological mixture SERS analysis, Microelectron. Eng. **144**, 37–41 (2015)

21.250 G. Simone, N. Malara, V. Trunzo, G. Perozziello, P. Neuzil, M. Francardi, L. Roveda, M. Renne, U. Prati, V. Mollace, A. Manz, E. Di Fabrizio: Protein–carbohydrate complex reveals circulating metastatic cells in a microfluidic assay, Small **9**(12), 2152–2161 (2013)

21.251 G. Perozziello, P. Candeloro, F. Gentile, A. Nicastri, A. Perri, M.L. Coluccio, A. Adamo, F. Pardeo, R. Catalano, E. Parrotta, H.D. Espinosa, G. Cuda, E. Di Fabrizio: Microfluidics and nanotechnology: Towards fully integrated analytical devices for the detection of cancer biomarkers, RSC Adv. **4**(98), 55590–55598 (2014)

21.252 G. Perozziello, R. Catalano, M. Francardi, E. Rondanina, F. Pardeo, F. De Angelis, N. Malara, P. Candeloro, G. Morrone, E. Di Fabrizio: A microfluidic device integrating plasmonic nanodevices for Raman spectroscopy analysis on trapped single living cells, Microelectron. Eng. **111**, 314–319 (2013)

21.253 G. Perozziello, J. Møllenbach, S. Laursen, E. di Fabrizio, K. Gernaey, U. Krühne: Lab on a chip automates in vitro cell culturing, Microelectron. Eng. **98**, 655–658 (2012)

21.254 G. Simone, G. Perozziello: UV/VIS transparent optical waveguides fabricated using organic–inorganic nanocomposite layers, J. Nanosci. Nanotechnol. **11**, 2057–2063 (2011)

21.255 F. De Angelis, F. Gentile, F. Mecarini, G. Das, M. Moretti, P. Candeloro, M.L. Coluccio, G. Cojoc, A. Accardo, C. Liberale, R.P. Zaccaria, G. Perozziello, L. Tirinato, A. Toma, G. Cuda, R. Cingolani, E. Di Fabrizio: Breaking the diffusion limit with super-hydrophobic delivery of molecules to plasmonic nanofocusing SERS structures, Nature Photonics **5**, 682–687 (2011)

22. Biological Molecules in Therapeutic Nanodevices

Stephen C. Lee, Bharat Bhushan

In this chapter, we discuss the incorporation of molecules into nanodevices as functional device components. Our primary focus is on biological molecules, although we also discuss the use of organic molecules as functional components of supramolecular nanodevices. Our primary device interest is in devices used in human therapy and diagnosis, though when it is informative, we discuss other nontherapeutic nanodevices containing biomolecular components. We discuss design challenges associated with devices built from prefabricated components (biological macromolecules) but that are not as frequently associated with fully synthetic nanodevices. Some design challenges (abstraction of device object properties, inputs, and outputs) can be addressed using existing systems engineering approaches and tools (including unified modeling language), whereas others (selection of optimal biological macromolecules from the billions available) have not been fully addressed. We discuss various assembly strategies applicable to biological macromolecules and organic molecules (self-assembly, chemoselective conjugation) and their advantages and disadvantages. We provide an example of a functional mesoscale device, a planar field-effect transistor (FET) protein sensor, that depends on nanoscale components for its function. We also use the sensor platform to illustrate how protein and other molecular engineering approaches can address nanoscale technological problems, and argue that protein engineering is a legitimate nanotechnology in this application. In developing the functional FET sensor, both direct adsorption of protein analyte receptors as well as linkage of receptors to the sensing surface through a polymer layer were tested. However, in the realized FET sensor, interfaces consist of a polymer layer linked to the semiconductor surface and to an analyte receptor (a protein). Nanotribology and other surface-science investigations of the interfaces revealed phenomena not previously documented for nanoscale protein interfaces (lubrication by directly adsorbed proteins, increases in friction force associated with polymer-mediated increases in sample compliance). Furthermore, the studies revealed wear of polymer and receptor proteins from semiconductor surfaces by an atomic force microscopy (AFM) tip which was not a concerted process, but rather depth of wear increased with increasing load on the cantilever. These studies also revealed that the polymer–protein interfaces were disturbed by nanonewton forces, suggesting that interfaces of immunoFET protein sensors translated to in vivo use must likely be protected from, or hardened to endure, abrasion from tissue. The results demonstrate that nanoscience (in this case, nanotribology) is needed to design and characterize functional planar immunoFET sensors, even though the sensors themselves are mesoscale devices. The results further suggest that modifications made to the sensor interfaces to address these nanoscale challenges may be best accomplished by protein and interfacial engineering approaches.

22.1 **Definitions and Scope** 694
22.1.1 Design Issues 695
22.1.2 Identification of Biomolecular Components 695
22.1.3 Design Paradigms 697
22.1.4 Utility and Scope of Therapeutic Nanodevices 698

22.2 **Assembly Approaches** 700
22.2.1 Low-Throughput Construction Methods 701
22.2.2 Supramolecular Chemistry and Self-Assembly 702
22.2.3 Chemoselective Conjugation 708
22.2.4 Unnatural Amino Acids to Support Chemoselective Conjugation of Biologically Produced Proteins 709

22.3 **Sensing Devices** 710
22.3.1 Planar FET Protein Sensors 710
22.3.2 Biotechnology Approaches to the *Fundamental Limitations* of Planar ImmunoFETs 712
22.3.3 Nanotribology of Protein-Sensing Interfaces on Micromachined Surfaces . 713

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_22

22.4 **Concluding Remarks: Barriers to Practice** 717
22.4.1 You Do Not Know What You Do Not Know: The Consequences of Certainty .. 717
22.4.2 Are Proteins and Molecules Legitimately Part of Nanotechnology? .. 718
References 718

22.1 Definitions and Scope

Nanotechnology is a field in rapid flux and development, as this volume shows, and definition of its metes and bounds, as well as identification of subdisciplines embraced by it, can be difficult and controversial. The term *nanotechnology* means many things to many people, and aspects of multiple disciplines, from physics to information technology to biotechnology, legitimately fall into the intersection of the Venn diagram of disciplines that defines nanotechnology. The breadth of the field allows almost any interested party to contribute to it, but the same ambiguity can render the field diffuse and amorphous. If nanotechnology embraces everything, what then is it? Consideration of the scope of the field may be useful.

To frame the discussion, we will define nanotechnology as the discipline that aims to satisfy desired objectives using materials and devices whose valuable properties are based on a specific nanometer-scale element of their structures. As opposed to nanoscience, nanotechnology is application oriented, so nanoscience is important to nanotechnology primarily to the extent that it is relevant to device design, function or application.

The meaning of *therapeutic* is self-explanatory and refers here to intervention in *human* disease processes (although many of the approaches discussed are equally applicable to veterinary medicine). We confine our discussion mostly to therapeutics used in vivo, because such applications clearly benefit from the low invasiveness that ultrasmall, but multipotent, nanotherapeutics potentially offer. It is debatable whether imaging, diagnostic or sensing devices can be considered therapeutic in this context, though as we will see, sensing/diagnostic functionalities are often inextricable elements of therapeutic nanodevices, and it is difficult to consider smart nanotherapeutics without discussion of their sensing capabilities. Therapeutics incorporating diagnostic capabilities (via their capability for sensing or imaging contrast delivery) are now recognized as their own class of drug entities, referred to as *theranostics* (see below).

Our definition of nanotechnology is both broader and narrower than more common definitions. First, our definition embraces macroscale structures whose useful properties derive from their nanoscale aspects. Second, we have a *device-centric* bias: we are interested in devices that perform multistep work processes. Third, consistent with our device-centric bias, the term *specific* (as in, *specific nanometer-scale elements*) is intended to exclude materials whose utility derives solely from properties inherent to being finely divided (high surface-to-volume ratios, for instance), or other material, chemical, and physical properties unless those properties contribute to specific device function. We made this exclusion based on our assessment that therapeutic nanodevices are more intriguing than nanomaterials per se (see below), although we will engage these attributes where they are germane to specific device or therapeutic applications. Fourth, our definition implies that limited nanotechnology has been available since the 1970s in the form of biotechnology. Based on their nanoscale structures, individual biological macromolecules (such as proteins) often exhibit the coordinated, modular multifunctionality that is characteristic of purpose-built devices (Fig. 22.1).

Biological macromolecules rely on the deployment of specific chemical functionalities with specific relative distributions in space with nanometer (and greater) resolution for their function, so the inclusion of molecular engineering aspects of biotechnology under the nanotechnology rubric is legitimate, despite the discomfort this may cause to traditionally trained engineers. As we will see, intervention in human disease often requires inclusion of biomolecules in therapeutic devices, as frequently no functional synthetic analogue of active proteins and nucleic acids is available. As we will discuss, specific nanoscale device problems also can be addressed with biotechnology and protein engineering approaches, so the legitimacy of inclusion of biotechnology in nanotechnology is now beyond debate.

An analogous argument can be made that organic chemistry is an early form of nanotechnology. Compared with organic small molecules, protein functional capabilities and properties are generally more complex and more dependent on their conformation in three-dimensional space at nanometer and subnanometer scales. The nanotechnology sobriquet, therefore, *may* be more appropriate to biotechnology than to organic chemistry. However, supramolecular chemistry

and therapeutic supramolecular devices depend on specific design features of organic molecules and assemblies thereof, so organic chemistry might be viewed as an even earlier version of nanotechnology, based on an argument very similar to that we use for biotechnology.

As described above, this chapter focuses primarily on nanoscale therapeutic devices as opposed to therapeutic nanomaterials. Devices are integrated functional structures and not admixtures of materials, compounds or substances. Devices exhibit desirable emergent properties inherent to their design, properties that emerge as the result of the spatial and/or temporal organization, coordination, and regulation of action of individual components. The organization of components in devices allows them to perform multistep, cogent work processes that cannot be mimicked by simple admixtures of individual components. In fact, if device functions can be mimicked well by simple mixtures of components, the labor involved in configuring and constructing a nanoscale device is not warranted. Our device definition thus excludes nanomaterials used as drug formulation excipients (pharmacologically inert materials included in formulations that improve pharmacophore uptake, biodistribution, pharmacokinetic, handling, storage or other properties), but embraces those same materials as integral components of drug-delivery or other clinical devices.

22.1.1 Design Issues

The biotechnology industry historically has focused on production of individual soluble protein and nucleic acid molecules for pharmaceutical use, with only limited attention paid to functional supramolecular structures [22.2–8]. This bias toward free molecules flies in the face of the obvious importance of integrated supramolecular structures in biology and, to the casual observer, may seem an odd gap in attention and emphasis on the part of biotechnologists. The bias toward single-molecule, protein therapeutics follows from the fact that biotechnology developed as an industrial activity, governed by market considerations. Of the potential therapeutics that might be realized from biotechnology, single-protein therapeutics are among the easiest to realize from both technical and regulatory perspectives, and so warrant extensive industrial attention. This is changing, however, and more complex entities (actual supramolecular therapeutic devices) have and will appear with increasing frequency in the 21st century.

New materials derived from micro-/nanotechnology provide the opportunity to complement the traditional limits of biotechnology by providing scaffolds that can support higher-level organization of multiple biomolecules to perform work activities that could not be performed by mixtures of free, soluble molecules. Such supramolecular structures have been called nanobiotechnological devices [22.9], nanobiological devices [22.2–8], and semisynthetic nanodevices, and they figure prominently in therapeutic nanotechnology. Some such devices have sensing and diagnostic capabilities, and therefore belong to the already discussed class of theranostic agents (see [22.9] for a recent example of a multifunctional theranostic nanoparticle, and [22.10] for a recent review of such multifunctional nanoparticulate magnetic resonance imaging contrast agents in cardiovascular disease).

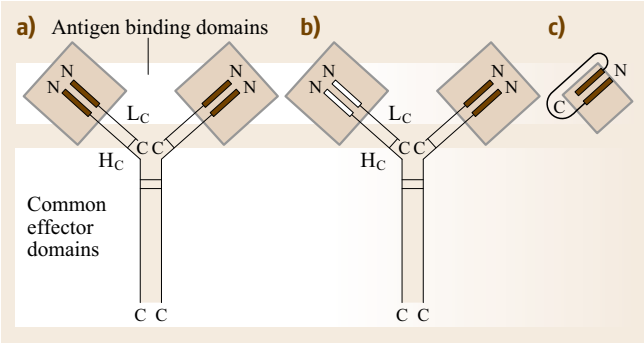


Fig. 22.1a–c Antibodies resemble purpose-built devices with distinct functional domains [22.1]. Native immunoglobulin class G (IgG) antibodies are composed of four polypeptide chains: two heavy chains (Hcs) and two light chains (Lcs), joined by interchain disulfide linkages (*lines* between Hc and Lc moieties). Amino and carboxy termini of individual polypeptide chains are indicated by N and C. Antigen-binding domains are responsible for specific antigen recognition, vary from antibody to antibody, and are indicated by the *thicker lines*. Common effector functions (Fc receptor binding, complement fixation, etc.) are delimited to domains of the antibodies that are constant from molecule to molecule. **(a)** A native IgG antibody is monospecific but bivalent in its antigen binding capacity. **(b)** An engineered, bispecific, bivalent antibody capable of recognizing two distinct antigens. **(c)** An engineered antibody fragment (single-chain fragment variable (Fv) or SCFv) that is monospecific and monovalent can recognize only one antigenic determinant and is engineered to lack common effector functions. This construct is translated as a single, continuous polypeptide chain (hence the name SCFv) because a peptide linker (indicated by the *connecting line* in the figure) is incorporated to connect the carboxy-end of the Hc fragment and the amino-end of the Lc fragment

22.1.2 Identification of Biomolecular Components

Design of nanodevices is similar to design of other engineered structures, providing that the special properties of the materials (relating to their nanoscale

aspects such as quantum, electrical, mechanical, biological properties, etc.), as well as their pharmacological properties, are considered. Therapeutics interact with patients on multiple levels, ranging from organismal to molecular, but it is reasonable to expect that most nanotherapeutics will interface with patients at the nanoscale [22.2–16]. Typically, this means interaction between therapeutics and patient biological macromolecules, supramolecular structures, organelles, cells or tissues, which in turn often dictates the incorporation of biological macromolecules (and other biostructures) into nanodevices [22.2–10]. Incorporating biological structures into (nanobiological) devices presents special challenges that do not occur in other aspects of engineering.

Unlike fully synthetic devices, semibiological nanodevices must incorporate prefabricated biological components (proteins, nucleic acids or derivatives thereof), and therefore intact nanodevices are seldom made entirely de novo. As a corollary, knowledge of properties of nanobiological device components is usually incomplete, as the molecules were not made by human design, so their properties are not known a priori and must be discovered. Therefore the range of activities inherent to any nanobiological device design may be much less obvious and less well defined than for fully synthetic devices. Further complicating the issue, the activities of biological molecules are often multifaceted (many genes and proteins exhibit pleiotropic activities), and the full range of functionality of individual biological molecules in interactions with other biological systems (as in nanotherapeutics) is often not known. This makes the design and prototyping of biological nanodevices an empirically intensive, iterative process [22.2–16].

Paralleling the paucity of information typically available about individual proteins, the number of distinct natural proteins in the biologic world is unknown but exponentially high (certainly in excess of 10^{13} distinct molecules [22.17]). When one considers engineered proteins, particularly those made by high-throughput mutagenesis methods, the number of existing protein sequences rises additional orders of magnitude. Most of these molecules remain to be discovered, so their individual properties (that might be critical to device designers, such as functional pH and temperature ranges, ionic requirements, cofactor requirements, radiation tolerance, resistance to degradation, etc.) are mostly unknown. Among the relative handful of proteins that are known, properties are typically incompletely known, and those properties that are known often are not those of greatest interest in selecting a protein as a nanodevice component. In fact, existing protein databases focus primarily on pharmacological properties or evolutionary relationships between proteins. We have proposed the building of databases useful specifically for nanobiotechnology, though this has yet to occur [22.5, 17, 18]. Existing protein databases, based as they are on phylogenetic, protein sequence, protein structure or protein primary function information, are not satisfactory for nanodevice design, if such design is to be realized as a discipline in and of itself, and if the immense power of biologic nanotechnology is to be realized to any significant extent. As it is, inventors of nanobiotechnological devices often rely on their personal knowledge of biology and biochemistry to select biological components for devices. It is obvious then, that the devices designed by even the most sophisticated biologists are very likely to be suboptimal. As a corollary, this means that nanobiological device designers must have significant biological expertise. This constitutes a huge barrier to entry for nonbiologists who could otherwise contribute importantly to the field. Thus, the lack of an appropriate protein database and accompanying search tools implies an immense opportunity cost for the field of biologic nanotechnology.

Biological macromolecules have properties, particularly those relating to their stability, that can limit their use in device contexts. In general, proteins, nucleic acids, lipids, and other biomolecules are more labile to physical/biochemical insult than are many synthetic materials. With the possible exceptions of topical agents or oral delivery and endosomal uptake of nanotherapeutics (both involving exposure to low pH), device lability in the face of *physical* insult is generally a major consideration only in ex vivo settings (relating to device storage, sterilization, etc.), because physical conditions that would destroy the device would be bad for the patient as well. However, living organisms remodel themselves constantly in response to stress, development, pathology, and external stimuli. For instance, epithelial tissues and blood components are constantly eliminated and regenerated, and bone and vasculature are continuously remodeled. The metabolic facilities responsible (circulating and tissue-bound proteases and other enzymes, various clearance organs, the immune system, etc.) can potentially process biological components of nanobiological therapeutic devices as well as endogenous materials, leading to partial or complete degradation of nanotherapeutic structure, function, or both. Furthermore, immune and wound responses protect the host against pathogenic organism incursions, by mechanisms that involve sequestering and degrading pathogens. Nanobiological therapeutics are subject to the actions of these host defense systems as well as normal remodeling processes. Various strategies to stabilize biomolecules and structures in heterologous in vivo environments are applicable to nanobiological therapeutics. Conversely, instability of

active biocomponents can offer a valuable and simple way to delimit the activity of nanotherapeutics containing biomolecules.

22.1.3 Design Paradigms

Several early attempts to codify the canonical properties of ideal nanobiological devices, and therapeutic nanodevices in particular, have been made [22.1, 5, 11–15, 19, 20]. These attempts to codify design constitute a limited set of design guidelines that are summarized in Table 22.1. Naturally occurring, functional biological components generally exist in the context of higher-order systems that support the organisms of which they are a part. In general, nanobiological devices contain biological components that retain their function in new (device) contexts. In other words, one must incorporate into the device enough of a functional biological unit (nucleic acid, protein, oligomeric protein complex, organelle, cell, etc.) to allow that unit to perform the function for which it was selected. If one wishes, for example, to appropriate the specific antigen-recognition property of an antibody for a device function (say, in targeting, as discussed later), it is not necessary to incorporate the entire 150 000 atomic mass unit (AMU) antibody, the bulk of which is devoted to functions other than antigen recognition (Fig. 22.1) [22.21], but it *is* critical to incorporate the approximately 20 000 AMU of the antibody essential for specific antibody–antigen binding (the variable domain, Fig. 22.1).

Device function is the result of the summed and various activities of biological and synthetic device components. The nanobiological device designer can exert control over the relative organization of biological device components, which allows biomolecular components of devices abstracted from their native context to contribute to overall device functions that are entirely different from those in which they participated in their original, organismal contexts.

All of these features are illustrated in the bacmid, or Bac-to-Bac, system, a commercially available molecular cloning device ([22.22–24] and Fig. 22.2). Bacmid configures prokaryotic and eukaryotic genetic elements from multiple sources into a device for producing recombinant eukaryotic viral genomes in bacteria, a function that is unprecedented in nature. The system is feasible because of the modularity of the genetic elements involved and because of the strict control of the relative arrangement of genetic elements allowed by recombinant deoxyribonucleic acid (DNA) technology. Analogous cloning devices based on comparable arrangements of bacterial and eukaryotic regulatory and structural genes are reviewed in [22.18]. Other nucleic acid devices using genetic control elements from phylogenetically various sources are being developed to preprogram the micro- and nanoscale architectural properties and physiological behavior of living things [22.25, 26].

Bacmid provides an example of a nontherapeutic nanobiological device and illustrates some specific design approaches for building functional devices with biocomponents. Bacmid complies fully with those design recommendations of Table 22.1 that are not therapeutic device specific. However, Table 22.1 provides guidelines only. Systems engineering approaches commonly used in software and computer engineering provide a more rigorous framework to consider nanodevice design [22.17, 18]. Systems tools such as unified modeling language (UML) allow more precise depiction of nanobiological devices than do most text descriptions. A simplified UML use-case of bacmid is presented in Fig. 22.3.

Table 22.1 Some ideal characteristics of nanodevices [22.1, 5, 11–15, 19, 20]

(a) Characteristics of all nanobiological devices
Biological molecules must retain function.
Device function is the result of the summed activities of device components.
The relative organization of device components drives device function.
Device functions can be unprecedented in the biological world.
(b) Desirable characteristics of therapeutic platforms
Therapeutics should be minimally invasive.
Therapeutics should have the capacity to target sites of disease.
Therapeutics should be able to sense disease states in order to: • Report conditions at the disease site to clinicians • Administer metered therapeutic interventions.
Therapeutic functions should be segregated into standardized modules.
Modules should be interchangeable to tune therapeutic function.

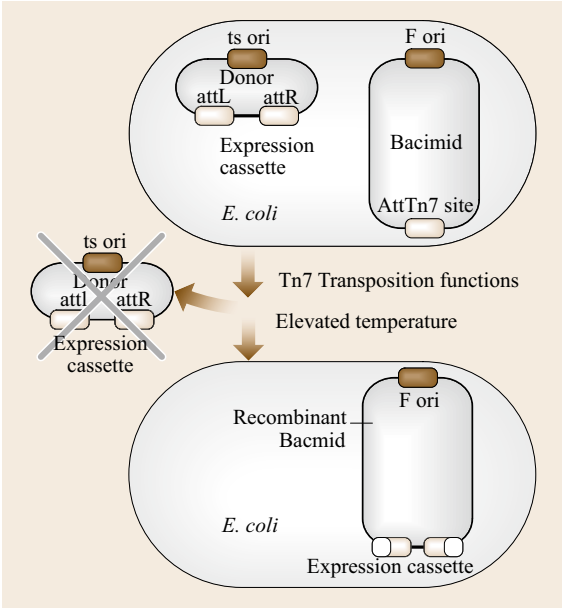


Fig. 22.2 The bacmid molecular cloning system as it is represented in the molecular biology literature. Bacmid is a molecular device designed to allow efficient production of recombinant insect viruses (baculovirus) in *Escherichia coli* [22.22–24]. Baculovirus is replicated in *E. coli* by the F plasmid origin of replication (*F ori*), and as such, is called a bacmid. The bacmid also includes an engineered transposable DNA element 7 (Tn7) attachment site isolated from the chromosome of an enteric bacteria (*Att*Tn7). *Att*Tn7 can receive Tn7 elements transposed from other cellular locations. A donor plasmid (donor) is replicated by a temperature-sensitive plasmid pSC101 origin of replication (*ts ori*). The donor also incorporates an expression cassette containing both the gene of interest for ultimate expression in insect cells and a selectable genetic marker operable in *E. coli*. The expression cassette is flanked by DNA sequences (*att*L and *att*R) that are recognized by the Tn7 transposition machinery. Tn7 transposition machinery resides elsewhere in the same *E. coli* cell. When donor plasmid is introduced into *E. coli* containing bacmid, Tn7 transposition machinery causes the physical relocation of expression cassettes from donor plasmid to bacmid. Unreacted donor plasmid is conveniently removed by elevating the incubation temperature, causing the ts pSC101 replicon to cease to function and, in turn, causing the donor to be lost. If selection for the genetic markers within the expression cassette is applied at this point, the only *E. coli* that survive are those containing recombinant bacmid (i. e., those that have received the gene for insect cell expression by transposition from the donor). Recombinant bacmid are conveniently isolated from *E. coli* and introduced into insect cell culture, where expression of the gene of interest occurs (after [22.25])

Most importantly, UML forces designers to abstract knowledge of component (object) functions and properties and to express them in terms of object inputs and outputs. This makes device designs modular by explicitly identifying object properties and limitations: any object producing the same outputs and accepting the same inputs can be substituted for the original component in the functional nanodevice. By this means, UML notation transforms nanodevice designs from one-off individual designs into general designs for a class of functionally similar devices [22.18]. In general, biologist inventors have intuitively used these systems design approaches to design devices, as we discuss [22.18]. However, UML-enforced explicit callout of component properties that are necessary and sufficient for device function makes design of isofunctional devices easier. An argument might be made that, in some cases, UML expression of a device design, an example of which is shown in Fig. 22.3, may render related devices (that are implied in the UML depiction) obvious. If so, broad use of UML in nanobiotechnology might have significant intellectual property implications.

In nanotherapeutic applications, devices should be noninvasive and target therapeutic payloads to sites of disease to maximize therapeutic benefit while minimizing undesired side-effects. This, of course, implies the existence of therapeutic effector functions in these nanodevices, to give devices the ability to remediate a physiologically undesirable condition. Beyond that, several desirable attributes relate to sensing of biomolecules, cells or physical conditions (sensing disease itself, identification of residual disease, and potentially responding to intrinsic or externally supplied triggers for payload release). Other properties relate to communication between device subunits (for instance, between sensor and effector domains of the device) or between the device and an external operator. With appropriate design, device functions can be modular, as illustrated by the early, hypothetical dendrimer-based therapeutic shown in Fig. 22.4a [22.11, 16, 27–29].

22.1.4 Utility and Scope of Therapeutic Nanodevices

Therapeutic nanotechnology will be useful when the underlying biology of the disease states involved is amenable to intervention at the nanoscale. While several disease states and physiological conditions (cancer, vaccination, cardiovascular disease, etc.) are particularly accessible to nanoscale interventions, some nanotechnological approaches may be applicable more broadly, in indications we currently cannot predict. Much as was the case with the introduction of recombi-

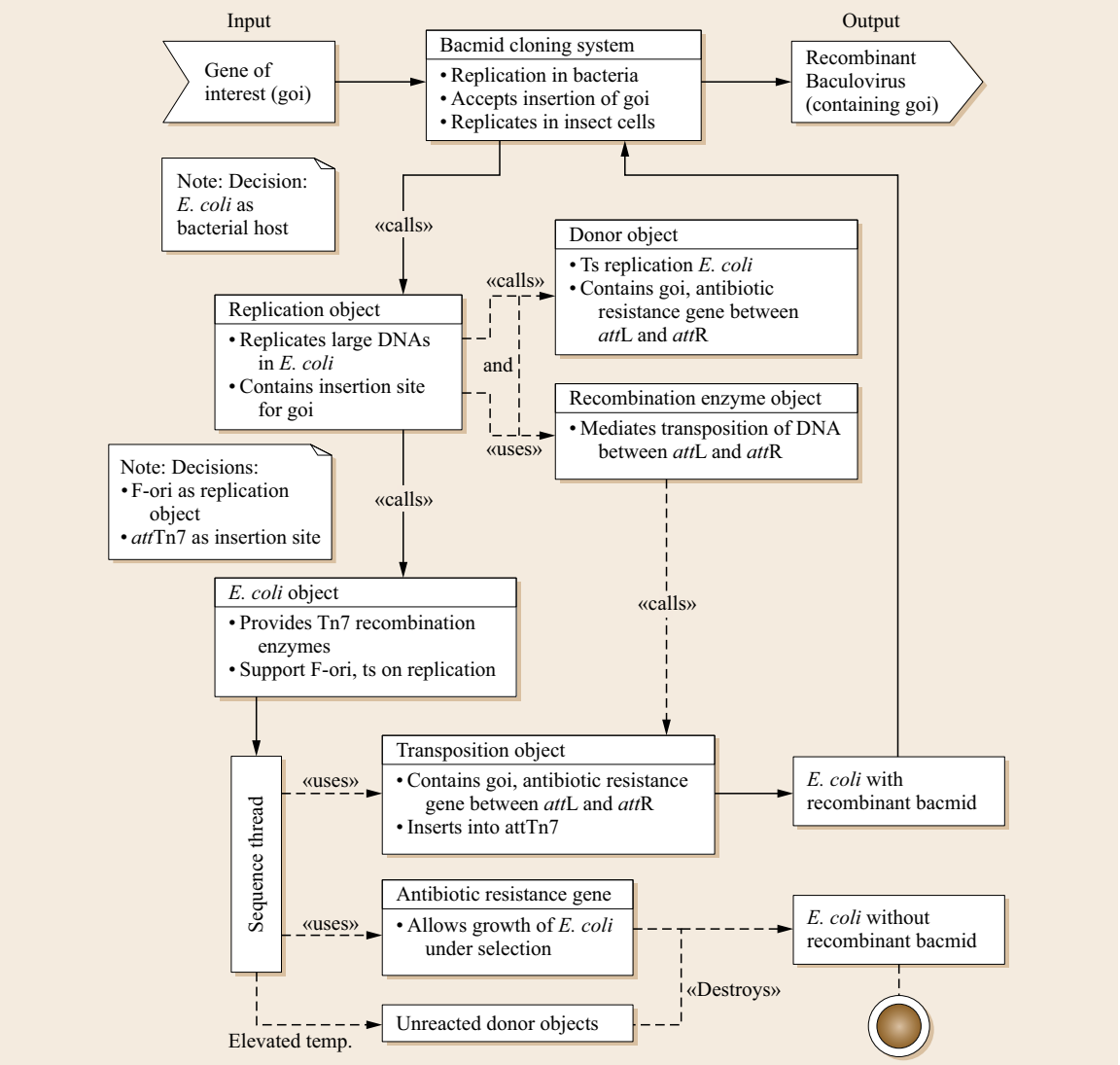


Fig. 22.3 The bacmid system represented as a use-case with unified modeling language (UML). Objects are represented by *boxes*, split horizontally into compartments. The *topmost compartment* contains the object name, the *middle compartment* contains the object attributes, and the *bottom compartment* contains the object operations (if different from the object attributes). Notes contain expository information such as design decisions. Sequence threads indicate serial events in process flow. Process flow is indicated by *solid arrows*. *Dashed arrows* indicate communication between objects and, for the purposes of illustrating bacmid, are labeled as *uses* (the object utilizes a second object), *calls* (the object elicits an action from a second object), and *destroys* (a step that causes the loss or destruction of one or more objects). End of process is indicated. The abstraction of the function of individual objects driven by UML may make other analogous devices (similar to bacmid, but substituting one or more isofunctional objects) obvious [22.17, 18]. In protein-containing devices, systematic identification of isofunctional objects could be facilitated using protein databases modified as suggested in [22.17] (after [22.18])

nant protein therapeutics, nanotherapeutics may present regulatory and pharmacoeconomic challenges related to their novelty and their cost of goods (COGs). However, there is little doubt that nanobiological devices providing clear patient benefit, and whose production, regulatory approval, and distribution are amenable to feasible business models, will enter clinical practice.

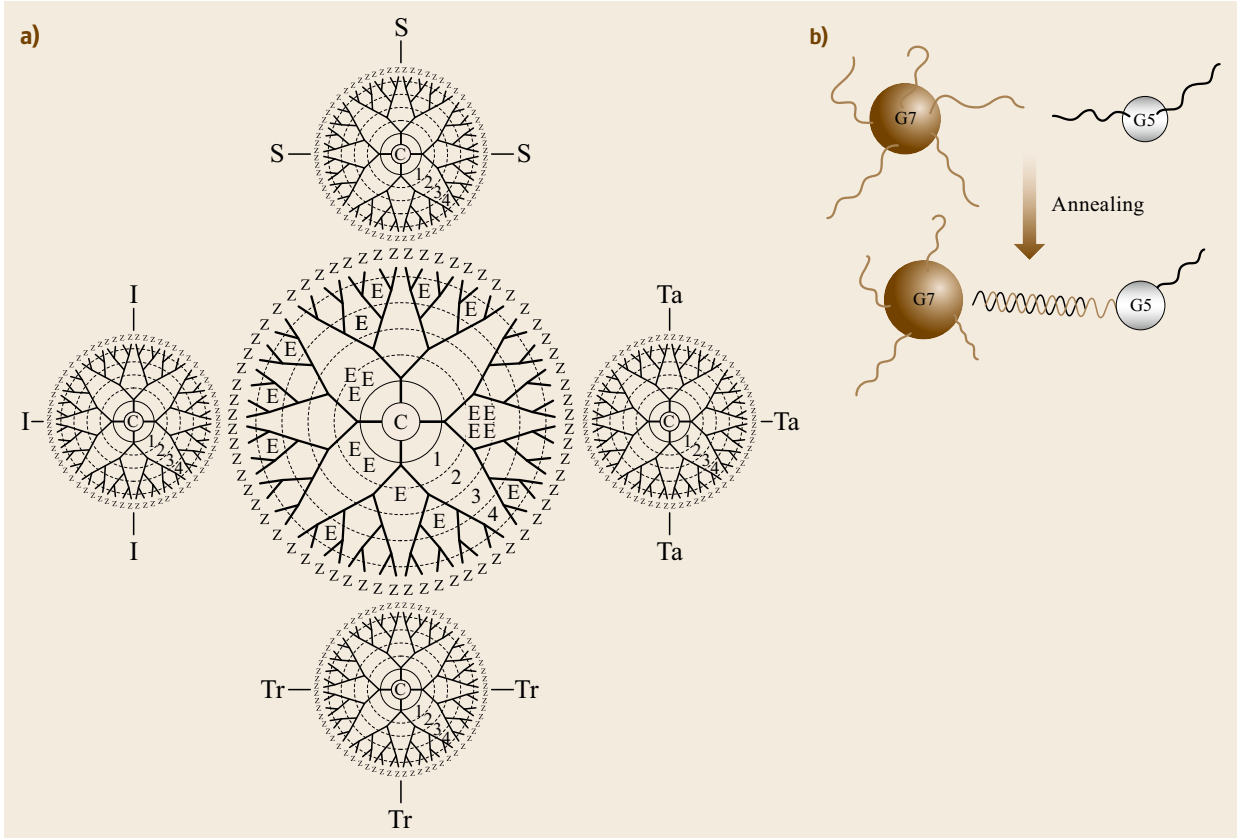


Fig. 22.4 (a) A hypothetical, modular nanotherapeutic agent patterned after the dendrimer-based cluster agent for oncology of *Baker* [22.27–29]. Self-assembling, multidendrimer structures as shown are sometimes referred to as tecto(dendrimers). Each polyamidoamine (PAMAM) dendrimer subunit is grown from an initiator core (C), and the tunable surface groups of the dendrimers are represented by Z. In this device, each dendrimer subunit has a specific, dedicated function in the device: the central dendrimer encapsulates small-molecule therapeutics (E), whereas other functional components are segregated to other dendrimer components. These include biochemical targeting/tethering functions (Ta), therapeutic triggering functions to allow activation of prodrug portions of the device by an external operator (Tr), metal or other constituents for imaging (I), and sensing functions (S) to mediate intrinsically controlled activation or release of therapeutic. This design constitutes a therapeutic platform [22.14, 16] because of its modular design. The depicted device is only one possible configuration of an almost infinite number of analogous therapeutics that can be tuned to fit particular therapeutic needs by interchanging functional modules (after [22.16]). (b) A more sophisticated assembly strategy for dendrimer therapeutics that utilizes self-assembly of biomolecules [22.30–32]. Individual dendritic polymers (spheres representing generation 5 and generation 7 PAMAM dendrimers) are conjugated with single-stranded (ss) oligonucleotides (*light and dark gray lines*). When oligonucleotides of two dendritic polymers are complementary, individual oligos hybridize, forming double-stranded (ds) DNA complexes, linking the two dendritic polymers. An interesting feature of this assembly system is the rigidity of short dsDNA (< 50 base pair) segments. This allows assembly of objects at precise, tunable nanoscale distances (after [22.30])

22.2 Assembly Approaches

Assembly of components into devices is amenable to multiple approaches. In the case of devices comprising a single molecule or processed from a single crystal (some microfabricated structures, single polymers, or grafted polymeric structures) assembly may not be an issue. Integration of multiple, separately microfabricated components may sometimes be necessary and may sometimes drive the need for assembly, even for silicon devices. Furthermore, many therapeutic nanodevices contain multiple, chemically diverse components

that must be assembled precisely to support their harmonious contribution to device function.

22.2.1 Low-Throughput Construction Methods

Low-throughput device construction methods are more applicable to construction of prototypes for research purposes than to production of commercial therapeutics. For example, direct-write technologies can obtain high (nanometer-scale) resolution; electron-beam (e-beam) lithography is a technique requiring no mask and can yield resolutions on the order of tens of nanometers, depending on the resist materials used [22.33]. Resolution in e-beam lithography is ultimately limited by electron scattering in the resist and electron optics, and like most direct-write approaches, e-beam lithography is limited in its throughput. Parallel approaches involving simultaneous writing with up to 1000 shaped e-beams are under development [22.33] and may mitigate limitations in manufacturing rate.

Atomic force microscopy (AFM) approaches utilize an ultrafine cantilever tip (typically with tip diameters of 50 nm or less) in contact with, or tapping, a surface or stage. The technique can be used to image molecules, analyze molecular biochemical properties (such as ligand–receptor affinity [22.34]), or manipulate materials at the nanoscale. In the latter mode, force microscopy has been used to manipulate atoms to build individual nanostructures since the mid-1980s, though the manufacturing throughput of *manual* placement of atoms and nanoscale components by force microscopy is limited, even with highly multiplexed arrays of cantilevers.

Dip-pen nanolithography (DPN) is a force microscopy methodology that can achieve high-resolution features (features of 100 nm or less) in a single step. In DPN, the AFM tip is coated with molecules to be deployed on a surface, and the molecules are transferred from the AFM tip to the surface as the coated tip contacts it. DPN also can be used to functionalize surfaces with two or more constituents and is well suited for deployment of functional biomolecules on synthetic surfaces with nanoscale precision [22.35, 36]. DPN suffers the limitations of synthetic throughput typical of AFM construction strategies.

Much as multibeam strategies might improve throughput in e-beam lithography [22.34], multiple tandem probes may significantly increase assembly throughput for construction methods that depend on force microscopy, but probably not sufficiently to allow manufacture of bulk quantities of nanostructures, as will likely be needed for consumer nanotherapeutic devices. As standard of care evolves increasingly toward tailored courses of therapy [22.37], individual therapeutics will become increasingly multicapable and powerful. Potentially, fewer copies of a nanotherapeutic may be required per patient, and each patient's nanotherapeutic may be tailored to him/her. It is not inconceivable that this might make relatively low-throughput synthesis/assembly methods practical, though this remains to be seen. If tailored therapeutics become the standard, the pharmaceutical industry will be irrevocably changed, with the concept of the *blockbuster* that can produce multiple billions of dollars in revenue per year through sales of a single agent to a broad population made no longer relevant. Business models in the pharmaceutical industry may be changed beyond recognition, and in the fullness of time, pharmacies could begin to resemble the formularies of old, with the capacity to make a specific preparation for an individual patient on site, but now using sophisticated bio- and nanotechnologies.

Individualized therapies for one patient are likely to differ from those for other patients by their incorporation of engineered molecules that are immunologically, toxicologically or physiologically tolerable by each individual host. Device functional properties will be tuned to integrate with individual host physiologies. One facile way to achieve patient-specific theranostics might be to design a modular device whose properties are tuned by substitution of biochemically isofunctional components, but whose other properties (immunological, toxicological, etc.) are suited to individual patients. This sounds like the modular nanodevice construction strategy we advocate [22.17, 18]. Should this come to pass, pharmaceutical companies will have to find a way to claim as intellectual property immense collections of similar function, but individually tailored, nanotherapeutic devices. Patenting each device individually is economically and logistically impossible, and a patent on a single specific nanotherapeutic might be easily *invented around* by competitors. One possible solution may be to structure patent claims broadly, around UML representations of nanotherapeutics that explicitly call out the critical characteristics of objects to identify alternate, isofunctional objects that might be substituted for the objects of the device in its original design [22.18]. This may capture a group of analogous, related theranostics, rather than an individual nanotherapeutic, or it may make some or all of the analogous devices obvious in the eyes of the patent examiner. This would likely preclude the proliferation of devices, each captured by its own patent, but template off a single central invention, as occurred with analogues of bacmid [22.18].

For the moment, though, ideal manufacturing approaches for nanotherapeutic devices resemble either

industrial polymer chemistry, occurring in bulk, in convenient buffer systems, or in massively parallel industrial microfabrication approaches. In any case, therapy for a single patient may involve multiple billions of individual nanotherapeutic units, so each individual nanotherapeutic structure must require only minimal input from a human technician. Self-assembly, when feasible, allows device construction without ongoing human intervention.

22.2.2 Supramolecular Chemistry and Self-Assembly

Exploitation of the self-assembly properties of dendrimers and its exploitation to build a modular therapeutic has already been alluded to in Fig. 22.4a. Self-assembly has long been recognized as a potentially critical labor-saving approach to the construction of nanostructures [22.38], and many organic and inorganic materials have self-assembly properties that can be exploited to build structures with controlled configurations. Self-assembly processes are usually driven by thermodynamic forces and generally result in structures that are not covalently linked. Intra-/intermolecular forces driving assembly can be electrostatic or hydrophobic interactions, hydrogen bonds, and van der Waals interactions between and within subunits of the self-assembling structures or the assembly environment. Thus, final configurations are limited by the ability to *tune* the properties of the subunits and control the assembly environment to generate particular structures.

Highly hydrophobic carbon nanotubes spontaneously assemble into higher-order [22.39] structures (nanoropes and multiwall carbon nanotubes) as the result of hydrophobic interactions between individual tubes [22.39]. C_{60} fullerenes and single-wall carbon nanotubes (SWCNT) also spontaneously assemble (Fig. 22.5) into higher-order nanostructures called *peapods* [22.40], in which fullerene molecules are encapsulated in nanotubes. The fullerenes of peapods modulate the local electronic properties of the SWCNT in which they are encapsulated and may allow tuning of carbon nanotube electrical properties, perhaps as in one-dimensional (1-D) carbon nanotube field-effect transistors (FETs) [22.41]. Several self-assembled carbon structures and carbon-structure-containing devices are depicted in Fig. 22.5.

In drug delivery, the most familiar self-assembled nanostructures are micelles [22.42–44]. These structures are formed from the association of block copolymer subunits (Fig. 22.6), each individual subunit containing hydrophobic and hydrophilic domains. Micelles spontaneously form when the concentration of their subunits exceeds the critical micelle concentration (CMC) in a solvent in which one of the polymeric domains is insoluble (Fig. 22.6). The CMC is determined by the insoluble polymeric domain, and can be adjusted by control of the chemistry and length of the immiscible domain, as well as by control of solvent conditions. Micelles formed at low concentrations from low-CMC polymers are stable at high dilution. Micelles formed from polymer monomers with high CMCs can dissociate upon dilution, a phenomenon that might be exploited to control release of therapeutic cargos. If desired, micelles can be stabilized by covalent cross-linking to generate shell-stabilized structures [22.42–44]. Size dispersity and other properties of micelles can be manipulated by control of solvent conditions, incorporation of excipients (to modulate polymer packing properties), temperature, and agitation. From the standpoint of size, reasonably monodisperse preparations (polydispersities of 1−5%) of nanoscale micellar structures can be prepared.

The immense versatility of industrial polymer chemistry allows micellar structures to be tuned chemically to suit the task at hand. They can be modified for targeting (by appending ligands that recognize particular targeting sites to their surfaces) or to support higher-order assembly of micelles. They can be made to imbibe therapeutic or other molecules for delivery and caused to dissociate or disgorge themselves of payloads at desired times or bodily sites under the influence of local physical or chemical conditions. The tunability of these and other properties at the level of monomeric polymer subunits (as well as the level of assembled higher-order structures) makes micelles potentially powerful nanoscale vehicles for the delivery of drugs or imaging contrast agents.

In its most sophisticated manifestation, synthetic polymer self-assembly strategies can allow the generation of three-dimensional structures of highly defined nanoscale morphology. The production of controlled self-assembled structures can be affected by synthesis and assemble specific organic chemical compounds of controlled structure and chirality. Nanoscale molecular assemblies of this nature are said to be the result of supramolecular chemistries. Supramolecular chemistries exploit designer knowledge of molecular geometries, intramolecular interactions (hydrophobic, metal chelation, hydrogen bonding, and dipole interactions), and intramolecular packing properties to drive formation of nanoscale structures of controlled shape [22.45, 46]. As the referenced reviews [22.45, 46] show, supramolecular chemistry can produce space-filling nanostructures of stunning regularity, beauty, and elegance. However, what is most exciting in the context of therapeutic nanodevices (in our opinion) is the

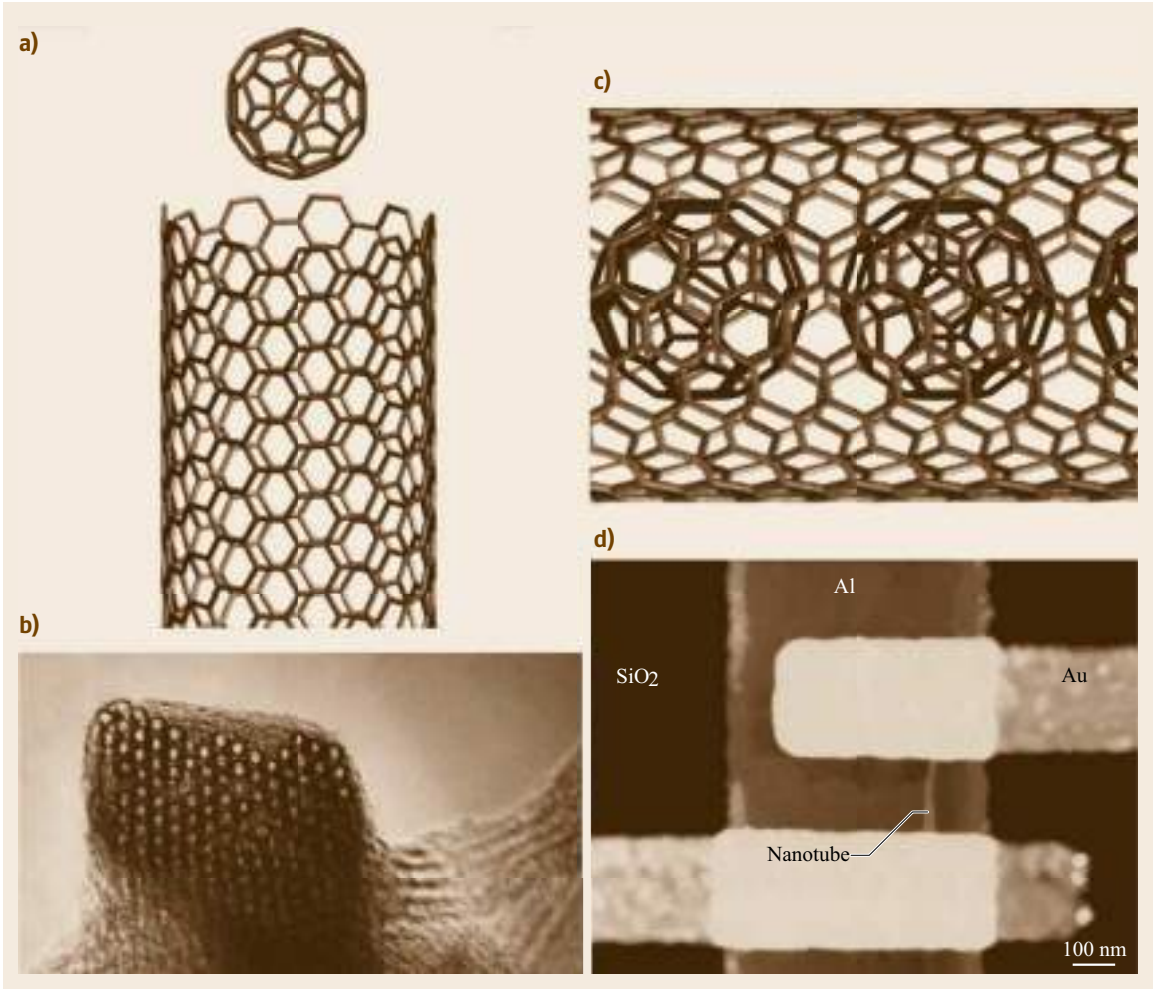


Fig. 22.5 **(a)** Shown to scale are two highly defined carbon nanostructures: a C_{60} fullerene and (10,10) single-wall carbon nanotube (SWCNT). **(b)** A self-assembled nanorope composed of carbon nanotubes that assemble by virtue of hydrophobic interactions [22.39–41]. **(c)** Schematic depiction of another self-assembled carbon nanostructure (a peapod) consisting of the fullerenes and SWCNT of **(a)**, with the fullerenes encapsulated in the SWCNT [22.40]. Fullerene encapsulation in the peapod modulates the local electronic properties of the SWCNT. **(d)** A nanotube field-effect transistor (FET) consisting of gold source and drain electrodes on an aluminum stage with a carbon nanotube serving as the FET channel [22.41] (after [22.39–41])

tunable biological activity of some supramolecular assemblies.

Incorporation of pharmacophores into supramolecular structures can render the structures biologically active. *Tysseling-Mattiace* and colleagues recently reported a supramolecular assembly (IKVAV peptide amphiphile, or IKVAV PA) with biologic activity in vitro and in vivo [22.47, 48]. IKVAV PA is based on amphiphilic polymer monomers, as are micelles; but unlike micelles, IKVAV PA is assembled from polymeric monomers whose chemistries are tuned to allow them to preferentially assemble into linear nanoscale filaments rather than spheres (Fig. 22.7). IKVAV PA nanofibers further coalesce into a gel under physiological salt and pH conditions. Thus, aqueous solutions of IKVAV PA spontaneously form a gel when injected in vivo.

The designers incorporated the peptide isoleucine–lysine–valine–alanine–valine (IKVAV) into the polar head group of each amphiphilic polymer monomer. IKVAV is a neuroactive peptide derived from the protein laminin, and is presented at immensely high valency on the surface of assembled IKVAV PA nanofilaments. In the central nervous system, IKVAV peptide is known to inhibit differentiation to glial cells and promote neu-

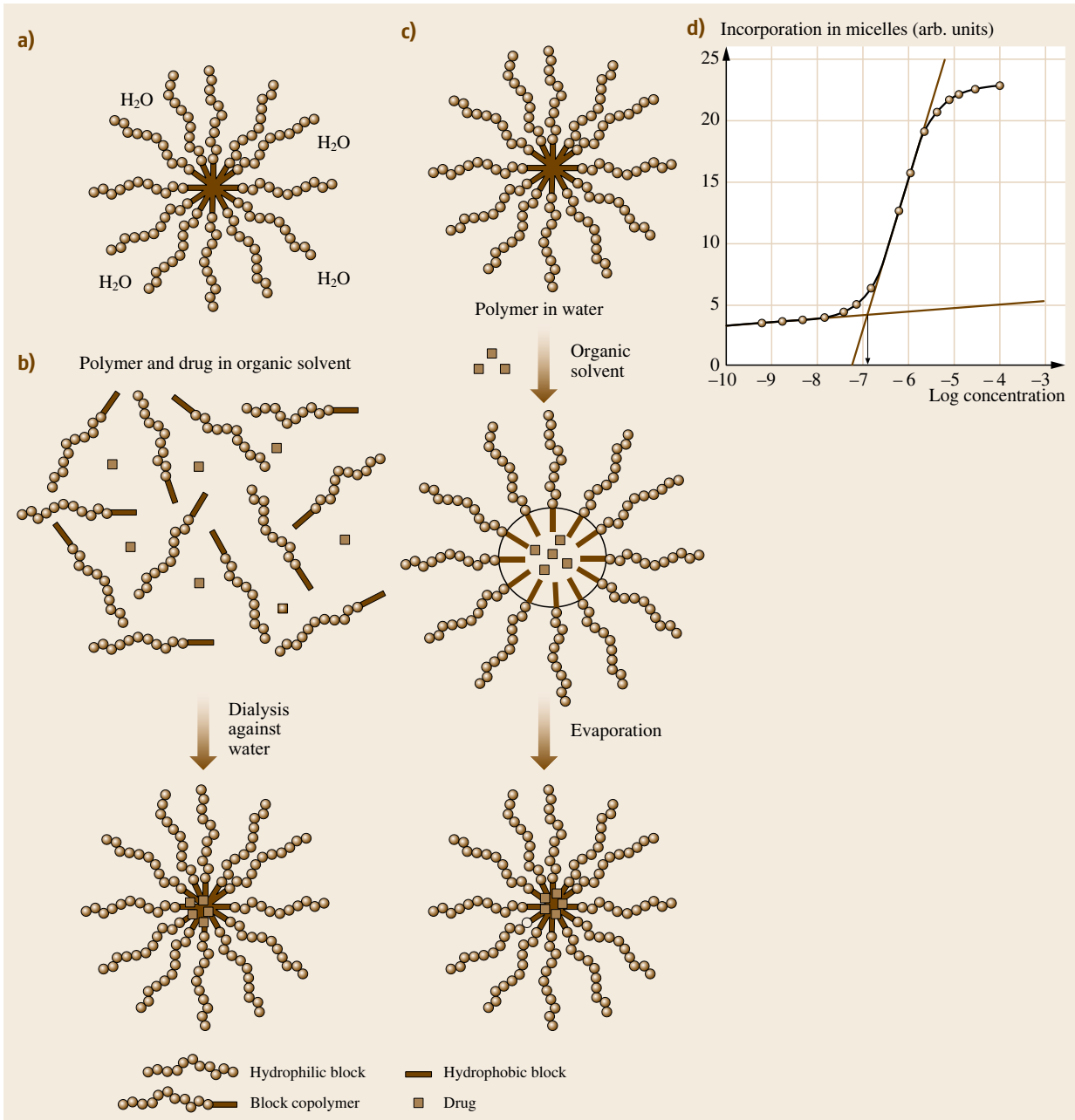


Fig. 22.6a–d Micellar drug-delivery vehicles and their self-assembly from block copolymers [22.42–44]. **(a)** Morphology of a micelle in aqueous buffer. Hydrophobic and hydrophilic polymer blocks, copolymers containing the hydrophilic polymer blocks, copolymers containing the blocks, micelles generated from the block copolymers, and (hydrophobic) drugs for encapsulation in the micelles are indicated. **(b)** Micelle self-assembly and charging with drug occurring simultaneously when the drug–polymer formulation is transitioned from organic to aqueous solvent by dialysis. **(c)** Preformed micelles can be passively imbibed with drugs in organic solvent. Organic solvent is then removed by evaporation, resulting in compression of the (now) drug-bearing hydrophobic core of the micelle. **(d)** An illustration of concentration-driven micelle formation. At and above the critical micelle concentration (CMC), block copolymer monomers assemble into micelles rather than exist as free block copolymer molecules. The *arrow* indicates the CMC for this system (after [22.42–44])

ronal outgrowth. Glial cells have a primary role in laying down scar tissue after spinal cord injury, and scar tissue is a key inhibitor of neurite outgrowth to repair the break.

In a murine model of induced spinal cord injury, IKVAV PA was injected into the site of spinal breaks. Relative to controls, injected animals exhibited reduced neurite apoptosis, reduced glial cell differentiation and scar tissue formation, and increased transit of the site of injury by descending (motor) neurons and by ascending (sensory) neurons (Fig. 22.7). These histological improvements in treated animals were accompanied by behavioral improvements (significant recovery of injury-induced motor and sensory deficits) that persisted longitudinally after the IKVAV PA was known to have been cleared from the site of injury. These improvements in motor and sensory function were unprecedented in the spinal injury model, and will be highly significant if translated to human therapy.

Furthermore, equimolar amounts of IKVAV peptide injected into the sites of injury produced no significant neurological improvements relative to control animals. That is, the structure provided by the amphiphilic polymer assembly is necessary to IKVAV PA therapeutic function. Consequently, IKVAV PA fits our definition of a nanodevice, and consists of an assembly module (the amphiphilic polymer monomers) and a bioactive module (IKVAV).

Design of self-assembling polymeric units that form specific nanoscale structures requires organic chemical synthetic capabilities and modeling skills that are not widely distributed in the physical and biological nanotechnology communities, but they do constitute genuine nanotechnologies, in as much as supramolecular chemistry facilitates multicomponent functional structures whose functions depend on structurally defined nanoscale components. The potential of supramolecular structures in therapy is immense.

Supramolecular chemistry of synthetic peptides has provided pharmacologically active antimicrobial structures (Fig. 22.8) [22.49, 50]. In the anti-infective architecture, individual peptide components are flat, circular molecules. The planar character of the toroidal subunits is a consequence of the alternating chirality of alternating DL amino acids (AAs) in the primary sequence of the peptide rings, currently only possible for synthetic peptides. Ribosomes recognize and incorporate into nascent polypeptides only L amino acids, and so, as a result of AA chirality and bond strain, peptides made by ribosomes cannot be flat, closed toroids like those of the peptidyl anti-infective agents. Alternation of D and L AAs is not possible in proteins made by ribosomal synthesis, though as we discuss below, this may

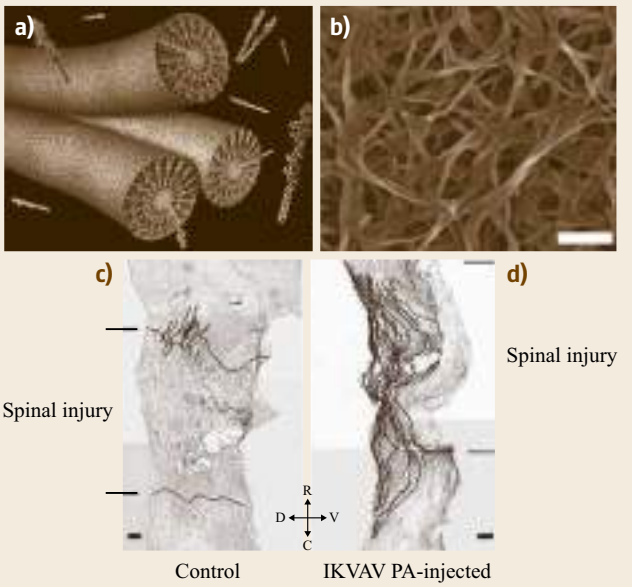


Fig. 22.7a–d The self-assembling neuroactive agent IKVAV PA (IKVAV peptide amphiphile) promotes regeneration of axons across sites of mechanically induced spinal injury in mice. **(a)** Schematic representation showing individual PA molecules assembled into a bundle of nanofibers interwoven to produce IKVAV PA. **(b)** Scanning electron micrograph image showing the network of nanofibers in vitro; *scale bar* in **(b)** indicates 200 nm. **(c,d)** Representative tracings of descending motor axon fibers within a distance of 500 μm rostral of the lesion in vehicle-injected **(c)** and IKVAV PA-injected **(d)** mice. The *dotted lines* demarcate the borders of the area of spinal cord injury. *Colored lines* indicate descending motor neurons impinging on the lesion. *Scale bars* in **(c,d)** indicate 100 μm. R – rostral; C – caudal; D – dorsal; V – ventral. IKVAV PA also promotes lesser, but still significant, regeneration of sensory neurons (not shown [22.48]). IKVAV PA-induced improvements in regeneration of axons are associated with significant behavioral improvements in mice. The self-assembly properties of IKVAV PA is essential to its function: IKVAV peptide alone does not promote regeneration [22.47, 48] (after [22.48])

change. Much as in α-helical domains of ribosomally synthesized proteins, however, the AA R-groups (which are of varying hydrophobic or hydrophilic chemical specificities) are arranged in the plane of the closed D, L rings, extending out from the center of the rings. Hydrogen bonds between individual rings govern self-assembly of the toroids into rod-like stacks, while the R-groups dominate interactions between multiple stacks of toroids and other macromolecules and structures. The planar toroidal subunits can be administered as monomers and self-assemble into multitoroid rods (Fig. 22.8) at the desired site of action (in biological membranes). The peptide toroid R-groups are

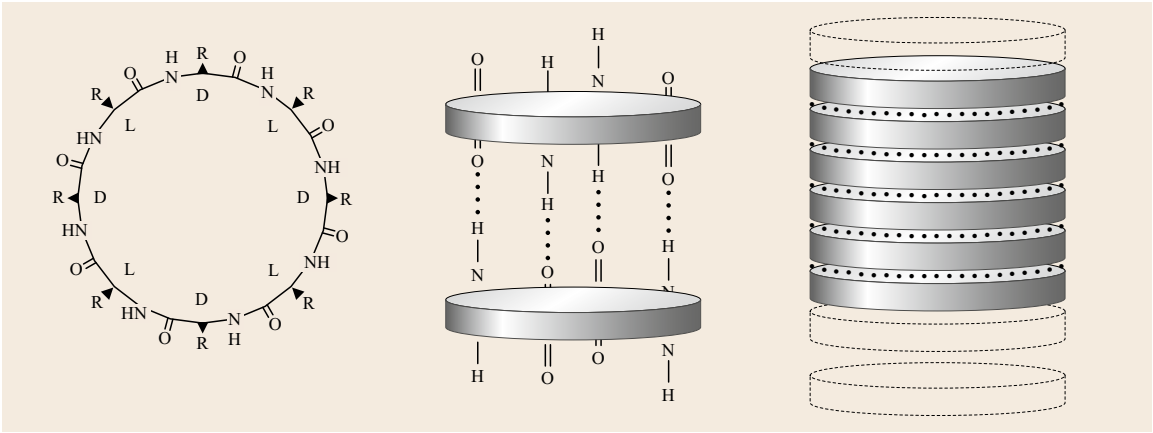


Fig. 22.8 A self-assembling peptide antibiotic nanostructure [22.49, 50]. Peptide linkages and the α-carbons and their pendant R-groups are indicated. The synthetic peptide rings are planar as a result of the alternating chirality (D or L) of their amino acid (AA) constituents. R-groups of AAs radiate out from the center of the toroid structure. Individual toroids self-assemble (stack) as a result of hydrogen-bonding interaction between amine and carboxy groups of the peptide backbones of adjacent toroids. The surface chemistry of multitoroid stacks is tuned at the level of the AA sequence and, therefore, R-group content of the synthetic peptide rings. The chemical properties of the stacked toroid surfaces allow them to intercalate into the membranes of pathogenic organisms, with lethal consequences. The specific membrane preferences for intercalation of the compound are tuned by control of the R-group content of the toroids (after [22.49, 50])

chemically tuned so that the rod structures into which they spontaneously assemble intercalate preferentially in specific lipid bilayers (i. e., in pathogen versus host membranes). Moreover, the assembled rods may undergo an additional level of self-assembly into multirod structures, spanning pathogen membranes [22.49, 50]. Intercalation of the self-assembled rods into pathogen membranes reduces the integrity of pathogen membranes selectively, potentially rendering agents toxic to pathogens but not to hosts.

Other self-assembling, nanoscale antibiotics with morphologies more like that of micellar structures [22.42–44] have been described (N8N antimicrobial nanoemulsion [22.51]). They, as well as the self-assembling D, L peptide toroids [22.49, 50], represent critically needed, novel antibacterial agents. Resistance to traditional, microbially derived antibiotics often is tied to detoxifying functions associated with secondary metabolite synthesis; these detoxifying functions are essential for the viability of many antibiotic-producing organisms [22.52]. The genes encoding such detoxifying functions are rapidly disseminated to other microorganisms, accounting for the rapid evolution of drug-resistant organisms that has bedeviled antimicrobial chemotherapy for the last half-century. Synthetic nanoscale antibiotics, such as peptide toroids [22.49, 50] and N8N antimicrobial nanoemulsion [22.51], act by mechanisms entirely distinct from those of traditional secondary metabolite antibiotics, and presumably no native detoxifying genes exist for these supramolecular antibiotics *because* they are structures not produced by evolution. Therefore, novel nanoscale antimicrobials may not be subject to the unfortunately rapid rise in resistant organisms associated with most secondary metabolite antibiotics, though this remains to be seen. As bacterial infection continues to reemerge as a major cause of morbidity and mortality in the developed world, as a consequence of increasing antibiotic-resistant pathogens, novel, nanoscale antibiotics will become increasingly important.

Biological macromolecules undergo self-assembly at multiple levels, and like all instances of such assembly, biological self-assembly processes are driven by thermodynamic forces. Some biomolecules undergo intramolecular self-assembly (as in protein folding from linear peptide sequences, Fig. 22.9). Higher-order structures are, in turn, built by self-assembly of smaller self-assembled subunits (for instance, structures assembled by hybridization of multiple oligonucleotides, enzyme complexes, fluid mosaic membranes, ribosomes, organelles, cells, tissues, etc.). Assemblies of higher-order structures, such as oligomeric proteins, are driven by the same forces that drive intramolecular protein folding.

Naturally occurring proteins are nonrandom copolymers of 20 chemically distinct amino acid (AA) subunits. The precise order of AAs (i. e., via interactions between AA side-chains) drives the linear polypeptide

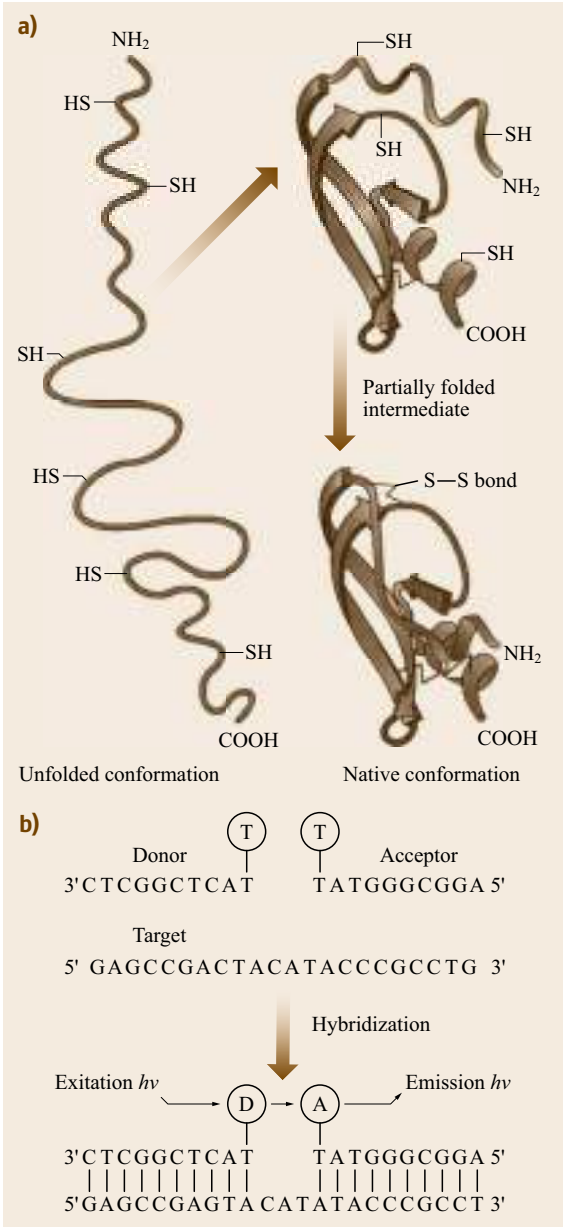


Fig. 22.9a,b Self-assembly of biological macromolecules. **(a)** Linear peptide chains (with amino- and carboxy-ends, as well as sulfhydryl groups of cysteine residues indicated) undergo a multistep folding process that involves the formation of secondary structures (α-helices, indicated by heavy helical regions, and β-sheet regions, indicated by the heavy arrows) that themselves associate into a tertiary structure. Final conformation is stabilized by the formation of intrachain disulfide linkages involving cysteine thiol groups. **(b)** A fluorescence transfer device that depends on self-assembly of biomolecules. The device is composed of donor (D) and acceptor (A) molecules brought into close proximity (within a few angstroms) by the base-pair hybridization of complementary oligonucleotides. When the structure is assembled, acceptor and donor are energetically coupled, and fluorescence transfer can occur [22.53, 54] (after [22.54]) ◄

chains to form specific secondary structures (the α-helices and β-sheet structures seen in Fig. 22.9). The secondary structures have their own preferences for association, which in turn leads to the formation of the tertiary and quaternary structures that constitute the folded protein structures. In its entirety, this process produces consistent structures that derive their biological functions from strict control of the deployment of chemical specificities (the AA side-chains) in three-dimensional space.

Biomolecules can be used to drive assembly of nanostructures, either as free molecules or when conjugated to heterologous nanomaterials (Fig. 22.9). For instance, three-dimensional nanostructures can be made by DNA hybridization [22.53–56]. Such DNA nanostructures can exhibit tightly controlled topographies but have limited integrity in terms of geometry [22.56], due to the flexibility of double-stranded (ds)DNA of 200 or more base pairs in length.

Often the domains of biomolecules responsible for assembly and recognition are small, continuous, and discrete enough that they can be abstracted from their native context as assembly modules and appended to other nanomaterials for direct formation of controlled nanoscale architectures. Antibodies and other specific biological affinity reagents can be used to assemble hybrid nanostructures (Fig. 22.9). Oligonucleotides are particularly intriguing in this application.

Unlike longer DNA segments, shorter dsDNA segments (of 50 base pairs or less) are rigid. The rigidity of short ds oligonucleotides has been cleverly exploited to build multidendrimer complexes (Fig. 22.4b) [22.30–32]. Dendrimers in these complexes are held at very precise distances from one another, defined by the length of the dsDNA segments. Nanometer-scale interdendrimer distances measured by AFM and other methods are within 10% of distances predicted from the dsDNA length [22.30, 31]. The length of the ds oligonucleotide is defined by the DNA sequence, and is easily tuned. This assembly strategy is currently being used to build multidendrimer anticancer devices, analogous to those envisioned in Fig. 22.4a, albeit with greater synthetic control and stability. The implications of the strategy goes well beyond dendrimer complexes, and might be used to impose precise distance relationships

on nanoscale components or between nanoscale components and surfaces in devices.

22.2.3 Chemoselective Conjugation

Several chemoselective bioconjugate approaches have arisen from the field of protein semisynthesis [22.20, 57, 58]. These protein synthetic chemistries allow site-specific conjugation of polypeptides to heterologous materials in bulk, as the result of conjugation between exclusive, mutually reactive electrophile–nucleophile pairs, one on the polypeptide, the other on its conjugation partner. Chemoselective conjugation strategies have been applied to the synthesis of multiple nanobiological devices [22.1, 1–8, 15, 57–60].

Proteins are profoundly dependent on their three-dimensional shapes for their activities: chemical derivatization at critical AA sites can profoundly and negatively impact protein bioactivity. Because conjugation can be directed to preselected sites via chemoselective approaches, and since the sites of conjugation in the protein can be chosen because the proteins tolerate adducts at those positions, proteins coupled to nanomaterials by such chemoselective methodologies often retain their biological activity. In contrast, protein bioactivity in conjugates often is lost or profoundly impaired when proteins are coupled to nanomaterials using promiscuous chemistries. For instance, promiscuous chemistries (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) conjugation [22.61]) used to conjugate cytokines to nanoparticles tend to inactivate human interleukin (hIL)-3 and other cytokines. The same protein–particle bioconjugates retain bioactivity if judiciously chosen chemoselective conjugation strategies are used [22.1, 59, 60].

The potential utility of chemoselective conjugation for incorporation of active biological structures into semisynthetic nanodevices cannot be overestimated. In fact, some protein engineering methods that can change protein topography, and therefore the spatial relationships between proteins and the device nanocomponents to which they are conjugated, can be used to control protein–device spatial arrangements only in circumstances where the association between the protein and device is nonrandom and oriented, as can occur via chemoselective conjugation.

Protein sequences are said to be circularly permuted (CP) when their parental amino- and carboxy-ends are ligated together, and new amino- and carboxy-ends are introduced elsewhere in the protein sequence. There are as many potential CP variants of a given protein as there are AAs in the protein sequence. In each distinct variant, the CP sequence initiates (and ends) at a different AA of the parental sequence.

The fact that many proteins tolerate circular permutation of their primary AA sequences came as a surprise to molecular biologists. Nonetheless, multiple examples of biofunctional CP proteins have been documented [22.62–71]. We recently developed a method that displays all possible CP variants of a protein on the surfaces of phage particles. Phage-presented CP proteins can be conveniently screened for function by their affinity for a target (for a known ligand of the parent protein, for instance). We call this process scanning circular permutation [22.72, 73], and it allows functional screening of every possible CP variant of a protein of interest.

One of the remarkable findings of the scanning CP work was the finding that a large fraction (as many as $\frac{1}{3}$, and possibly more) of the distinct CP variants of a test protein were bioactive [22.72]. Many of the new amino- and carboxy-ends of functional CP variants fell within known secondary structural domains of the parent protein, and yet CP variants interrupting those secondary structural domains exhibited full bioactivity. This all suggests that the linear order of protein AA and secondary structural domains as they occur in natural proteins may not be requisite for protein structure and function, and further that multiple biofunctional CP variants of many proteins may be possible [22.62–72]. Scanning CP [22.72] provides a high-throughput way to identify biofunctional CP variants and to build collections of CP versions of a single parent sequence.

When proteins are conjugated to a polymer interface by a single, specific amino acid (as by chemoselective conjugation [22.20, 57, 58]), their orientations are more or less regular (limited by the conformational freedoms of the protein and interfacial polymers themselves). Typically, chemoselective conjugation is limited solely to protein N-termini, producing consistent orientation relative to the surface. Circular permutation alters protein topology to modulate distances between protein functional domains (such as antigen-combining sites of SCFvs or single chain fragment variables) and N-termini (where conjugation uniquely occurs, and so proximal to the surface). Together, chemoselective conjugation and circular permutation modulate protein orientation on surfaces to allow tuning of distance to surfaces and protein functional domains (like antigen combining sites) [22.72, 73]. This effect is illustrated for a hypothetical antibody fragment (so-called single-chain Fv or SCFv) in Fig. 22.10. The potential utility of changing these distances is discussed in the context of a nanobiotechnological sensor system below. In this sensor, certain key nanoscale distances are the primary determinants of overall sensor function, and we will show how circular permutagenesis and other molecu-

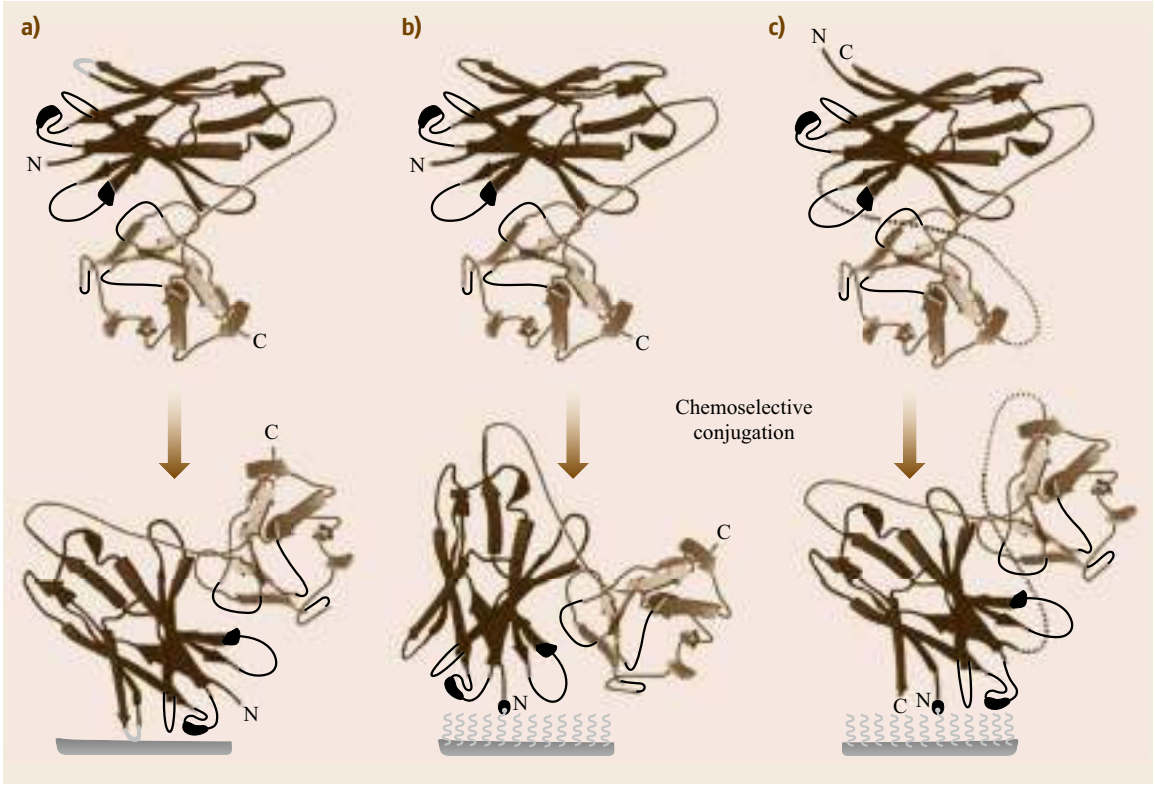


Fig. 22.10a–c Examples of protein engineering manipulations that could potentially alter the orientations of proteins (SCFvs) deployed on micromachined surfaces. Ribbon diagrams of SCFvs (VH–VL configuration) to be deposited on a SiO_2 surface. VH is shown in *dark brown*, VL in *light brown*. Complementarity-determining regions (CDRs) of VH and VL are shown in *black*. CDRs direct specific antigen recognition. Micromachined metal oxide surface is represented by a *brown bar*. A polymeric layer on the surface is represented by *wavy lines*. N- and C-ends of SCFvs are indicated. Chemoselective ligation between N-ends of SCFvs and polymer is indicated. **(a)** Affinity peptide-SCFv: surface-specific affinity peptide selected from a display library (for instance, SiO_2- [22.74, 75] or Al_2O_3-binding [22.76] peptide) is inserted into the SCFv antibody fragment (*gray line*). The affinity peptide binds to the surface, effectively orienting the SCFv and determining the proximity of the CDRs to the SiO_2 surface. **(b)** Parent SCFv: chemoselective conjugation of a modified SCFv (with an N-terminal aldehyde; [22.1, 59, 60, 72, 73] to the polymer layer. Note the position of the VH CDRs. **(c)** A circularly permuted variant of a SCFv: chemoselective conjugation of a circularly permuted [22.72, 73], but otherwise comparable, SCFv. In **(b)** and **(c)** , note that chemoselective conjugation produces a consistent orientation of SCFvs and that, relative to the parent SCFv, CP alters the proximity of the CDRs to the surface (after [22.73])

lar biology approaches can be used to address problems occurring at the nanoscale.

22.2.4 Unnatural Amino Acids to Support Chemoselective Conjugation of Biologically Produced Proteins

Early protein chemoselective conjugation methods depended on either protein or peptide chemical synthesis to introduce a nucleophile or electrophile into the protein sequence for chemoselective conjugation, or on some postsynthesis processing of a native AA to a chemically reactive group [22.20, 57, 58]. Typically, the reactive group has been at an end of a polypeptide (most typically, at the N-end). Recently, more elegant and versatile methods to introduce reactive groups to polypeptides via ribosomal synthesis have developed. These methods utilize biologically produced proteins made in expression host systems engineered to introduce unnatural AAs (UAAs) at specific sites in proteins. UAAs can have a variety of functional side-chains, including chromophores, fluors, photoreactive groups or chemically reactive groups. To the extent that reactive functionalities of UAAs can participate in chemoselective conjugations with other chemical groups, this approach can be used to prepare proteins for chemo-

selective conjugation to other surfaces or nanoscale moieties. Importantly, since the UAAs are introduced by ribosomal protein synthesis, the position of the UAA, and therefore the site of chemoselective conjugation, is not limited to the end of the protein sequence. Chemoselective conjugation can potentially be made to occur at any AA of the polypeptide sequence. Such UAA-incorporating systems have been developed in prokaryotic, lower eukaryotic [22.77, 78], and mammalian expression systems [22.79], potentially allowing proteins with UAAs to be synthesized in host systems of varying expression properties or with host-specific posttranslational modifications.

These systems for ribosomal introduction of UAAs at specific sites in biologically produced proteins have several general features in common [22.77–79]. They exploit suppression of nonsense codons (of which there are three in *E. coli*, and which, in wild-type strains, stop protein translation) by a mutant transfer ribonucleic acid (RNA) (a so-called suppressing tRNA) whose anticodon loop is homologous to the nonsense codon that will drive UAA insertion into nascent protein. Not only must this mutant tRNA recognize the nonsense codon, but it must also have the capacity to be charged with the desired UAA by a corresponding aminoacyl-tRNA synthase. Generating aminoacyl-tRNA synthases that can recognize both the suppressing tRNA and the desired UAA requires extensive genetic engineering effort. To ensure specificity of UAA incorporation, the aminoacyl-tRNA synthase cannot recognize any native AA, and cannot direct charging of any of the host's other tRNAs.

As current (2008) metabolic engineering capabilities go, UAA-incorporating systems must rank as the most sophisticated described to date. Nonetheless, their output is engineered protein with reactive groups at predetermined AA positions within them. These reactive groups can be conjugated chemoselectively to other nanoscale components, to effect changes in the spatial relationship of particular protein functional domains and the components to which proteins are conjugated. A collection of proteins differing from one another only by the position of the UAA that participates in chemoselective conjugation could have a utility to nanotechnologists that closely parallels that of a collection of circularly permuted proteins. As Fig. 22.10 shows, circular permuted proteins chemoselectively conjugated to surfaces can be used to tune distance between underlying surfaces and protein functional domains. Incorporation of UAAs and chemoselective conjugation therefore might also be used for analogous purposes in nanobiological devices.

22.3 Sensing Devices

The need for *smart* therapy is a key theme of therapeutic nanotechnology, and of pharmacology as a whole. We discussed above the emerging class of theranostic agents that include sensing or diagnostic functions.

Drugs with narrow therapeutic windows should ideally be delivered only to their desired site of action and be pharmacologically active only when their activity is needed. These strategies can limit undesired secondary effects of therapy, some of which can be debilitating or life threatening. One possible approach to this issue is the incorporation of sensing capability (specifically, the capacity to recognize appropriate contexts for therapeutic activity) into nanotherapeutic devices. Sensing capability may allow self-regulation of a therapeutic device, reporting to an external clinician/device operator (as in imaging applications, see below), or both.

Biosensors are typically considered to be multifunctional, multicomponent devices [22.80]. Usually a biosensor system is composed of a signal transducer, a sensor interface, a biological detection (bioaffinity) agent, and an associated assay methodology, with each system component governed by its inherent operational considerations. Signal transducers are moieties that are sensitive to a physicochemical change in their environment and that undergo some detectable change in chemistry, structure or state as the result of analyte (the thing to be sensed) recognition. Analytes for nanotherapeutic application could be biomolecules, such as proteins, small molecules (organic or inorganic), and ions (salts or hydrogen ions), or physical conditions (such as redox state or temperature). Interfaces are the sensor components that interact directly with the analyte. For sensor use in nanotherapeutic devices, immobilized or otherwise captured biological molecules (proteins, nucleic acids) often constitute the active parts of biosensor interfaces. Whatever the chemical nature of the interface, it determines the selectivity, sensitivity, and stability of the sensing system and also is a dominant determinant of sensor operational limits. Assay methodology determines the need (or lack thereof) for analyte tracers, the number of analytical reagents, and the complexity and rapidity of the sensing process.

22.3.1 Planar FET Protein Sensors

To illustrate the potential of biotechnology to solve nanoscale problems, we will focus on protein-sensing immunoFETs (a field-effect transistor sensor which

binds analyte via an antibody molecule or fragment thereof on its sensing surface). Field-effect transistors (FETs) consist of a current source and current drain separated from each other by a semiconductor channel through which current flows. A *gate* to which electrical bias is applied is positioned above the channel to modulate channel conductance. Depending on gate bias, current flow through the FET channel is increased or decreased (Fig. 22.11) [22.81–85]. Electrical properties of FET sensing channels can be modulated by any proximal electric field/charge acting as a gate, allowing configuration of sensors [22.81–91], including, in principal, protein sensors. Receptor molecules (antibody fragments, peptides, aptamers, etc.) that recognize protein analyte molecules are deployed on the sensing channel, and analytes bind to the interface via the receptors. Charges of protein analytes induce a dipole between the surface and the underlying depletion region of the semiconductor, eliciting a gating effect which affects current through the FET. Differently charged analytes interact differentially with charge carriers in the FET, producing analyte-charge-specific responses. FET sensors can be constructed from nanowires (so-called 1-D FETs). One-dimensional structures *may* have much enhanced sensitivity due to their high surface-area-to-volume ratios [22.82–85]. An example of a 1-D FET was shown in Fig. 22.5, and 1-D FETs much like those pictured are widely accepted as nanotechnological devices, though meso- to macroscale planar FET sensors often are not. The reason for considering 1-D FETs as *real* nanotechnology is that 1-D devices contain a nanoscale component (the semiconducting nanowire). They also additionally contain nanoscale receptor molecules. Thus, although physical nanotechnologists often chafe (very unreasonably) at this argument, planar FETs are nanodevices for the same reason as 1-D nanowire FETs are. Planar FETs, though they themselves may be meso- to macroscale, depend on obligately nanoscale components (biomolecular receptors for analytes) for their function. Here we focus on *planar* FETs as a platform to demonstrate how biotechnology can solve a problem occurring at the nanoscale.

The potential utility of FET protein sensing in physiological environments is obvious, but has only recently been realized [22.90, 91]. FET sensor sensitivity to analyte charge is limited by multiple factors [22.81–91], though the key problem cited is typically regarded to be shielding by buffer ions in the high-salt (150 mM Na^+) physiological environment (Fig. 22.12) [22.81–84, 86]. This led to an assessment that planar immunoFETs, owing to the size of antibodies (10–12 nm) and the maximal distance over which analyte charges were thought to be detectable by the FET (sometimes called the Debye length, < 2 nm), were fundamentally infeasible. The

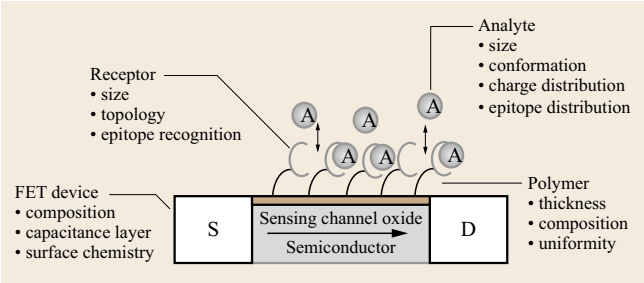


Fig. 22.11 A protein-sensing FET and some of the parameters influencing its function. Current source and drain (S, D) are shown, as is the semiconductor (*gray box*). Capacitance layer/sensing channel oxide is shown in *brown*. Current flowing from the source to the drain in the semiconductor (I_{SD}) is indicated by the *black arrow*, and increases or decreases in response to charges brought to the sensing channel surface as a result of analyte binding. Receptors (SCFvs, streptavidin, etc.) are shown as *gray C-shaped lines* and bind analyte (*gray balls, labeled A*) specifically but reversibly and are conjugated to the sensing channel by an interfacial polymer. Parameters that might be chosen judiciously to optimize the sensor are shown. Note that interfacial optimization occurs primarily on the nanoscale, with specific changes to molecular sizes and orientations

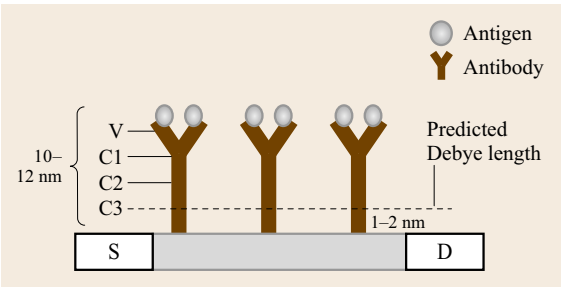


Fig. 22.12 The biochemical basis of the classical analysis interpreted as indicating the infeasibility of planar immunoFETs operating in physiological buffer. Artist's conception of antibodies on FET sensing channel. Antibody domains (constant 1, 2, 3 as C1, C2, C3 and variable as V) are indicated, as is the predicted Debye length in physiological buffer (150 mM salt, about 1–2 nm). FET source and drain (S and D) are shown. Shown is the classical (but wrong, see text) depiction of antibodies adsorbed onto a FET sensing channel solely by their C3 domains. This flawed representation is central to the classical objection to the feasibility of immunoFETs (after [22.81])

classical analysis of immunoFETs reasonably considers ion shielding, but is so fatally flawed in its consideration of receptor protein structures on FET surfaces that it is irrelevant. To biochemists, the fallacies of the classical model are obvious, and biochemists encountering the classical model immediately dismiss it as meaningless,

though much of the sensing community has clung tenaciously to the moribund model. We have discussed the fatal flaws of the classical analysis elsewhere [22.90, 91] and will repeat the discussion only briefly:

1. Use of intact antibody as receptors as in Fig. 22.12 is unnecessary: *much* smaller (10×), antigen-binding antibody fragments are available.
2. Antibodies do not adsorb exclusively at C3 domain (as shown in Fig. 22.12): antibodies bind surfaces in a distribution of orientations.
3. Antibodies contain a flexible region at the crotch of the *Y* called the *hinge* that allows antigen-combining sites vast freedom in their relative positions. The implicit assumption illustrated in Fig. 22.12 is that antibodies are rigid bodies, which is wrong.
4. The effect of antigen (analyte) three-dimensional (3-D) shape and charge distribution on charge proximity to the sensing channel is not considered in classical analysis.
5. The diversity available in native and phage-presented antibody repertoires was not exploited. The impact on sensor properties of analyte receptors (antibodies or fragments thereof) of differing valency, conformations, antigen (analyte) affinity or analyte epitope recognition properties were not considered.
6. Improvement of antibody receptor properties by protein engineering is not considered in the classical analysis.
7. Data directly shows that planar FET protein sensors can detect a biologically important analyte at concentrations which occur in real biological systems under physiological salt conditions [22.90, 91].

Suffice it to say that, if planar immunoFETs are not feasible, the reasoning of the classical assessment has little to do with why they might be infeasible. However, we know from our work that planar FET protein sensors operating at physiological salt concentrations are feasible [22.90, 91], though we show that the distance between the analyte charge and the sensing channels is a key parameter for sensor sensitivity. We will focus on biotechnology approaches to engineer the nanometer distance between bound analyte charges and FET sensing surfaces as an existence proof of the utility of biotechnology to nanotechnology.

22.3.2 Biotechnology Approaches to the *Fundamental Limitations* of Planar ImmunoFETs

First, in the classical model [22.81–84, 86], chemoselective conjugation of receptors (antibodies) to sensing interfaces is not considered. Chemoselective conjugation would have the desirable effect of orienting the antibody layer consistently, from one antibody molecule to the next. Even though the classical representation (as mimicked in Fig. 22.12) shows antibodies adsorbed onto sensing surfaces in an oriented fashion, it is well known to immunologists that antibodies adsorb to surfaces in a stochastic distribution of orientations, binding to the surfaces via a random distribution of antibody surface regions. That is to say that the antibody layer as contemplated in the classical assessment is misrepresented, and the lack of orientation of the antibodies at real interfaces would interfere with many of the available biotechnology strategies to control the distance between the analyte charge and the sensing surface. However, the entire classical assessment of immunoFETs [22.81–84, 86] is based on the incorrect representation of the extent of order in the protein orientation shown in Fig. 22.12. This critical flaw alone invalidates the classical assessment, but there is much more wrong with the assessment than just this point.

Secondly, use of intact antibodies (IgG) as immunoFET receptors is unnecessary: specific epitope recognition function can be isolated on fragments several-fold smaller than intact antibodies [22.21, 92–99]. Single-chain Fv antibody fragments (single-chain fragment variables, SCFvs) are less than half the size of intact antibodies. Single-domain (camelid) antibody fragments, called variable heavy–heavy (VHH), are smaller still: about 10% of the mass of intact antibodies [22.97–99]. There are multiple methods to convert existing antibody genes (isolated from mammalian B-cells) to genes that will direct biologic production of either SCFvs or VHHs recognizing the same antigen (analyte), as well as methods to isolate de novo SCFvs or VHHs that are directed to antigens of interest that are well known to those skilled in the art [22.97–99]. The simple expedient of using an antibody fragment as an immunoFET receptor reduces the distance between the sensing surface and the bound analyte charge to nearly the calculated distance over which buffer counterion shielding is expected to occur (Fig. 22.12 [22.81–84, 86]). Again, failure to consider this fact is in itself sufficient to invalidate the classical assessment. Antibody fragments are convenient substrates for chemoselective conjugation, and we exploit the protein orientation chemoselective approaches in our SCFv layers to engineer analyte–semiconductor distances.

Thirdly, no consideration is given in the classical model to engineering the antibody molecules or fragments thereof to optimize their function as immunoFET receptors. This was a major oversight, even at the time when the classical model was promulgated in the early 1990s. The tools available for antibody engineering are

vast and sophisticated. We have discussed a few of them above (receptor circular permutagenesis [22.72, 73] and introduction of UAAs [22.77–79], both useful in the context of chemoselective conjugation of the antibody or fragments thereof to the sensing interface). As Fig. 22.10 illustrates, the effect of circular permutagenesis of a SCFv is to change the orientation of its antigen-combining sites to the site of chemoselective conjugation. Effectively, the orientation (relative to the immunoFET sensing channel) and therefore the distance charges of analytes bound to antibody fragments is different when analytes are bound to different CP antibody variants. Since charges of protein analytes are not very mobile, the critical distance between analyte charges and sensing surfaces is different with different circularly permuted variants of a single antibody fragment. In the context of the immunoFET, it should be possible to identify those circularly permuted antibody fragments that produce the greatest sensor sensitivity, presumably because they minimize the distance between bound analyte charges and FET sensing surfaces. Similarly, the optimal AA position for a UAA in an antibody fragment could be determined for an antibody fragment used as a FET receptor from the standpoint of immunoFET sensitivity. The approach could tune the distance between underlying surfaces and analyte charges, much as circular permutagenesis can.

There are still many other biotechnology approaches that might be used to optimize receptors for immunoFETs that have been ignored to date. We and multiple others have identified peptides that can specifically bind microfabricated surfaces [22.74–76]. Such peptides might be inserted into SCFv or VHH sequences and used to adhere the SCFv or VHH to the micromachined surface, as we suggest [22.73, 74]. Presumably, variable positions of peptide insertion could lead to differential orientations of antibody fragments on the immunoFET surface, with the potential benefits in terms of FET sensitivity already described for circular permutagenesis or UAA insertion described above. This has the potentially great advantage of adhering immunoFET receptors directly to the sensing surface, but our experience with direct adsorption of receptor proteins on micromachined surfaces have proven direct adsorption to be unsatisfactory in our hands (see below).

The foregoing is a very incomplete list of potential biotechnology approaches to addressing a limitation correctly identified (buffer counterion shielding of analyte charges) but incorrectly analyzed [22.81–84, 86]. Many more approaches are possible, although we will not attempt to enumerate them here. The point is that planar immunoFETs have a nanoscale limitation that can be addressed by biotechnology. Hence, in this application, biotechnology is a bona fide nanotechnology.

22.3.3 Nanotribology of Protein-Sensing Interfaces on Micromachined Surfaces

In FET protein sensors, the critical importance of the sensing interface is clear, implying the need to characterize the interface, and nanotribology and surface characterization have led to the discovery of additional interfacial design parameters and properties that must be considered in immunoFET design. Parameters of interest were thickness of interfaces (related to, but distinct from, the critical distance between bound analyte charges and semiconductor surfaces), surface roughness (smoothness), and interfacial robustness (resistance to wear). Resistance to wear is a critical issue if sensors are to be subjected to abrasion, as by tissue in an in vivo environment.

These studies were performed on silicon surfaces with a thermally grown oxide layer (silica surfaces [22.100–103]) and on a sputter-coated aluminum layer with a native oxide (aluminum surfaces [22.103]). Silicon surfaces simulate the sensing surfaces of a metal–oxide–semiconductor field-effect transistor (MOSFET), whereas aluminum simulates the sensing surface of an AlGaN (aluminum–gallium nitride) heterojunction FET. Native Al_2O_3 oxide which develops on sputter-coated aluminum surfaces is also the most prevalent oxide on the surfaces of AlGaN/GaN heterojunction FETs. All studies were performed with the model receptor protein streptavidin, and not with an antibody fragment.

Lee et al. and *Bhushan* et al. studied the step-by-step morphological changes resulting on silicon surfaces during deposition of a model protein (streptavidin) [22.100, 101]. These studies revealed, among other things, that streptavidin adsorbed to silicon interacted with the surface very weakly. Most of the adsorbed streptavidin could be removed from the surface by rinsing with an aqueous buffer [22.101]. Streptavidin adherence by adsorption to micropatterned asperities on the silicon surface and to the edges of silicon fragments was modestly stronger than to the flat silicon surface, although adhesion was still weak [22.100, 101]. Adhering streptavidin to the silicon surface via a biotinylated silane polymer (3-aminopropyltriethoxysilane (APTES), Fig. 22.13) significantly enhanced resistance to wear [22.101–103].

Ellipsometry of APTES-coated silicon surfaces showed the thickness of the polymer layer to be several-fold the expected monolayer thickness [22.103]. The unexpected thickness of the film likely represents multilayering by APTES, a phenomenon that had been previously reported [22.104–106]. Despite the fact that the sensor with a multilayered APTES layer was suf-

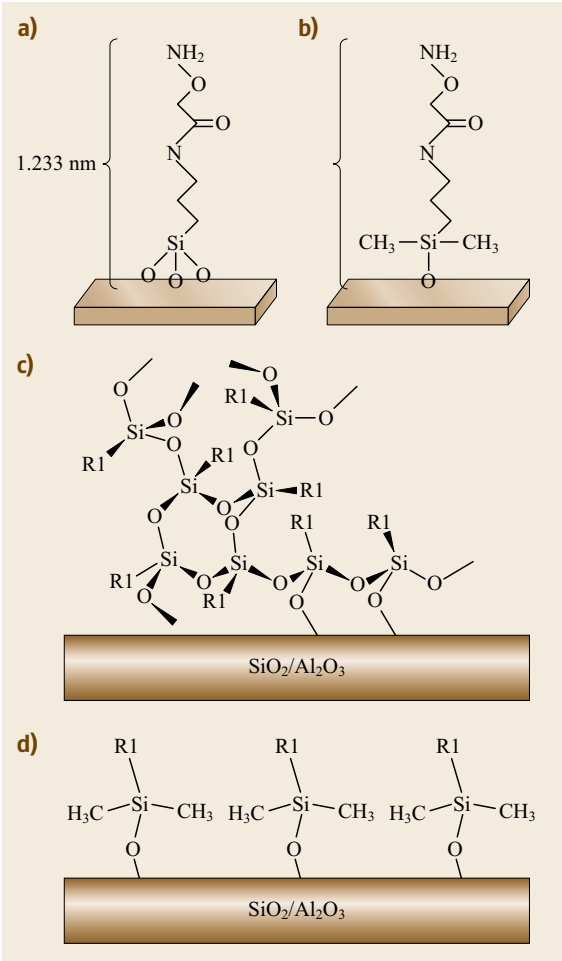


Fig. 22.13a–d Tri- and monoalkoxy silane constituents of the sensor interface on micromachined substrates. In **(a)** and **(b)**, oxyamine-terminated polymers for chemoselective conjugation are shown. **(a)** shows the trialkoxy interfacial polymer and **(b)** its monoalkoxy analogue. As determined by summing bond lengths, the length of polymer in nm is shown. Both ellipsometry (Fig. 22.14) and AFM analysis show that interfaces made with the monoalkoxy version of polymers are thinner and less rough, presumably due to reduced intramolecular silane polymerization [22.103–106]. **(c)** shows a hypothetical film made from trialkoxy polymer of **(a)**, and **(d)** shows a hypothetical film film made with the monoalkoxy polymer of **(b)**. R1 is ($CH_2CH_2NHCOCH_2ONH_2$) (after [22.103])

ficiently sensitive to detect analyte [22.90, 91], thinner interfaces would theoretically increase sensor sensitivity to analyte. Multilayering is related to the trivalency of APTES siloxane residues and is influenced by the deposition method [22.104–106]. Consequently, APTES polymer films were also constructed using a more optimized deposition protocol [22.104]. The more optimal method produced thinner APTES films, as expected (Fig. 22.14). However, peak-to-valley values for APTES films deposited by either protocol were substantially greater than the summed bond lengths of APTES [22.100–103]. The monosiloxane reagent, aminopropyldimethylethoxysilane (APDMES, Fig. 22.13), which has a much diminished capacity to produce multilayers, was also used to build polymer films on silicon. Strikingly, APDMES produced the thinnest films, with thicknesses comparable to the summed bond lengths of the APDMES polymer (Fig. 22.14).

The thicker silane films (based on APTES) were also substantially rougher than the APDMES film. APTES films applied by our original procedure were roughest, followed by APTES films deposited by the more optimal procedure, and APDMES films were the smoothest and most regular. This roughness ranking was conserved after subsequent steps in interface construction (biotinylation of and streptavidin binding to the biotinylated silane polymer, Fig. 22.14 [22.102, 103]). The results observed on silicon surfaces were replicated on aluminum surfaces, with APDMES providing a smoother, thinner interfacial polymer layer than did APTES [22.103]. Furthermore, the rank order of roughness following biotinylation was also conserved for APTES films deposited by our two protocols and APDMES films on aluminum [22.103].

Bhushan et al. [22.103] studied friction and wear of streptavidin deposited by physical adsorption and via a biotinylated APTES polymer layer on the silicon surface. The coefficient of friction for streptavidin-adsorbed silicon surfaces is less than for uncoated silicon [22.103], because the streptavidin acts as a lubricant film. Consistent with this interpretation, coefficients of friction were dependent on the concentration of streptavidin (that is, on the surface density of adsorbed streptavidin molecules, which increases when higher concentrations of streptavidin were used in coating the surfaces), and decreased at higher streptavidin concentrations. At higher densities of adsorbed streptavidin, the surface is more uniform and smoother, and the silicon substrate is more fully covered with protein than at lower concentration. This implies that the streptavidin forms a nearly continuous lubricant film at higher concentration. When streptavidin is linked to the surface through a biotinylated silane polymer, the coefficient of friction increases. This follows from the increased compliance of the underlying silane polymer. When normal load is applied to the surface, the surface is compressed, owing to the increased mechanical compliance afforded by interfacial biotinylated polymer length, flexibility or layering, all resulting in a larger contact area between the AFM tip and in-

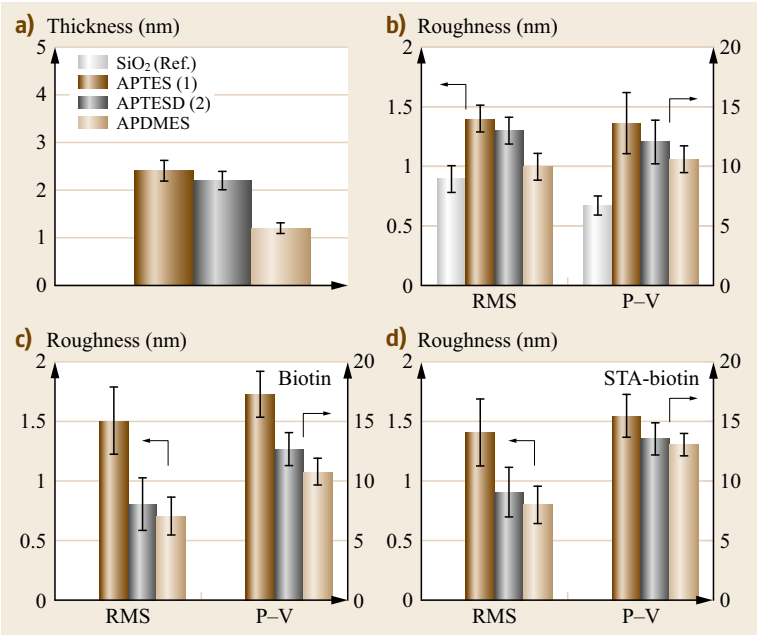


Fig. 22.14a–d Representation of nanotribological properties of films on silicon/SiO_2 surfaces. APTES (1) films were deposited by the method used in our published HFET (heterojunction field effect transistor) protein sensors [22.91, 92], and APTES (2) and APDMES films were deposited by the improved method described in [22.104]. **(a)** Silane film thickness (ellipsometry), **(b)** surface roughness (root mean square (RMS) and peak-to-valley (P–V) values) of silane films, **(c)** surface roughness (RMS and P–V values) of the same three silane films following biotinylation, and **(d)** surface roughness (RMS and P–V values) of the same three silane films following streptavidin binding to the biotin (after [22.103])

terfacial molecules. Since the size of streptavidin is much larger than that of the biotinylated silane polymer, the tightly packed streptavidin molecules cause very little lateral deflection of the tip. In streptavidin surfaces deposited by linkage to a polymer bound to the silicon surface, high contact area and low lateral deflection cause the friction force to increase relative to the same applied normal load on streptavidin surfaces deposited by adsorption [22.102, 103]. Together, these effects produce a *cushioning effect* that runs counter to the lubricating effect observed when protein (streptavidin) is directly adsorbed to the surfaces, increasing the coefficients of friction for streptavidin deposited by a polymer interface. The observations of streptavidin's lubricating properties and polymer–protein interfacial cushioning properties have been replicated on aluminum surfaces [22.103]. We think these are likely general properties of many protein interfaces on micromachined surfaces. We expect that many proteins (such as SCFvs adsorbed onto various metal oxide surfaces) will produce analogous lubricating properties, and that cushioning-mediated friction increases will occur for many polymer–protein interfaces, regardless of the composition of the underlying smooth, micromachined surfaces.

Interestingly, streptavidin bound to silicon surfaces via biotinylated silane polymers had differential resistance to wear, with APTES interfaces (which exhibit the most multilayering) being less robust than interfaces based on APDMES (with limited/no multilayering). The most highly multilayered polymer interfaces had, by definition, larger extents of intermonomer cross-linking, and were less resistant to wear by cantilevers [22.103]. This was interpreted by analogy to graphene layers in graphite, in that bonds between monomers of the polymer layer were more numerous than bonds between the polymer layer and the silicon substrate. Hence, when the polymer layer tore under stress from the cantilever, a larger section of the cross-linked polymer sheet tore free. It was expected that, if the polymer layer were a true, un-cross-linked, self-assembled monolayer, individual polymer monomers would have no way to transfer force applied on them to their neighboring monomers, and so would be individually plucked off the interface by the tip.

These studies also [22.102, 103] occasioned further insights into the mechanism of wear of polymer–protein interfaces on micromachined surfaces. The polymer–protein interfaces (biotinylated silane polymer–streptavidin) on silicon were sufficiently delicate that their wear could not be investigated by contact-mode AFM at any load tested. Contact mode always immediately stripped the surface, even at the lowest loads achievable. Consequently, wear experiments were performed at varying loads in tapping mode, wherein the free amplitude (and hence the load) of the cantilever was systematically varied (Fig. 22.15). As one might expect, as load increased, wear increased. The finding implies that wear of APDMES–streptavidin interfaces on silicon is not a concerted process, but proceeds sequentially more deeply as the applied force increases. That said, all tested interfaces were very delicate, and

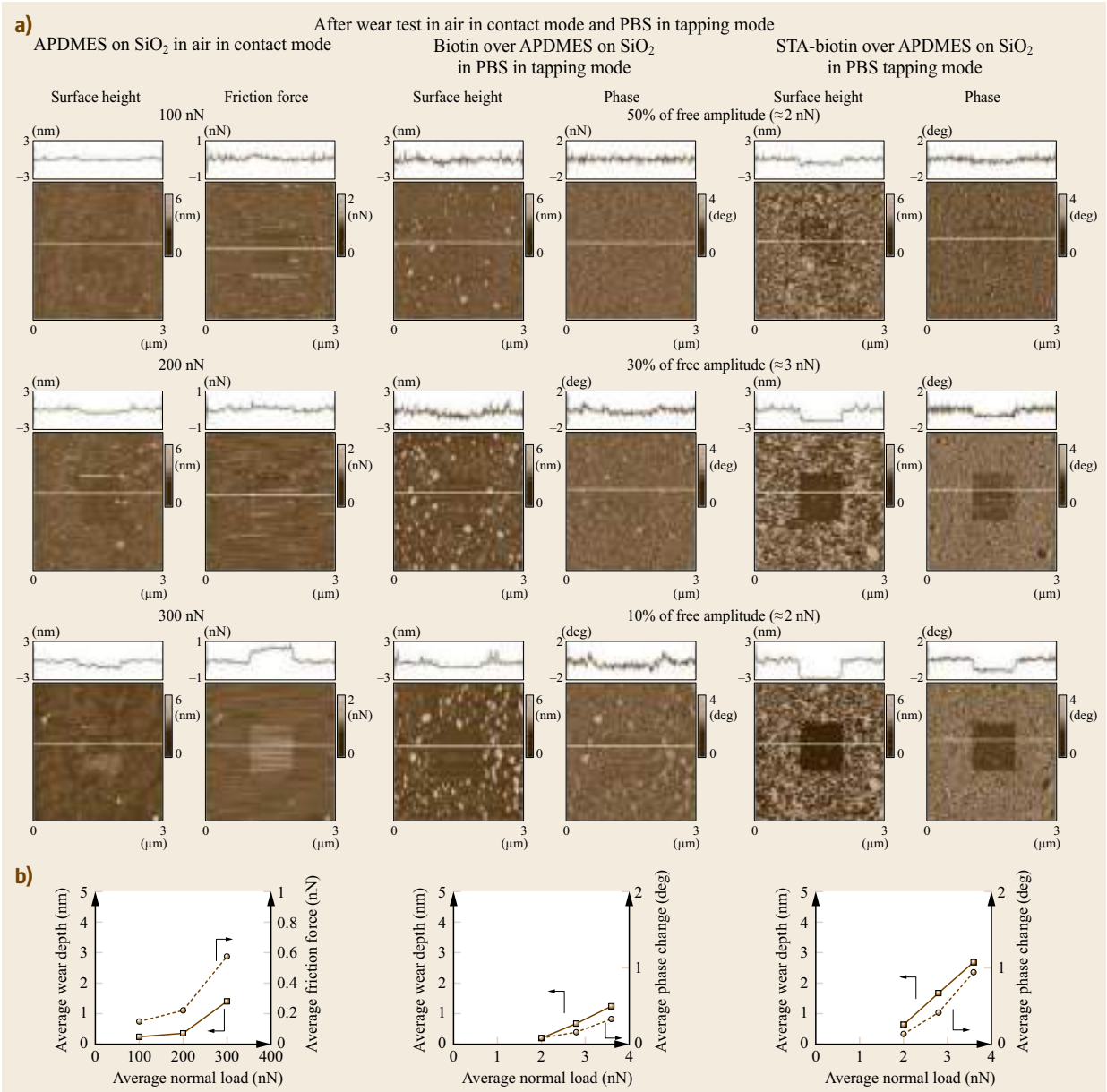


Fig. 22.15a,b Differential wear of interfacial polymer–protein films by AFM operated in tapping mode. **(a)** AFM surface height and friction force/phase angle images and cross-sectional profiles obtained after wear testing in air in contact mode/phosphate buffered saline (PBS) in tapping mode at a range of normal loads and free amplitudes, and **(b)** plot of average wear depth and average friction force/phase angle as a function of average normal load, for APDMES on silicon/SiO_2, biotin over APDMES on silicon/SiO_2, and on STA–biotin on APDMES and silicon/SiO_2 (after [22.103])

required only nanonewton forces to remove them from the surfaces.

It can reasonably be expected that unprotected immunoFET sensors in vivo will experience at least nanonewton abrasive forces. Such forces are likely to be sufficient to strip receptor layers from the immunoFET, and thereby cause sensor failure. This implies that the sensor polymer–receptor interfaces must either be hardened somehow to resist encountered abrasive insult, or that the sensor interfaces must be sheltered from direct, abrasive insult by tissue. This problem is not satisfactorily solved, but is currently under inves-

tigation. However, the point remains that the primacy of the importance of the immunoFET interface drove surface-science and nanotribological study of the interface. As a consequence, we documented properties new to nanoscale tribology of protein interfaces (lubricating and cushioning effects). Nanotribological characterization of models of immunoFET surfaces also revealed another nanoscale parameter that must be addressed before the immunoFETs can be deployed in in vivo clinical or research applications: the robustness to mechanical insult (or lack thereof) of FET receptor layers.

22.4 Concluding Remarks: Barriers to Practice

The effort to produce nanotherapeutic devices is highly interdisciplinary, with a sweep of knowledge that is difficult for any one investigator to master fully. The effort required constitutes a major barrier to entry to the field. Previous versions of this chapter ended with a discussion of the complexity of biology (being much greater than encountered in traditional engineering disciplines), the fashion in which biologists prioritize information, and cultural differences between engineers and biological scientists [22.107]. The chapter discussed some of the ways in which biologists and engineers hold and disseminate information (approaches so vastly different as to interfere with design of nanobiological devices). It also discussed the (then largely unmet) need for significant biological education for biomedical engineers, at least those who will be involved in producing devices that interact with living systems in vivo, or contain components derived from living systems. The consensus regarding the need for biologic training for biomedical engineers appears to be shifting in favor of more biology, in no small part as a result of the increasing prevalence of biologic components in biomedical nano- and microdevices and attempts to transfer those technologies to the clinic. However, there are still conundrums related to how biomedical and other engineers see themselves and see biology that result in lost opportunities.

22.4.1 You Do Not Know What You Do Not Know: The Consequences of Certainty

As discussed above, planar immunoFETs and similar bioFETs operating at physiological salt concentrations were regarded as infeasible from 1991 up until recently [22.81–84, 86]. However, there are now examples of such planar FET protein sensors that can detect protein analytes in physiological buffers at analyte concentrations known to occur in vivo [22.90, 91]. While this might seem to be a breakthrough in sensing technology, this development was delayed for years longer than it needed to be by the intransigent adherence of biosensor specialists to a flawed assessment of feasibility [22.81–84, 86].

The irony was that, with the right FET platform, planar FET sensing of proteins in physiological buffers has always been possible. The sensing community could have discovered this fact 20 years ago, and potentially could have moved to making functional immunoFET sensors for clinical use at that time, but for stubborn adherence to a flawed model of planar immunoFET interface structure. Until as little as a year ago, funding agencies and journals responded to applications or papers discussing planar FETs for protein sensing at physiological salt concentrations as if the documents reflected the proposer's or author's naivete. The direct response was often that the issue of planar FET sensors had been dealt with since 1991, and did not warrant further consideration. This response might have made sense had the classical assessment been theoretically sound, or if it were unsound in some way that required research to discover. However, the assessment was unsound from its first enunciation, largely because it did not take into account knowledge about protein and antibody structure that was available long before 1991. In the original proposal of the assessment, this may have been a result of the (then) poorly appreciated interdisciplinary nature of biosensor technology, or naivete regarding antibody structures and properties.

The classical assessment was promulgated by electrical engineers who had little knowledge of antibody structure, so they made mistakes in aspects of the assessment dependent on understanding antibodies. In its general form, this problem is a pitfall to which all interdisciplinary scientists are vulnerable. By definition, one cannot know what one does not know, and to some extent ignorance must be forgiven, as it is the default human condition. However, it should not have taken a working planar bioFET to overturn the classical assessment. The classical assessment was *dead on arrival*. Immunologists knew on cursory inspection that the classical assessment was nearly without relevance to real protein sensing. Unfortunately, engineers did not. In many cases, they failed to reexamine the assumptions of the classical assessment, in favor of reiterating its conclusion. If anything was done improperly, it was in failing to consider reasoned discussion of the flaws

in the assessment (as above). Doing so may have cost the sensor field 20 years.

This is mentioned without the intent to embarrass persons who proposed or adhered to the classical assessment, but rather to point out that dogmatic certainty has consequences. Error is an ongoing hazard in any interdisciplinary field. As biomedical nanotechnologists, we mention this story with trepidation for our own pronouncements.

22.4.2 Are Proteins and Molecules Legitimately Part of Nanotechnology?

There is a school of thought that holds that nanotechnology deals solely with quantum and other phenomena that manifest only when matter is subdivided into nanoscale particles or structures. This reasoning would exclude supramolecular chemistry depending on organic molecules and devices built from protein or nucleic acid functional components from nanotechnology. If this reasoning is accepted as valid, nanotechnology so defined may have little intersection with human health.

As the reader might assume, such reasoning does not appeal to us. We hope we have presented a compelling case that molecules are essential components of nanotherapeutics. Organic or biologic macromolecules (proteins, nucleic acids, etc.) are necessary for many interactions between living things and therapeutic devices. In part, this is because these molecules are dimensionally and chemically appropriate to interact with the biocomponent of the patient (i. e., the argument of *speaking* to the patient's body in a language it understands). Also, synthetic nanostructures with the precise chemical and topological complexity of biological macromolecules cannot yet be produced or tuned with the same efficacy as biotechnology produces and tunes the properties of proteins (i. e., the argument that biomolecules are the only show in town). Moreover, with the examples of the FET protein sensor and the IKVAV PA amphiphile, we have shown that biotechnology and organic chemistry (respectively) can be used to address problems occurring at the nanoscale. In this sense, they fall reasonably under the rubric of nanotechnology.

The molecular nanocomponents of the immunoFET and IKVAV PA are tuned to suit (at least partially) using systematic knowledge of nanocomponent properties. It is our opinion that engineering is any activity that uses knowledge of component properties to control the behavior of materials and devices containing those components to achieve a desired end. In the case of the biological molecules with which we work, knowledge of properties is frequently fragmentary, limiting our ability to design individual protein molecules de novo, and driving us to use molecular genetic screening techniques to identify variant proteins with desired properties. Nonetheless, we see ourselves as engineers because we use knowledge of our materials (proteins) to its limits to optimize functional devices (immunoFETs, magnetic resonance imaging (MRI) contrast agents). In the future, protein structure is likely to become more amenable to quantitative modeling and design, and the inclusion of protein engineering as a genuine engineering discipline will become less controversial.

At any rate, fortunately for us, and for biomedical nanotechnologists, the restrictive view of the definition of nanotechnology and engineering is less widely held than in the past. These fields will likely become increasingly welcoming to biologists as time passes.

Acknowledgments. The authors acknowledge the contribution to the chapter by Philip D. Barnes. This work was supported in part by grants from the Program for Homeland Security of the Ohio State University (program no. 14 525) and award GRT00011123 (project no. 60 015 533) to S.C.L. from the National Science Foundation.

References

22.1 S.C. Lee, R. Parthasarathy, T. Duffin, K. Botwin, T. Beck, G. Lange, J. Zobel, D. Jansson, D. Kunneman, E. Rowold, C.F. Voliva: Antibodies to PAMAM dendrimers: Reagents for immune detection assembly and patterning of dendrimers. In: *Dendrimers and Other Dendritic Polymers*, ed. by D. Tomalia, J. Frechet (Wiley, London 2001) pp. 559–566

22.2 S.C. Lee: Biotechnology for nanotechnology, Trends Biotechnol. **16**, 239–240 (1998)

22.3 S.C. Lee: Engineering the protein components of nanobiological devices. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 67–74

22.4 S.C. Lee: How a molecular biologist can wind up organizing nanotechnology meetings. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998)

22.5 S.C. Lee: The nanobiological strategy for construction of nanodevices. In: *Biological Molecules in*

Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 3–14

22.6 S.C. Lee: A biological nanodevice for drug delivery. In: *Natl. Sci. Technol. Counc. IWGN Workshop Rep.: Nanotechnol. Res. Dir. Int. Technol. Res. Inst., World Technol. Div.* (Kluwer, Baltimore 1999) pp. 91–92

22.7 S.C. Lee, R. Parthasarathy, K. Botwin: Proteinpolymer conjugates: Synthesis of simple nanobiotechnological devices, Polym. Prepr. **40**, 449–450 (1999)

22.8 L. Jelinski: Biologically related aspects of nanoparticles, nanostructured materials and nanodevices. In: *Nanostructure Science and Technology*, ed. by R.W. Siegel, E. Hu, M.C. Roco (Kluwer, Dordrecht 1999) pp. 113–130

22.9 B.R. Smith, J. Heverhagen, M. Knopp, P. Schmalbrock, J. Shapiro, M. Shiomi, N. Moldovan, M. Ferrari, S.C. Lee: Magnetic resonance imaging of atherosclerosis in vivo using biochemically targeted ultrasmall superparamagnetic iron oxide particles (SPIONs), Biomed. Microdevices **9**, 719–728 (2007)

22.10 A.J. Nijdam, T.R. Nicholson III, J. Shapiro, B.R. Smith, J.T. Heverhagen, P. Schmalbrock, M.V. Knopp, A. Kebbel, D. Wang, S.C. Lee: Biochemically targeted nanoparticulate contrast agents for magnetic resonance imaging diagnosis of cardiovascular disease, Curr. Nanosci. **5**, 88–102 (2009)

22.11 J.R. Baker Jr.: Therapeutic nanodevices. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 173–183

22.12 R. Duncan: Drug targeting: Where are we now and where are we heading?, J. Drug Target. **5**, 1–4 (1997)

22.13 R. Duncan, S. Gac-Breton, R. Keane, Y.N. Sat, R. Satchi, F. Searle: Polymer-drug conjugates, PDEPT and PELT: Basic principles for design and transfer from the laboratory to clinic, J. Cont. Release **74**, 135–146 (2001)

22.14 D.S. Goldin, C.A. Dahl, K.L. Olsen, L.H. Ostrach, R.D. Klausner: Biomedicine. The NASA-NCI collaboration on biomolecular sensors, Science **292**, 443–444 (2001)

22.15 S.C. Lee: Dendrimers in nanobiological devices. In: *Dendrimers and Other Dendritic Polymers*, ed. by D. Tomalia, J. Frechet (Wiley, London 2001) pp. 548–557

22.16 J.R. Baker Jr., A. Quintana, L. Piehler, M. Banazak-Holl, D. Tomalia, E. Racka: The synthesis and testing of anti-cancer therapeutic nanodevices, Biomed. Microdevices **3**, 61–69 (2001)

22.17 K.D. Bhalerao, E. Eteshola, M. Keener, S.C. Lee: Nanodevice design through the functional abstraction of biological macromolecules, Appl. Phys. Lett. **87**, 143902–143904 (2005)

22.18 S.C. Lee, K. Bhalerao, M. Ferrari: Object oriented design tools for supramolecular devices and biomedical nanotechnology, Ann. New York Acad. Sci. **1013**, 110–123 (2004)

22.19 J.M. Harris, N.E. Martin, M. Modi: Pegylation: A novel process for modifying pharmacokinetics, Clin. Pharmacokin. **40**, 539–551 (2001)

22.20 S.B.H. Kent: Building proteins through chemistry: Total chemical synthesis of protein molecules by chemical ligation of unprotected protein segments. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 75–92

22.21 C.A. Janeway, P. Travers, M. Walport, J.D. Capra: *Immunobiology* (Elsevier, London 1999)

22.22 S.C. Lee, M.S. Leusch, V.A. Luckow, P. Olins: Method of producing recombinant viruses in bacteria, US Patent 5348886 (1993)

22.23 M.S. Leusch, S.C. Lee, P.O. Olins: A novel hostvector system for direct selection of recombinant baculoviruses (bacmids) in E. coli, Gene **160**, 191–194 (1995)

22.24 V.A. Luckow, S.C. Lee, G.F. Barry, P.O. Olins: Efficient generation of infectious recombinant baculoviruses by site-specific, transposon-mediated insertion of foreign DNA into a baculovirus genome propagated in E. coli, J. Virol. **67**, 4566–4579 (1993)

22.25 T. Gardner, C.R. Cantor, J.J. Collins: Construction of a genetic toggle switch in E. coli, Nature **403**, 339–342 (2000)

22.26 J. Hasty, F. Isaacs, M. Dolnik, D. McMillen, J.J. Collins: Designer gene networks: Towards fundamental cellular control, Chaos **11**, 107–220 (2001)

22.27 S. Uppuluri, D.R. Swanson, L.T. Piehler, J. Li, G. Hagnauer, D.A. Tomalia: Core shell tecto(dendrimers). I. Synthesis and characterization of saturated shell models, Adv. Mater. **12**, 796–800 (2000)

22.28 A.K. Patri, I.J. Majoros, J.R. Baker Jr.: Dendritic polymer macromolecular carriers for drug delivery, Curr. Opin. Chem. Biol. **6**, 466–471 (2002)

22.29 A. Quintana, E. Raczka, L. Piehler, I. Lee, A. Myc, I. Majoros, A.K. Patri, T. Thomas, J. Mule, J.R. Baker Jr.: Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor, Pharma. Res. **19**, 1310–1316 (2002)

22.30 Y. Choi, T. Thomas, A. Kotlyar, M. Islam, J. Baker Jr.: Synthesis and functional evaluation of DNA-assembled polyamidoamine dendrimer clusters for cancer cell-specific targeting, Chem. Biol. **12**, 35–43 (2005)

22.31 Y. Choi, A. Mecke, B.G. Orr, M.M. Banaszak Holl, J.R. Baker Jr.: DNA-directed synthesis of generation 7 and 5 PAMAM dendrimer nanoclusters, Nano Lett. **4**, 391–397 (2004)

22.32 D.G. Mullen, A.M. Desai, J.N. Waddell, X.-M. Cheng, C.V. Kelly, D.Q. McNerny, I.J. Majoros, J.R. Baker Jr., L.M. Sander, B.G. Orr, M.M. Banaszak Holl: The implications of stochastic synthesis for

the conjugation of functional groups to nanoparticles, Bioconjug. Chem. **19**, 1748–1752 (2008)
22.33 T.R. Groves, D. Pickard, B. Rafferty, N. Crosland, D. Adam, G. Schubert: Maskless electron beam lithography: Propects, progress and challenges, Microelectron. Eng. **61**, 285–293 (2002)
22.34 M. Guthold, R. Superfine, R. Taylor: The rules are changing: Force measurements on single molecules and how they relate to bulk reaction kinetics and energies, Biomed. Microdevices **3**, 9–18 (2001)
22.35 L.M. Demers, D.S. Ginger, S.-J. Park, Z. Li, S.-W. Chung, C.A. Mirkin: Direct patterning of modified oligonucleotides on metals and insulatos by dip-pen nanolithography, Science **296**, 1836–1838 (2002)
22.36 K.-B. Lee, S.-J. Park, C.A. Mirkin, J.C. Smith, M. Mrksich: Protein nanoarrays generated by dip-pen nanolithography, Science **295**, 1702–1705 (2002)
22.37 M. Ferrari, J. Liu: The engineered course of treatment, Mech. Eng. **123**, 44–47 (2001)
22.38 K.E. Drexler: *Engines of Creation: The Coming Era of Nanotechnology* (Anchor Books, New York 1986)
22.39 J. Cumings, A. Zetti: Low-friction nanoscale linear bearing realized frommultiwall carbon nanotubes, Science **289**, 602–604 (2000)
22.40 D.J. Hornbaker, S.-J. Kahng, S. Mirsa, B.W. Smith, A.T. Johnson, E.J. Mele, D.E. Luzzi, A. Yazdoni: Mapping the one-dimensional electronic states of nanotube peapod structures, Science **295**, 828–831 (2002)
22.41 C. Dekker: Carbon nanotubes as molecular quantum wires, Phys. Today **28**, 22–28 (1999)
22.42 M.-C. Jones, J.-C. Leroux: Polymeric micelles – a new generation of colloidal drug carriers, Eur. J. Pharma. Biopharma. **48**, 101–111 (1999)
22.43 I. Uchegbu: Parenteral drug delivery: 1, Pharma. J. **263**, 309–318 (1999)
22.44 I. Uchegbu: Parenteral drug delivery: 2, Pharma. J. **263**, 355–359 (1999)
22.45 J.D. Hartgerink, E.R. Zubarev, S.I. Stupp: Supramolecular one-dimensional objects, Curr. Opin. Solid State Mater. Sci. **5**, 355–361 (2001)
22.46 L.C. Palmer, Y.S. Velichko, M. Olvera De La Cruz, S.I. Stupp: Supramolecular self-assembly codes for functional structures, Philos. Trans. R. Soc. A **365**, 1417–1433 (2007)
22.47 G.A. Silva, C. Catherine, K.L. Niece, E. Beniash, D.A. Harrington, J.A. Kessler, S.I. Stupp: Selective differentiation of neural progenitor cells by high-epitope density nanofibers, Science **303**, 1352–1355 (2004)
22.48 V.M. Tysseling-Mattiace, V. Sahni, K.L. Niece, D. Birch, C. Czeisler, M. Fehlings, S.I. Stupp, J.A. Kessler: Self-assembling nanofibers inhibit glial scar formation and promote axon elongation after spinal cord injury, J. Neurosci. **28**, 3814–3823 (2008)
22.49 S. Fernandez-Lopez, H.-S. Kim, E.C. Choi, M. Delgado, J.R. Granja, A. Khasanov, K. Kraehenbuehl, G. Long, D.A. Weinberger, K.M. Wilcoxen, M. Ghardiri: Antibacterial agents based on the cyclic D,L-alpha-peptide architecture, Nature **412**, 452–455 (2001)
22.50 A. Saghatelian, Y. Yokobayashi, K. Soltani, M.R. Ghadiri: A chiroselective peptide replicator, Nature **409**, 777–778 (2001)
22.51 T. Hamouda, A. Myc, B. Donovan, A.Y. Shih, J.D. Reuter, J.R. Baker Jr.: A novel surfactant nanoemulsion with a unique non-irritant topical antimicrobial activity against bacteria, enveloped viruses and fungi, Microbiol. Res. **156**, 1–7 (2001)
22.52 J. Davies: Aminoglycoside-aminocyclitol antibiotics and their modifying enzymes. In: *Antibiotics in Laboratory Medicine*, ed. by V. Lorian (Williams and Wilkins, Baltimore 1984) pp. 474–489
22.53 M.J. Heller: Utilization of synthetic DNA for molecular electronic and photonic-based device applications. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 59–66
22.54 Z. Ma, S. Taylor: Nucleic acid triggered catalytic drug release, Proc. Natl. Acad. Sci. USA **97**, 11159–11163 (2000)
22.55 R.C. Merkle: Biotechnology as a route to nanotechnology, Trends Biotechnol. **17**, 271–274 (1999)
22.56 N.C. Seeman, J. Chen, Z. Zhang, B. Lu, H. Qiu, T.-J. Fu, Y. Wang, X. Li, J. Qi, F. Liu, L.A. Wenzler, S. Du, J.E. Mueller, H. Wang, C. Mao, W. Sun, Z. Shen, M.H. Wong, R. Sha: A bottom-up approach to nanotechnology using DNA. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 45–58
22.57 G. Lemieux, C. Bertozzi: Chemoselective ligation reactions with proteins, oligosaccharides and cells, Trends Biotechnol. **16**, 506–512 (1998)
22.58 R. Offord, K. Rose: Multicomponent synthetic constructs. In: *Biological Molecules in Nanotechnology: The Convergence of Biotechnology, Polymer Chemistry and Materials Science*, ed. by S.C. Lee, L. Savage (IBC, Southborough 1998) pp. 93–105
22.59 S.C. Lee, R. Parthasarathy, K. Botwin, D. Kunneman, E. Rowold, G. Lange, J. Zobel, T. Beck, T. Miller, W. Hood, J. Monahan, R. Jansson, J.P. McKearn, C.F. Voliva: Biochemical and immunological properties of cytokines conjugated to dendritic polymers, Biomed. Microdevices Biomems Biomed. Nanotechnol. **6**, 191–201 (2004)
22.60 S.C. Lee, R. Parthasarathy, T. Duffin, K. Botwin, T. Beck, G. Lange, J. Zobel, D. Kunneman, E. Rowold, C.F. Voliva: Recognition properties of antibodies to PAMAM dendrimers and their use in immune detection of dendrimers, Biomed. Microdevices **3**, 51–57 (2001)
22.61 G.T. Hermanson: *Bioconjugate Chemistry* (Academic, San Diego 1996)

22.62 S. Topell, R. Glockshuber: Circular permutation of the green fluorescent protein, Meth. Mol. Biol. **183**, 31–48 (2002)

22.63 A. Rojas, S. Garcia-Vallve, J. Palau, A. Romeu: Circular permutations in proteins, Biologia **54**, 255–277 (1999)

22.64 T.U. Schwartz, R. Walczak, G. Blobel: Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor beta subunit, Protein Sci. **13**, 2814–2818 (2004)

22.65 U. Heinemann, M. Hahn: Circular permutation of polypeptide chains: Implications for protein folding and stability, Prog. Biophys. Mol. Biol. **64**, 121–143 (1996)

22.66 A. Buchwalder, H. Szadkoski, K. Kirschner: A fully active variant of dihydrofolate reductase with a circularly permuted sequence, Biochemistry **31**, 1621–1630 (1992)

22.67 L.S. Mullins, K. Wesseling, J.M. Kuo, J.B. Garrett, F.M. Raushel: Transposition of protein sequences: Circular permutation of ribonuclease T1, J. Am. Chem. Soc. **116**, 5529–5533 (1994)

22.68 M. Hahn, K. Piotukh, R. Borriss, U. Heinemann: Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis, Proc. Natl. Acad. Sci. USA **91**, 10417–10421 (1994)

22.69 Y.R. Yang, H.K. Schachznan: Aspartate transcarbamoylase containing circularly permuted catalytic polypeptide chains, Proc. Natl. Acad. Sci. USA **90**, 11980–11984 (1993)

22.70 X. Lin, G. Koelsch, J.A. Loy, J. Tang: Rearranging the domains of pepsinogen, Protein Sci. **4**, 159–166 (1995)

22.71 M.L. Vignais, C. Corbier, G. Mulliert, C. Branlant, G. Branlant: Circular permutation within the coenzyme binding domain of the tetrameric glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*, Protein Sci. **4**, 994–1000 (1995)

22.72 E. Eteshola, C.D. Van Valkenburgh, S. Merlin, E. Rowold, J. Adams, R. Ibdah, L.E. Pegg, A. Donelly, E. Rowold, J. Klover, S.C. Lee: Screening of a library of circularly permuted proteins on phage to manipulate protein topography, J. Nanoeng. Nanosyst. **219**, 45–55 (2006)

22.73 E. Eteshola, M.T. Keener, M.A. Elias, J. Shapiro, L.J. Brillson, B. Bhushan, S.C. Lee: Engineering functional protein interfaces for immunologically modified field effect transistors (ImmunoFETs) by molecular genetic means, J. R. Soc. Interface **5**, 123–127 (2008)

22.74 E. Eteshola, L. Brillson, S.C. Lee: Selection and characteristics of peptides that bind thermally grown silicon dioxide films, Biomol. Eng. **22**, 202–204 (2005)

22.75 R.R. Naik, L.L. Brott, S.J. Clarson, M.O. Stone: Silica-precipitating peptides isolated from a combinatorial phage display peptide library, J. Nanosci. Nanotechnol. **2**, 95–100 (2002)

22.76 E.M. Krauland, B.R. Peelle, K.D. Wittrup, A.M. Belcher: Peptide tags for enhanced cellular and protein adhesion to single-crystalline sapphire, Biotechnol. Bioeng. **97**, 1009–1020 (2007)

22.77 L. Wang, J. Xie, P.G. Schultz: Expanding the genetic code, Annu. Rev. Biophys. Biomol. Struct. **35**, 225–249 (2006)

22.78 J. Xie, P.G. Schultz: A chemical toolkit for proteins – an expanded genetic code, Nat. Rev. Mol. Cell Biol. **7**, 775–782 (2006)

22.79 N. Hino, A. Hayashi, K. Sakamoto, S. Yokoyama: Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code, Nat. Protoc. **1**, 2957–2962 (2006)

22.80 K. Rogers: Principles of affinity-based biosensors, Mol. Biotechnol. **14**, 109–129 (2000)

22.81 M.J. Schoning, A. Poghossian: Recent advances in biologically sensitive field-effect transistors (BioFETS), Analyst **127**, 1137–1151 (2002)

22.82 P. Bergveld, J. Hendrikes, W. Olthuis: Theory and application of the material work function for chemical sensors based on the field effect principle, Meas. Sci. Technol. **9**, 1801–1808 (1998)

22.83 W. Olthius, P. Bergveld, J. Kruise: The exploitation of ISFETs to determine acid-base behavior of proteins, Electrochim. Acta **43**, 3483–3488 (1997)

22.84 R.B. Schasfoort, R.P. Kooyman, P. Bergveld, J. Greve: A new approach to ImmunoFET operation, Biosens. Bioelectron. **5**, 103–124 (1990)

22.85 Y. Cui, Q. Wei, H. Park, C.M. Lieber: Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species, Science **293**, 1289 (2001)

22.86 P. Bergveld: A critical evaluation of direct electrical protein detection methods, Biosens. Bioelectron. **6**, 55–72 (1991)

22.87 J.I. Hahm, C.M. Lieber: Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors, Nano Lett. **4**, 51–54 (2004)

22.88 A. Star, J.C.P. Gabriel, K. Bradley, G. Gruner: Electronic detection of specific protein binding using nanotube FET devices, Nano Lett. **3**, 459–463 (2003)

22.89 G. Zheng, F. Patolsky, Y. Cui, W.U. Wang, C.M. Lieber: Multiplexed electrical detection of cancer markers with nanowire sensor arrays, Nat. Biotechnol. **23**, 1294–1301 (2005)

22.90 J. Shapiro, S. Gupta, E. Eteshola, M. Elias, X. Wen, W. Lu, L.J. Brillson, S.C. Lee: Challenges in optimization of nanobiotechnological devices illustrated by partial optimization of a protein biosensor. In: *Proc. 2nd Int. Congr. Nanobiotechnol. Nanomed.*, *NanoBio 2007* (Int. Association Nanotechnology, San Jose 2007)

22.91 S. Gupta, M. Elias, X. Wen, J. Shapiro, L. Brillson, W. Lu, S.C. Lee: Detection of clinically relevant levels of biological analyte under physiologic buffer using planar field effect transistors, Biosens. Bioelectron. **24**, 505–511 (2008)

22.92 K. Decanniere, A. Desmyter, M. Lauwereys, M.A. Ghahroudi, S. Muyldermans, L. Wyns: A single-domain antibody fragment in complex with rnase a: Non-canonical loop structures and nanomolar affinity using two CDR loops, Struct. Fold Des. **7**, 361–370 (1999)

22.93 K. Decanniere, T.R. Transue, A. Desmyter, D. Maes, S. Muyldermans, L. Wyns: Degenerate interfaces in antigen-antibody complexes, J. Mol. Biol. **313**, 473–478 (2001)

22.94 A. Desmyter, K. Decanniere, S. Muyldermans, L. Wyns: Antigen specificity and high affinity binding provided by one single loop of a camel single-domain antibody, J. Biol. Chem. **276**, 26285–26290 (2001)

22.95 A. Muruganandam, J. Tanha, S. Narang, D. Stanimirovic: Selection of phage-displayed llama single-domain antibodies that transmigrate across human blood-brain barrier endothelium, FASEB J. **16**, 240–242 (2002)

22.96 S. Muyldermans: Single domain camel antibodies: Current status, Mol. Biotechnol. **74**, 277–302 (2001)

22.97 S. Muyldermans, M. Lauwereys: Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies, J. Mol. Recogn. **12**, 131–140 (1999)

22.98 L. Riechmann, S. Muyldermans: Single domain antibodies: comparison of camel VH and camelised human VH domains, J. Immunol. Meth. **231**, 25–38 (1999)

22.99 M.S. Hayden, L.K. Gilliland, J.A. Ledbetter: Antibody engineering, Curr. Opin. Immunol. **9**, 201–212 (1997)

22.100 B. Bhushan, D.R. Tokachichu, M.T. Keener, S.C. Lee: Morphology and adhesion of biomolecules on silicon based surfaces, Acta Biomater. **1**, 327–341 (2005)

22.101 S.C. Lee, M.T. Keener, D.R. Tokachichu, B. Bhushan, P.D. Barnes, B.J. Cipriany, M. Gao, L.J. Brillson: Protein binding on thermally grown silicon dioxide, J. Vac. Sci. Technol. B **23**, 1856–1865 (2005)

22.102 B. Bhushan, D.R. Tokachichu, M.T. Keener, S.C. Lee: Nanoscale adhesion, friction and wear studies on silicon based surfaces, Acta Biomater. **2**, 39–49 (2006)

22.103 B. Bhushan, K.J. Kwak, S. Gupta, S.C. Lee: Nanoscale adhesion, friction and wear studies of biomolecules on polymer-coated silica and alumina based surfaces, J. R. Soc. Interface **6**, 719–733 (2009)

22.104 Y. Han, D. Mayer, A. Offenhausser, S. Ingebrandt: Surface activation of thin silicon oxides by wet cleaning and silanization, Thin Solid Films **510**, 175–180 (2006)

22.105 K. Kallury, P.M. MacDonald, M. Thompson: Effect of surface water and base catalysis on the silanization of silica by (aminopropyl)alkoxysilanes studied by x-ray photoelectron spectroscopy and 13C cross-polarization/magic angle spinning nuclear magnetic resonance, Langmuir **10**, 492–499 (1994)

22.106 J.H. Moon, J.W. Shin, S.Y. Kim, J.W. Park: Formation of uniform aminosilane thin layers: An imine formation to measure relative surface density of the amine group, Langmuir **12**, 4621–4624 (1996)

22.107 S.C. Lee, M. Reugsegger, P.D. Barnes, B.R. Smith, M. Ferrari: Therapeutic nanodevices. In: *Springer Handbook of Nanotechnology*, ed. by B. Bhushan (Springer, Berlin, Heidelberg 2007) pp. 461–504

Part E Nanometrology

Part E Nanometrology

23 Scanning Probe Microscopy – Principle of Operation, Instrumentation and Probes
Bharat Bhushan, Columbus, USA
Othmar Marti, Ulm, Germany

24 Low-Temperature Scanning Probe Microscopy
Mehmet Z. Baykara, Ankara, Turkey
Markus Morgenstern, Aachen, Germany
Alexander Schwarz, Hamburg, Germany
Udo D. Schwarz, New Haven, USA

25 Biomedical Sensing with the Atomic Force Microscope
Constanze Lamprecht, Linz, Austria
Jürgen Strasser, Linz, Austria
Melanie Koehler, Louvain-la-Neuve, Belgium
Sandra Posch, Linz, Austria
Yoojin Oh, Linz, Austria
Rong Zhu, Linz, Austria
Lilia A. Chtcheglova, Linz, Austria
Andreas Ebner, Linz, Austria
Peter Hinterdorfer, Linz, Austria

26 Superresolution Microscopy
Tom D. Milster, Tucson, USA

23. Scanning Probe Microscopy – Principle of Operation, Instrumentation and Probes

Bharat Bhushan, Othmar Marti

Since the introduction of the Scanning Tunneling Microscope (STM) in 1981 and Atomic Force Microscope (AFM) in 1985, many variations of probe-based microscopies, referred to as Scanning Probe Microscopes (SPM), have been developed. While the pure imaging capabilities of SPM techniques are dominated by the application of these methods at their early development stages, the physics of probe-sample interactions and the quantitative analyses of tribological, electronic, magnetic, biological, and chemical surfaces have now become of increasing interest. In this chapter, we introduce various STM and AFM designs, various operating modes, various probes (tips), and AFM instrumentation and analyses.

23.1 **Scanning Tunneling Microscope** 727
23.1.1 The STM Design of Binnig et al. 727
23.1.2 Commercial STMs 728
23.1.3 STM Probe Construction 730
23.2 **Atomic Force Microscope** 731
23.2.1 The AFM Design of Binnig et al. 733
23.2.2 Commercial AFM 733
23.2.3 AFM Probe Construction 738
23.2.4 Friction Measurement Methods 742
23.2.5 Normal Force and Friction Force Calibrations of Cantilever Beams 745
23.3 **AFM Instrumentation and Analyses** 747
23.3.1 The Mechanics of Cantilevers 747
23.3.2 Instrumentation and Analyses of Detection Systems for Cantilever Deflections 749
23.3.3 Combinations for 3-D-Force Measurements 757
23.3.4 Scanning and Control Systems 758
23.4 **Conclusion** 763
References 763

The scanning tunneling microscope (STM) developed by Dr. Gerd Binnig and his colleagues in 1981 at the IBM Zurich Research Laboratory, Rueschlikon, Switzerland, was the first instrument capable of directly obtaining three-dimensional (3-D) images of solid surfaces with atomic resolution [23.1]. G. Binnig and H. Rohrer received a Nobel Prize in Physics in 1986 for their discovery. STMs can only be used to study surfaces that are electrically conductive to some degree. On the basis of their STM design, in 1985 *Binnig* et al. developed an atomic force microscope (AFM) to measure ultrasmall forces (less than 1 μN) present between the AFM tip surface and the sample surface [23.2] (also see [23.3]). AFMs can be used for measurement of all engineering surfaces that may be either electrically conductive or insulating. AFM has become a popular surface profiler for topographic and normal force measurements on the micro- to nanoscale.

AFMs have been modified in order to measure both normal and lateral forces, and are called lateral force microscopes (LFMs) or friction force microscopes (FFMs) [23.4–11]. FFMs have been further modified to measure lateral forces in two orthogonal directions [23.12–16]. A number of researchers have continued to improve the AFM/FFM designs and used them to measure adhesion and friction of solid and liquid surfaces on micro- and nanoscales [23.9–11, 17–33]. AFMs have been used for scratching, wear, and measurements of elastic/plastic mechanical properties (such as indentation hardness and modulus of elasticity) [23.9–11, 21, 23, 26–32, 34–45].

AFMs have been used for manipulation of individual atoms of Xenon [23.46], molecules [23.47], silicon surfaces [23.48], and polymer surfaces [23.49]. STMs have been used for formation of nanofeatures by local-

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_23

ized heating or by inducing chemical reactions under the STM tip [23.50–52] and nanomachining [23.53]. AFMs have been used for nanofabrication [23.10, 36, 54–56] and nanomachining [23.57].

STMs and AFMs are used at extreme magnifications ranging from 10^3 to $10^9\times$ in x-, y-, and z-directions for imaging macro to atomic dimensions with high resolution information and for spectroscopy. These instruments can be used in any environment including ambient air [23.2, 58], various gases [23.17], liquid [23.59–61], vacuum [23.1, 62], low temperatures (lower than about 100 K) [23.63–67], and high temperatures [23.68, 69]. Imaging in liquid allows the study of live biological samples and also eliminates water capillary forces present in ambient air located at the tip-sample interface. Low-temperature (liquid helium temperatures) imaging is useful for the study of biological and organic materials and the study of low-temperature phenomena such as superconductivity or charge-density waves. Low-temperature operation is also advantageous for high-sensitivity force mapping due to the reduction in thermal vibration. STMs and AFMs have also been used to image liquids such as liquid crystals and lubricant molecules on graphite surfaces [23.70–73].

The pure imaging capabilities of variations of probe-based microscope techniques, referred to as scanning probe microscopy (SPM), have dominated the application of these methods at their early development stages. However, the physics and chemistry of probe-sample interactions and the quantitative analyses of tribological, electronic, magnetic, biological, and chemical surfaces have now become of increasing interest. Nanoscale science and technology are strongly driven by SPMs, which allow investigation and manipulation of surfaces down to the atomic scale. With growing understanding of the underlying interaction mechanisms, SPMs have found application in many fields outside basic research fields. In addition, various derivatives of all these methods have been developed for special applications, some of them targeting far beyond microscopy.

A family of instruments based on STMs and AFMs, SPMs have been developed for various applications of scientific and industrial interest. These include STM, AFM, FFM (or LFM), scanning electrostatic force microscopy (SEFM) [23.74, 75], scanning force acoustic microscopy (SFAM) (or atomic force acoustic microscopy (AFAM)) [23.21, 22, 42, 45, 76–78], scanning magnetic microscopy (SMM) (or magnetic force microscopy (MFM)) [23.79–83], scanning near-field optical microscopy (SNOM) [23.84–87], scanning thermal microscopy (SThM) [23.88–90], scanning electrochemical microscopy (SEcM) [23.91], scanning Kelvin Probe microscopy (SKPM) [23.92–96], scanning chemical potential microscopy (SCPM) [23.89], scanning ion conductance microscopy (SICM) [23.97, 98], and scanning capacitance microscopy (SCM) [23.92, 99–101]. The subfamily of instruments that measure forces (e.g., AFM, FFM, SEFM, SFAM, and SMM) are also referred to as scanning force microscopes (SFM). Although these instruments offer atomic resolution and are ideal for basic research, they are most often used for cutting-edge industrial applications that do not require atomic resolution.

Commercial production of SPMs started with STM in 1987 and AFM in 1989 by Digital Instruments, Inc. (now Bruker Instruments, Inc.). For comparisons of SPMs with other microscopies including Transmission Electron Microscopy (TEM) and Scanning Electron Microscopy (SEM), see Table. 23.1 (Bruker Instruments, Inc.). Numbers of these instruments are equally divided into US, Japan, and Europe with industry/university and government lab splits of about 50/50, 70/30, and 30/70, respectively. It is clear that research and industrial applications of SPMs are rapidly expanding. For various examples, see [23.11, 102–110].

This chapter presents an overview of STM and AFM with various operating modes and various probes (tips) used in these instruments. Thereafter, details on AFM instrumentation and analyses are given.

Table 23.1 Comparison of various conventional microscopes with SPMs in 2015

	Optical	SEM/TEM	Confocal	SPM
Magnification	10^3	10^7	10^4	10^9
Instrument price (US $)	≈ $20k	≈ $500k	≈ $60k	≈ $150k
Technology age	200 yrs	50 yrs	30 yrs	30 yrs
Applications	Ubiquitous	Science & technology	New & unfolding	Science & technology
Approximate market 2015	$2B	$1.5B	$400M	$600M

23.1 Scanning Tunneling Microscope

The principle of electron tunneling was proposed by *Giaever* [23.111]. He proposed that if an electrical potential difference is applied to two metals separated by a thin insulating film, a current will flow across the barrier because of the ability of electrons to penetrate it. To be able to measure this tunneling current, the two metals must be spaced no more than 10 nm apart. *Binnig* et al. [23.1] introduced vacuum tunneling combined with lateral scanning. The vacuum provides the ideal barrier for tunneling. The lateral scanning allows surface imaging with exquisite resolution – laterally less than 1 nm and vertically less than 0.1 nm – sufficient to define the position of single atoms. The very high vertical resolution is obtained because the tunneling current varies exponentially with the distance between the metal tip and the scanned surface acting as two electrodes. Typically, tunneling current decreases by a factor of 2 as the separation is increased by 0.2 nm. Very high lateral resolution depends upon sharp tips. *Binnig* et al. [23.1] overcame two key obstacles in regard to damping external vibrations and moving the tunneling probe into close proximity to the sample. Their instrument is called the scanning tunneling microscope (STM). Today's STMs can be used in the ambient environment for atomic-scale imaging of surfaces. Excellent reviews on this subject are presented by *Hansma* and *Tersoff* [23.112], *Sarid* and *Elings* [23.113], *Durig* et al. [23.114], *Frommer* [23.115], *Güntherodt* and *Wiesendanger* [23.116], *Wiesendanger* and *Güntherodt* [23.117], *Bonnell* [23.118], *Marti* and *Amrein* [23.119], *Stroscio* and *Kaiser* [23.120], and *Güntherodt* et al. [23.121].

The principle of the STM is straightforward. A sharp metal tip (one electrode of the tunnel junction) is brought close enough (0.3−1 nm) to the surface to be investigated (second electrode) so that, at a convenient operating voltage (10 mV–1 V), the tunneling current varies from 0.2 to 10 nA, which is measurable. The tip is scanned over a surface at a distance of 0.3−1 nm, while the tunneling current between it and the surface is sensed. The STM can be operated in either the constant current mode or the constant height mode, Fig. 23.1. The left-hand column of Fig. 23.1 shows the basic constant current mode of operation. A feedback network changes the height of the tip z to keep the current constant. The displacement of the tip given by the voltage applied to the piezoelectric drives then yields a topographic map of the surface. Alternatively, in the constant height mode, a metal tip can be scanned across a surface at nearly constant height and constant voltage while the current is monitored, as shown in the right-hand column of Fig. 23.1. In this case, the feedback network responds only rapidly enough to keep the average current constant. The constant current mode is used generally for atomic-scale images. This mode is not practical for rough surfaces. A three-dimensional picture $z(x, y)$ of a surface consists of multiple scans $z(x)$ displayed laterally from each other in the y-direction. It should be noted that if different atomic species are present in a sample, the different atomic species may produce different tunneling currents for a given bias voltage. Thus, the height data may not be a direct representation of the topography of the surface of the sample.

23.1.1 The STM Design of Binnig et al.

Figure 23.2 shows a schematic of one of *Binnig* and *Rohrer*'s designs for operation in ultrahigh vacuum [23.1, 122]. The metal tip was fixed to rectangular piezodrives P_x, P_y, and P_z made out of commercial piezoceramic material for scanning. The sample is mounted on either a superconducting magnetic levitation or a two-stage spring system to achieve a stability of the gap width of about 0.02 nm. The tunnel current J_T is a sensitive function of the gap width d that is

$$J_T \propto V_T \exp(-A\phi^{1/2}d) ,$$

where V_T is the bias voltage, ϕ is the average barrier height (work function), and the constant $A = 1.025\,\mathrm{eV}^{-\frac{1}{2}}\,\mathrm{Å}^{-1}$. With a work function of a few eV, J_T

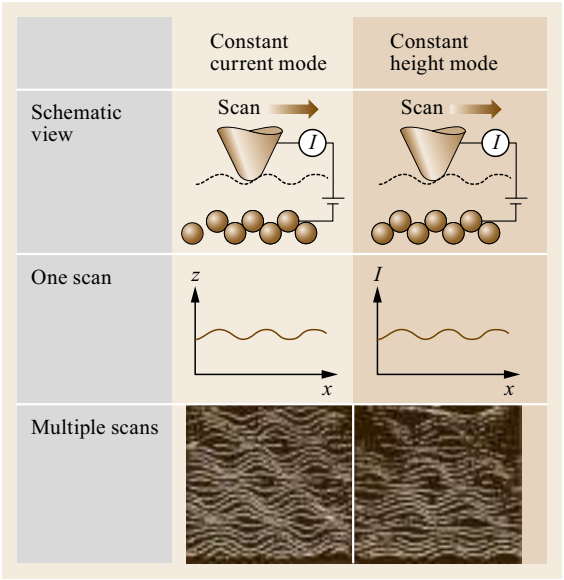


Fig. 23.1 STM can be operated in either the constant current or constant height mode. The *images* are of graphite in air [23.112]

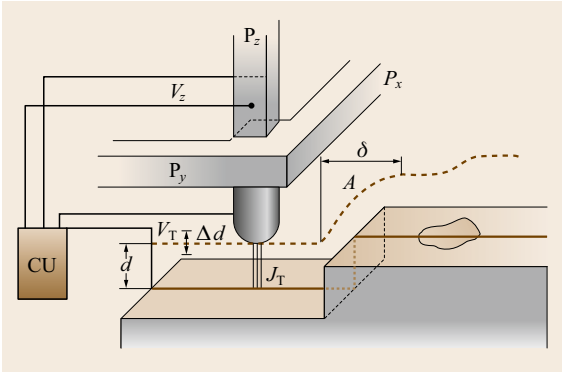


Fig. 23.2 Principle of operation of the STM made by *Binnig* and *Rohrer* [23.122]

changes by an order of magnitude for every angstrom change of d. If the current is kept constant to within, for example, 2%, then the gap d remains constant to within 1 pm. For operation in the constant current mode, the control unit (CU) applies a voltage V_z to the piezo P_z such that J_T remains constant when scanning the tip with P_y and P_x over the surface. At a constant work function ϕ, $V_z(V_x, V_y)$ yields the roughness of the surface $z(x, y)$ directly, as illustrated at a surface step at A. Smearing the step, δ (lateral resolution) is on the order of $(R)^{1/2}$, where R is the radius of curvature of the tip. Thus, a lateral resolution of about 2 nm requires tip radii of the order of 10 nm. A 1 mm diameter solid rod ground at one end at roughly 90° yields overall tip radii of only a few hundred nm, but with closest protrusion of rather sharp microtips on the relatively dull end yields a lateral resolution of about 2 nm. In situ sharpening of the tips by gently touching the surface brings the resolution down to the 1 nm range by applying high fields (on the order of 10^8 V/cm). For example, resolutions considerably below 1 nm were achieved by applying a high field for half an hour. Most experiments were carried out with tungsten wires, either ground or etched to a radius typically in the range of 0.1–10 µm. In some cases, in situ processing of the tips was carried out for further reduction of tip radii.

23.1.2 Commercial STMs

There are a number of commercial STMs available on the market. In 1987, Digital Instruments, Inc. (now Bruker Instruments) located in Santa Barbara, CA, introduced the first commercial STM, the Nanoscope I. In a later model, the Nanoscope IV STM for operation in ambient air, the sample is held in position while a piezoelectric crystal in the form of a cylindrical tube (referred to as a PZT (lead zirconate titanate) tube scanner) scans the sharp metallic probe over the surface in a raster pattern while sensing and outputting the tunneling current to the control station, Fig. 23.3. The digital signal processor (DSP) calculates the desired separation of the tip from the sample by sensing the tunneling current flowing between the sample and the tip. The bias voltage applied between the sample and the tip encourages the tunneling current to flow. The DSP completes the digital feedback loop by outputting the desired voltage to the piezoelectric tube. The STM operates in both the *constant height* and *constant current* modes depending on a parameter selection in the control panel. In the constant current mode, the feedback gains are set high, the tunneling tip closely tracks the sample surface, and the variation in the tip height required to maintain constant tunneling current is measured by the change in the voltage applied to the piezo tube. In the constant height mode, the feedback gains are set low, the tip remains at a nearly constant height as it sweeps over the sample surface, and the tunneling current is imaged.

Physically, the Nanoscope STM consists of three main parts: the head, which houses the piezoelectric tube scanner for three-dimensional motion of the tip and the preamplifier circuit (FET input amplifier) mounted on top of the head for the tunneling current; the base on which the sample is mounted; and the base support, which supports the base and head [23.10]. The base accommodates samples up to 10 mm by 20 mm and 10 mm in thickness. Scan sizes available for the STM are 0.7 µm (for atomic resolution), 12 µm, 75 µm and 125 µm square.

The scanning head controls the three-dimensional motion of the tip. The removable head consists of a piezo tube scanner, about 12.7 mm in diameter, mounted into an Invar shell used to minimize vertical thermal drifts because of good thermal match between the piezo tube and the Invar. The piezo tube has separate electrodes for x, y, and z which are driven by separate

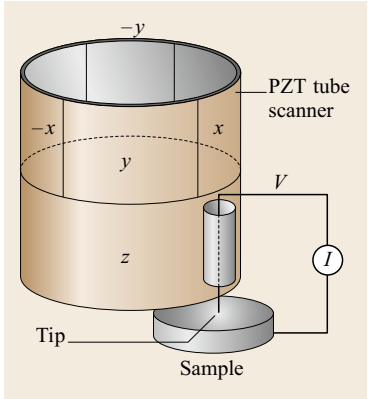


Fig. 23.3 Principle of operation of a commercial STM, a sharp tip attached to a piezoelectric tube scanner is scanned on a sample

drive circuits. The electrode configuration (Fig. 23.3) provides x and y motions which are perpendicular to each other, minimizes horizontal and vertical coupling, and provides good sensitivity. The vertical motion of the tube is controlled by the z electrode which is driven by the feedback loop. The x and y scanning motions are each controlled by two electrodes, which are driven by voltages of the same magnitudes, but opposite signs. These electrodes are called $-y$, $-x$, $+y$, and $+x$. Applying complimentary voltages allows a short, stiff tube to provide a good scan range without large voltages. The motion of the tip due to external vibrations is proportional to the square of the ratio of vibration frequency to the resonant frequency of the tube. Therefore, to minimize the tip vibrations, the resonant frequencies of the tube are high, about 60 kHz in the vertical direction and about 40 kHz in the horizontal direction. The tip holder is a stainless steel tube with a 300 μm inner diameter for 250 μm diameter tips, mounted in ceramic in order to keep the mass on the end of the tube low. The tip is mounted either on the front edge of the tube (to keep mounting mass low and resonant frequency high) (Fig. 23.3) or the center of the tube for large-range scanners, namely 75 and 125 μm (to preserve the symmetry of the scanning). This commercial STM accepts any tip with a 250 μm diameter shaft. The piezo tube requires x–y calibration which is carried out by imaging an appropriate calibration standard. Cleaved graphite is used for the small-scan length head while two-dimensional grids (a gold-plated rule) can be used for longer range heads.

The Invar base holds the sample in position, supports the head, and provides coarse x–y motion for the sample. A spring-steel sample clip with two thumb screws holds the sample in place. An x–y translation stage built into the base allows the sample to be repositioned under the tip. Three precision screws arranged in a triangular pattern support the head and provide coarse and fine adjustment of the tip height. The base support consists of the base support ring and the motor housing. The stepper motor enclosed in the motor housing allows the tip to be engaged and withdrawn from the surface automatically.

Samples to be imaged with STM must be conductive enough to allow a few nanoamperes of current to flow from the bias voltage source to the area to be scanned. In many cases, nonconductive samples can be coated with a thin layer of a conductive material to facilitate imaging. The bias voltage and the tunneling current depend on the sample. Usually they are set at a standard value for engagement and fine-tuned to enhance the quality of the image. The scan size depends on the sample and the features of interest. A maximum scan rate of 122 Hz can be used. The maximum scan rate is usually related to the scan size. A scan rate above 10 Hz is used for small scans (typically 60 Hz for atomic-scale imaging with a 0.7 μm scanner). The scan rate should be lowered for large scans, especially if the sample surfaces are rough or contain large steps. Moving the tip quickly along the sample surface at high scan rates with large scan sizes will usually lead to a tip crash. Essentially, the scan rate should be inversely proportional to the scan size (typically 2–4 Hz for 1 μm, 0.5–1 Hz for 12 μm, and 0.2 Hz for 125 μm scan sizes). Scan rate in length/time, is equal to scan length divided by the scan rate in Hz. For example, for 10 μm × 10 μm scan size scanned at 0.5 Hz, the scan rate is 10 μm/s. Typically, 256 × 256 data formats are most commonly used. The lateral resolution at larger scans is approximately equal to scan length divided by 256.

Figure 23.4 shows an example of STM images of an evaporated C_{60} film on gold-coated freshly cleaved mica taken at room temperature and ambient pres-

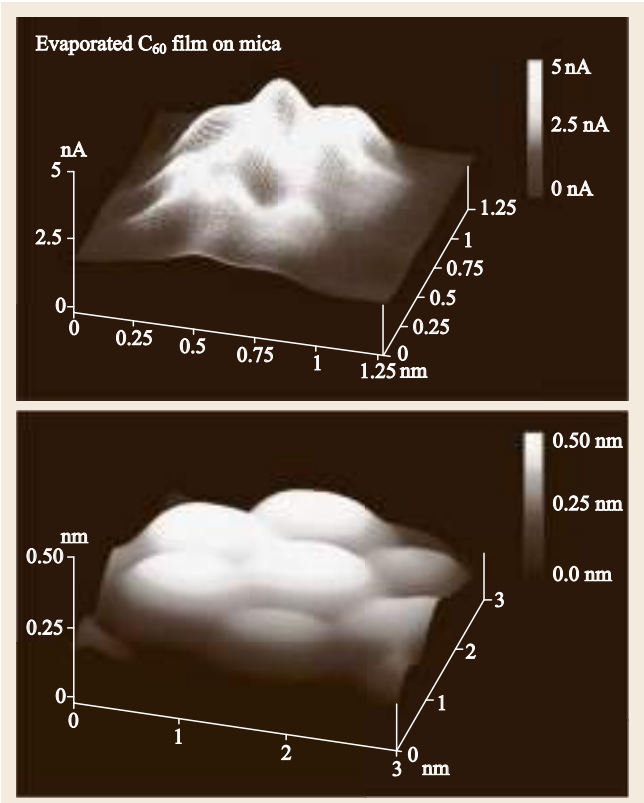


Fig. 23.4 STM images of evaporated C_{60} film on a gold-coated freshly cleaved mica using a mechanically sheared Pt-Ir (80/20) tip in constant height mode (after [23.123])

sure [23.123]. Images with atomic resolution at two scan sizes are obtained. Next, we describe STM designs that are available for special applications.

Electrochemical STM

Electrochemical STM is used to perform and monitor the electrochemical reactions inside the STM. It includes a microscope base with an integral potentiostat, a short head with a 0.7 μm scan range and a differential preamp and the software required to operate the potentiostat and display the result of the electrochemical reaction.

Standalone STM

Standalone STMs, which rest directly on the sample, are able to scan large samples. They are similar to the standard STM, except the sample base has been eliminated.

23.1.3 STM Probe Construction

The STM probe should have a cantilever integrated with a sharp metal tip with a low aspect ratio (tip length/tip shank) to minimize flexural vibrations. Ideally, the tip should be atomically sharp, but in practice, most tip preparation methods produce a tip with a rather ragged profile and that consists of several asperities with the one closest to the surface responsible for tunneling. STM cantilevers with sharp tips are typically fabricated from metal wires of tungsten (W), platinum-iridium (Pt-Ir), or gold (Au) and sharpened by grinding, cutting with a wire cutter or razor blade, field emission/evaporator, ion milling, fracture, or electrochemical polishing/etching [23.124, 125]. The two most commonly used tips are made from either a Pt-Ir (80/20) alloy or tungsten wire. Iridium is used to provide stiffness. The Pt-Ir tips are generally mechanically formed and are readily available. The tungsten tips are etched from tungsten wire with an electrochemical process, for example by using 1 molar KOH solution with a platinum electrode in an electrochemical cell at about 30 V. In general, Pt-Ir tips provide better atomic resolution than tungsten tips, probably due to the lower reactivity of Pt. But tungsten tips are more uniformly shaped and may perform better on samples with steeply sloped features. The tungsten wire diameter used for the cantilever is typically 250 μm with the radius of curvature ranging from 20 to 100 nm and a cone angle ranging from 10 to 60°, Fig. 23.5. The wire can be bent in an L shape, if required for use in the instrument. For calculations of normal spring constant and natural frequency of round cantilevers, see [23.113].

For imaging of deep trenches, high-aspect-ratio controlled geometry (CG) Pt-Ir probes are used, Fig. 23.6. These probes are electrochemically etched from Pt-Ir (80/20) wire and polished to a specific shape which is consistent from tip to tip. Probes have a full cone angle of approximately 15° and a tip radius of less than 50 nm. For imaging of very deep trenches (> 0.25 μm) and nanofeatures, focused ion beam (FIB)-milled CG probes with an extremely sharp tip (radius < 5 nm) are used. For electrochemistry, Pt-Ir probes are coated with a nonconducting film (not shown in the figure).

Pt alloy and W tips are very sharp and have high resolution, but are fragile and sometimes break when contacting a surface. Diamond tips have been used by *Kaneko* and *Oguchi* [23.126]. The diamond tip made conductive by boron ion implantation is found to be chip resistant.



Fig. 23.5 Schematic of a typical tungsten cantilever with a sharp tip produced by electrochemical etching

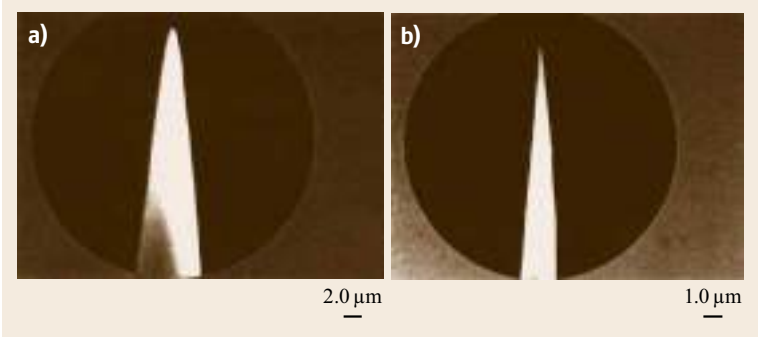


Fig. 23.6a,b Schematics of **(a)** CG Pt-Ir probe, and **(b)** CG Pt-Ir FIB-milled probe

23.2 Atomic Force Microscope

Like the STM, AFM relies on a scanning technique to produce very high resolution, 3-D images of sample surfaces. AFM measures ultrasmall forces (less than 1 nN) present between the AFM tip surface and a sample surface. These small forces are measured by measuring the motion of a very flexible cantilever beam having an ultrasmall mass. While STM requires that the surface to be measured be electrically conductive, AFM is capable of investigating surfaces of both conductors and insulators on an atomic scale if suitable techniques for measurement of cantilever motion are used. In the operation of high-resolution AFM, the sample is generally scanned, rather than the tip as in STM, because AFM measures the relative displacement between the cantilever surface and reference surface, and any cantilever movement would add vibrations. For measurements of large samples, AFMs are available where the tip is scanned and the sample is stationary. As long as the AFM is operated in the so-called contact mode, little if any vibration is introduced.

The AFM combines the principles of the STM and the stylus profiler, Fig. 23.7. In an AFM, the force between the sample and tip is detected rather than the tunneling current to sense the proximity of the tip to the sample. The AFM can be used either in static or dynamic mode. In the static mode, also referred to as repulsive mode or contact mode [23.2], a sharp tip at the end of a cantilever is brought in contact with a sample surface. During initial contact, the atoms at the end of the tip experience a very weak repulsive force due to electronic orbital overlap with the atoms in the sample surface. The force acting on the tip causes a cantilever deflection which is measured by tunneling, capacitive, or optical detectors. The deflection can be measured to within 0.02 nm, so for a typical cantilever spring constant of 10 N/m a force as low as 0.2 nN (corresponding normal pressure $\approx$ 200 MPa for a Si_3N_4 tip with radius of about 50 nm against single-crystal silicon) can be detected. (To put these numbers in perspective, individual atoms and human hair are typically a fraction of a nanometer and about 75 μm in diameter, respectively, and an eyelash and a drop of water have a mass of about 100 nN and 10 μN, respectively.)

In the dynamic mode of operation for the AFM, also referred to as attractive force imaging or noncontact imaging mode (noncontact-AFM or nc-AFM operation), the tip is brought in close proximity (within a few nm) to, and not in contact with the sample [23.127, 128]. The cantilever is deliberately vibrated either in amplitude modulation (AM) mode [23.74] or frequency modulation (FM) mode [23.74, 113, 129, 130]. Very weak van der Waals attractive forces are present at the tip-sample interface. In this technique, the normal pressure exerted at the interface is close to zero (desirable to avoid any surface deformation). In the two modes, surface topography is measured by laterally scanning the sample under the tip while simultaneously measuring the separation-dependent force or force gradient (derivative) between the tip and the surface, Fig. 23.7.

In the contact (static) mode, the interaction force between tip and sample is measured by measuring the cantilever deflection. In the noncontact (or dynamic) mode, the force gradient is obtained by vibrating the cantilever and measuring the shift of resonant frequency of the cantilever. To obtain topographic information, the interaction force is either recorded directly, or used as a control parameter for a feedback circuit that maintains the force or force derivative at a constant value. With an AFM operated in the contact mode, topographic images with a vertical resolution of less than 0.1 nm (as low as 0.01 nm) and a lateral resolution of about 0.2 nm have been obtained [23.3, 59, 131–135]. With a 0.01 nm displacement sensitivity, 10 nN to 1 pN forces are measurable. These forces are comparable to the forces associated with chemical bonding, e.g., 0.1 μN for an ionic bond and 10 pN for a hydrogen bond [23.2]. For further reading, see [23.113–115, 119, 121, 136–140].

Lateral forces being applied at the tip during scanning in the contact mode affect roughness measurements [23.141]. To minimize the effects of friction and other lateral forces in the topography measurements in contact-mode AFMs and to measure topography of soft surfaces, AFMs can be operated in the so-called tapping mode or amplitude modulation mode [23.35, 142].

STM is ideal for atomic-scale imaging. To obtain atomic resolution with AFM, the spring constant of the cantilever should be weaker than the equivalent spring

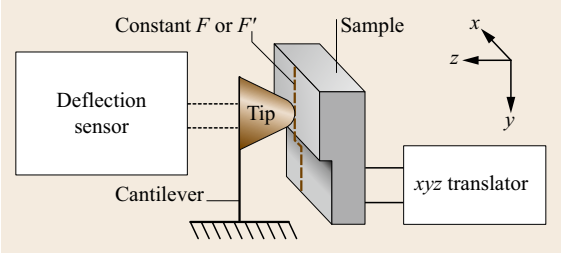


Fig. 23.7 Principle of operation of the AFM. Sample mounted on a piezoelectric scanner is scanned against a short tip and the cantilever deflection is measured, mostly, using a laser deflection technique. Force (contact mode) or force gradient (noncontact mode) is measured during scanning

between atoms. For example, the vibration frequencies ω of atoms bound in a molecule or in a crystalline solid are typically 10^{13} Hz or higher. Combining this with the mass of the atoms m, on the order of 10^{-25} kg, gives interatomic spring constants k, given by $\omega^2 m$, on the order of 10 N/m [23.136]. (For comparison, the spring constant of a piece of household aluminum foil that is 4 mm long and 1 mm wide is about 1 N/m.) Therefore, a cantilever beam with a spring constant of about 1 N/m or lower is desirable. Tips have to be as sharp as possible. Tips with a radius ranging from 20 to 50 nm are commonly available.

Atomic resolution cannot be achieved with these tips at the normal load in the nN range. Atomic structures at these loads have been obtained from lattice imaging or by imaging of the crystal periodicity. Reported data show either perfectly ordered periodic atomic structures or defects on a larger lateral scale, but no well-defined, laterally resolved, atomic-scale defects like those seen in images routinely obtained with STM. Interatomic forces with one or several atoms in contact are 20–40 or 50–100 pN, respectively. Thus, atomic resolution with AFM is only possible with a sharp tip on a flexible cantilever at a net repulsive force of 100 pN or lower [23.143]. Upon increasing the force from 10 pN, *Ohnesorge* and *Binnig* [23.143] observed that monoatomic steplines were slowly wiped away and a perfectly ordered structure was left. This observation explains why mostly defect-free atomic resolution has been observed with AFM. Note that for atomic-resolution measurements, the cantilever should not be too soft to avoid jumps, and also that measurements in the noncontact imaging mode may be desirable for imaging with atomic resolution.

The key component in AFM is the sensor for measuring the force on the tip due to its interaction with the sample. A cantilever (with a sharp tip) with extremely low spring constants is required for high vertical and lateral resolutions at small forces (0.1 nN or lower), but at the same time, a high resonant frequency (about 10–100 kHz) is necessary in order to minimize the sensitivity to vibration noise from the building in which the AFM is located close to 100 Hz. This requires a spring with extremely low vertical spring constant (typically 0.05–1 N/m) as well as low mass (on the order of 1 ng). Today, the most advanced AFM cantilevers are microfabricated from silicon or silicon nitride using photolithographic techniques. (For further details on cantilevers, see a later section). Typical lateral dimensions are on the order of 100 µm, with thicknesses on the order of 1 µm. The force on the tip due to its interaction with the sample is sensed by detecting the deflection of the compliant lever with a known spring constant. This cantilever deflection (displacement smaller than 0.1 nm) has been measured by detecting tunneling current similar to that used in STM in the pioneering work of *Binnig* et al. [23.2] and later used by *Giessibl* et al. [23.65], by capacitance detection [23.144, 145], piezoresistive detection [23.146, 147], and by four optical techniques namely:

1. By optical interferometry [23.4, 5, 148, 149] and with the use of optical fibers [23.66, 150]
2. By optical polarization detection [23.81, 151]
3. By laser diode feedback [23.152]
4. By optical (laser) beam deflection [23.6, 7, 62, 132, 133].

Schematics of the four more commonly used detection systems are shown in Fig. 23.8. The tunneling method originally used by *Binnig* et al. [23.2] in the first version of the AFM uses a second tip to monitor the deflection of the cantilever with its force-sensing tip. Tunneling is rather sensitive to contaminants, and the interaction between the tunneling tip and the rear side of the cantilever can become comparable to the interaction between the tip and sample. Tunneling is rarely used and was mentioned earlier for historical purposes. *Giessibl* et al. [23.65] used it for a low-temperature AFM/STM design. In contrast to tunneling, other deflection sensors are far away from the cantilever at distances of microns to tens of mm. The optical detection methods are believed to be more sensitive, reliable, and easily implemented than other techniques [23.113, 139]. The optical beam deflection method has the largest working distance, is insensitive to distance changes, and is capable of measuring angular changes (friction forces). Therefore, it is the most commonly used method in commercial SPMs.

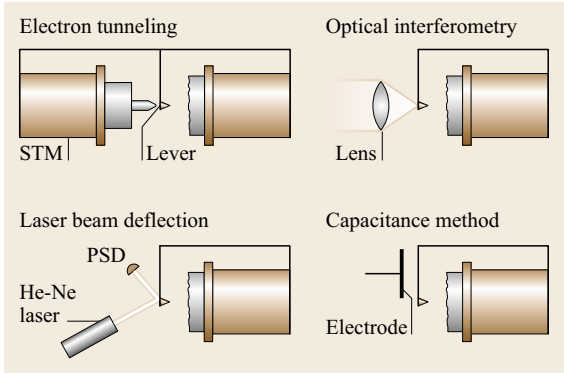


Fig. 23.8 Schematics of the four detection systems for measurement of cantilever deflection. In each set up, the sample mounted on the piezoelectric body is shown *on the right*, the cantilever *in the middle*, and the corresponding deflection sensor *on the left* (after [23.139])

Almost all SPMs use piezo translators to scan the sample or, alternatively, to scan the tip. An electric field applied across a piezoelectric material causes a change in the crystal structure, with expansion in some directions and contraction in others. A net change in volume also occurs [23.153]. The first STM used a piezo tripod for scanning [23.1]. The piezo tripod is one way to generate three-dimensional movement of a tip attached to its center. However, the tripod needs to be fairly large (≈ 50 mm) to get a suitable range. Its size and asymmetric shape makes it susceptible to thermal drift. Tube scanners are widely used in AFMs [23.154]. These provide ample scanning range within a small size. Control electronics systems for AFMs can use either analog or digital feedback. Digital feedback circuits are better suited for ultralow noise operation.

Images from AFMs need to be processed. An ideal AFM is a noise-free device that images a sample with perfect tips of known shape and has perfect linear scanning piezo. In reality, scanning devices are affected by distortions for which corrections must be made. The distortions can be linear and nonlinear. Linear distortions mainly result from imperfections in the machining of the piezo translators, causing cross talk from the z-piezo to the x- and y-piezos, and vice versa. Nonlinear distortions mainly result because of the presence of a hysteresis loop in piezoelectric ceramics. These may also result if the scan frequency approaches the upper frequency limit of the x- and y-drive amplifiers or the upper frequency limit of the feedback loop (z-component). In addition, electronic noise may be present in the system. The noise is removed by digital filtering in the real space [23.155] or in the spatial frequency domain (Fourier space) [23.156].

Processed data consists of many tens of thousands of points per plane (or data set). The output of the first STM and AFM images were recorded on an x–y chart recorder, with the z-value plotted against the tip position in the fast scan direction. Chart recorders have slow response so computers are used for display of the data. The data are displayed as a wire mesh display or a gray scale display (with at least 64 shades of gray).

23.2.1 The AFM Design of Binnig et al.

In the first AFM design developed by *Binnig* et al. [23.2], AFM images were obtained by measurement of the force on a sharp tip created by the proximity to the surface of the sample mounted on a 3-D piezoelectric scanner. Tunneling current between STM tip and the backside of the cantilever beam with attached tip was measured to obtain the normal force. This force was kept at a constant level with a feedback mechanism. The STM tip was also mounted on a piezoelectric element to maintain the tunneling current at a constant level.

23.2.2 Commercial AFM

A review of early AFM designs is presented by *Bhushan* [23.10]. There are a large number of commercial AFMs available on the market. Major manufacturers of AFMs for use in an ambient environment are: Bruker Instruments, Inc. (formerly Digital Instruments Inc. and later Veeco Instruments Inc.), Santa Barbara, California; Agilent Technologies, Santa Clara, California, Park Systems Corp., Suwon, Korea; Asylum Research, Santa Clara, California; NT-MDT Spectrum Instruments, Moscow, Russia; Nanosurf AF, Liestal, Switzerland; Seiko Instruments, Japan; and Olympus, Japan. A major manufacturer of AFM/STMs for use in UHV (ultra-high vacuum) and low-temperature (LT) environments is Scienta Omicron GMBH, Taunusstein, Germany.

We describe here two typical commercial AFMs – small-sample and large-sample AFMs – for operation in the contact mode, produced by Bruker Instruments, Santa Barbara, CA, with scanning lengths ranging from about 0.7 μm (for atomic resolution) to about 125 μm [23.8, 132, 135, 157]. The original design of these AFMs comes from [23.62]. Basically, the AFM scans the sample in a raster pattern while outputting the cantilever deflection error signal to the control station. The cantilever deflection (or the force) is measured using the laser deflection technique, Fig. 23.9. The DSP in the workstation controls the z-position of the piezo based on the cantilever deflection error signal. The AFM operates in both the *constant height* and *constant force* modes. The DSP always adjusts the height of the sample under the tip based on the cantilever deflection error signal, but if the feedback gains are low, the piezo remains at a nearly *constant height* and the cantilever deflection data is collected. With high gains, the piezo height changes to keep the cantilever deflection nearly constant (therefore the force is constant) and the change in piezo height is collected by the system.

To describe further the operation principles of the small-sample commercial AFM shown in Fig. 23.9a, the sample, generally no larger than 10 mm × 10 mm, is mounted on a PZT tube scanner which consists of separate electrodes to precisely scan the sample in the x–y plane in a raster pattern, and to move the sample in the vertical (z) direction. A sharp tip at the free end of a flexible cantilever is brought in contact with the sample. Features on the sample surface cause the cantilever to deflect in the vertical and lateral directions as the sample moves under the tip. A laser beam from a diode laser (5 mW max peak output at 670 nm)

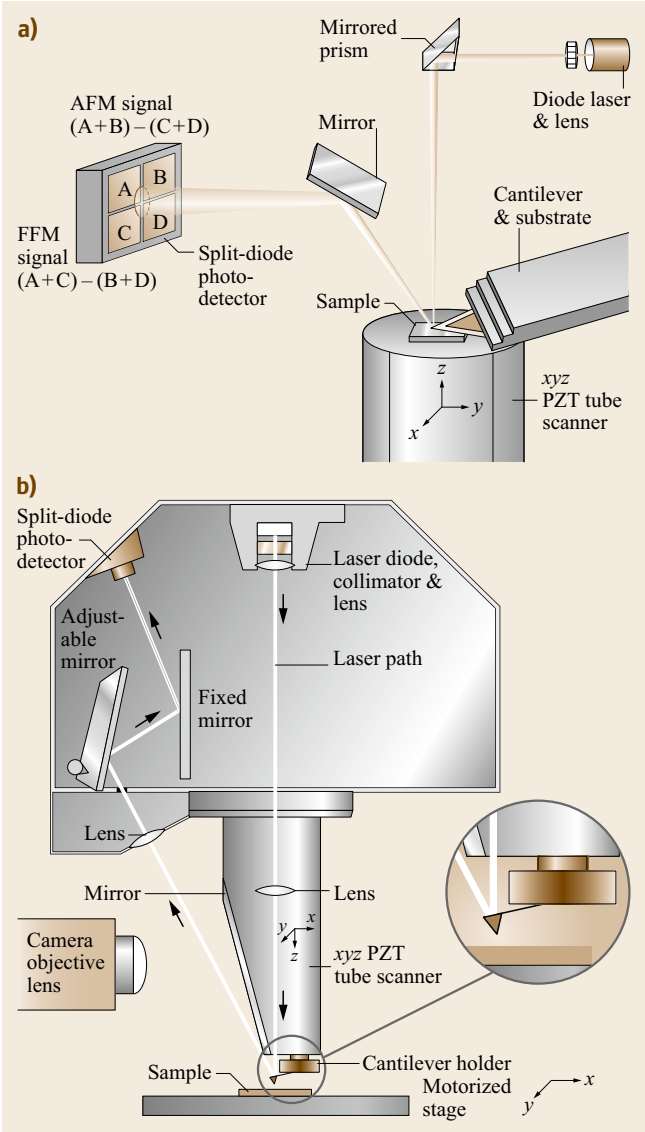


Fig. 23.9a,b Principles of operation of **(a)** a commercial small-sample AFM/FFM, and **(b)** a large-sample AFM/FFM

is directed by a prism onto the back of a cantilever near its free end, tilted downward at about 10° with respect to the horizontal plane. The reflected beam from the vertex of the cantilever is directed through a mirror onto a quad photodetector (split photodetector with four quadrants) (commonly called a position-sensitive detector (PSD), produced by Silicon Detector Corp., Camarillo, California). The differential signal from the top and bottom photodiodes provides the AFM signal, which is a sensitive measure of the cantilever vertical deflection. Topographic features of the sample cause the tip to deflect in the vertical direction as the sample is scanned under the tip. This tip deflection will change the direction of the reflected laser beam, changing the intensity difference between the top and bottom sets of photodetectors (AFM signal). In the AFM operating mode called the height mode, for topographic imaging or for any other operation in which the applied normal force is to be kept constant, a feedback circuit is used to modulate the voltage applied to the PZT scanner to adjust the height of the PZT. This ensures that the cantilever vertical deflection (given by the intensity difference between the top and bottom detector), and consequently normal load, will remain constant during scanning. The PZT height variation is thus a direct measure of the surface roughness of the sample.

In a large-sample AFM, both force sensors using the optical deflection method and scanning unit are mounted on the microscope head, Fig. 23.9b. Because of vibrations added by cantilever movement, lateral resolution of this design is somewhat poorer than the design in Fig. 23.9a in which the sample is scanned instead of the cantilever beam. The advantage of the large-sample AFM is that large samples can be measured readily.

Most AFMs can be used for topography measurements in the so-called tapping mode (intermittent contact mode), also referred to as dynamic force microscopy in the amplitude mode (AM) mentioned earlier. In the tapping mode, during scanning over the surface, the cantilever/tip assembly is sinusoidally vibrated by a piezo mounted above it, and the oscillating tip slightly taps the surface at the resonant frequency of the cantilever (70–400 kHz) with a constant (20–100 nm) oscillating amplitude introduced in the vertical direction with a feedback loop keeping the average normal force constant, Fig. 23.10. The oscillating amplitude is kept large enough so that the tip does not get stuck to the sample because of adhesive attractions. The tapping mode is used in topography measurements to minimize effects of friction and other lateral forces to measure topography of soft surfaces.

Topographic measurements are made at any scanning angle. At a first instance, scanning angle may not appear to be an important parameter. However, the friction force between the tip and the sample will affect the topographic measurements in a parallel scan (scanning along the long axis of the cantilever). Therefore, a perpendicular scan may be more desirable. Generally, one picks a scanning angle that gives the same topographic data in both directions. This angle may be slightly different than that for the perpendicular scan.

For measurement of friction force being applied at the tip surface during sliding, left hand and right hand sets of quadrants of the photodetector are used. In the so-called friction mode, the sample is scanned back and

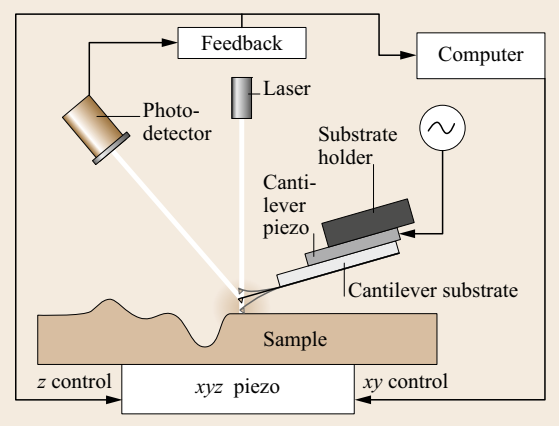


Fig. 23.10 Schematic of tapping mode used for surface roughness measurements

forth in a direction orthogonal to the long axis of the cantilever beam. A friction force between the sample and the tip will produce a twisting of the cantilever. As a result, the laser beam will be reflected out of the plane defined by the incident beam and the beam reflected vertically from an untwisted cantilever. This produces an intensity difference of the laser beam received in the left hand and right hand sets of quadrants of the photodetector. The intensity difference between the two sets of detectors (FFM signal) is directly related to the degree of twisting and hence to the magnitude of the friction force. This method provides three-dimensional maps of friction force. One problem associated with this method is that any misalignment between the laser beam and the photodetector axis introduces error in the measurement. However, by following the procedures developed by *Ruan* and *Bhushan* [23.157] in which the average FFM signal for the sample scanned in two opposite directions is subtracted from the friction profiles of each of the two scans, the misalignment effect is eliminated. By following the friction force calibration procedures developed by *Ruan* and *Bhushan* [23.157], voltages corresponding to friction forces can be converted to force units. The coefficient of friction is obtained from the slope of friction force data measured as a function of normal loads typically ranging from 10 to 150 nN. This approach eliminates any contributions due to adhesive forces [23.36]. For calculation of the coefficient of friction based on a single point measurement, friction force should be divided by the sum of applied normal load and intrinsic adhesive force. Furthermore, note that for a single asperity contact, the coefficient of friction is not independent of load. For further details, refer to a later section.

The tip is scanned in such a way that its trajectory on the sample forms a triangular pattern, Fig. 23.11. Scanning speeds in the fast and slow scan directions depend on the scan area and scan frequency. Scan sizes ranging from less than 1 nm × 1 nm to 125 μm × 125 μm and scan rates from less than 0.5 to 122 Hz typically can be used. Higher scan rates are used for smaller scan lengths. For example, scan rates in the fast and slow scan directions for an area of 10 μm × 10 μm scanned at 0.5 Hz are 10 μm/s and 20 nm/s, respectively.

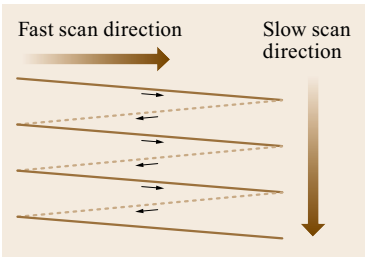


Fig. 23.11 Schematic of triangular pattern trajectory of the AFM tip as the sample is scanned in two dimensions. During imaging, data are recorded only during scans along the *solid scan lines*

We now describe the construction of a small sample AFM in more detail. It consists of three main parts: the optical head that senses the cantilever deflection, a PZT tube scanner that controls the scanning motion of the sample mounted on its one end, and the base that supports the scanner and head and includes circuits for the deflection signal, Fig. 23.12a. The AFM connects directly to a control system. The optical head consists of a laser diode stage, a photodiode stage preamp board, a cantilever mount and its holding arm, and a deflection beam reflecting mirror, Fig. 23.12b. The laser diode stage is a tilt stage used to adjust the position of the laser beam relative to the cantilever. It consists of the laser diode, collimator, focusing lens, baseplate, and the x and y laser diode positioners. The positioners are used to place the laser spot on the end of the cantilever. The photodiode stage is an adjustable stage used to position the photodiode elements relative to the reflected laser beam. It consists of the split photodiode, the base plate, and the photodiode positioners. The deflection beam reflecting mirror is mounted on the upper left in the interior of the head which reflects the deflected beam toward the photodiode. The cantilever mount is a metal (for operation in air) or glass (for operation in water) block which holds the cantilever firmly at the proper angle, Fig. 23.12d. Next, the tube scanner consists of an Invar cylinder holding a single tube made of piezoelectric crystal which provides the necessary three-dimensional motion to the sample. Mounted on top of the tube is a magnetic cap on which the steel sample puck is placed. The tube is rigidly held at one end with the sample mounted on the other end of the tube. The scanner also contains three fine-pitched screws which form the mount for the optical head. The optical head rests on the tips of the screws which are used to adjust the position of the head relative to the sample.

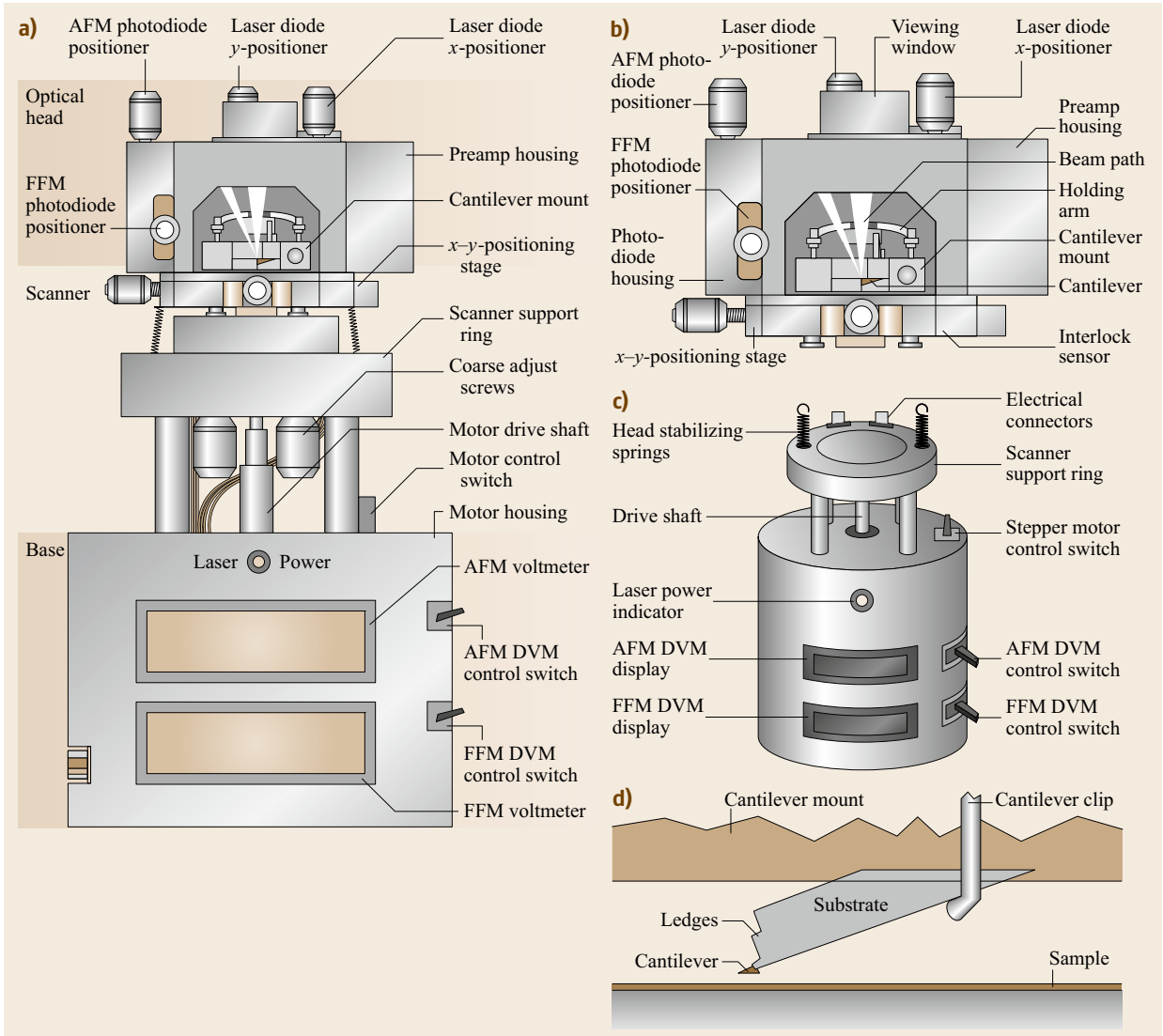


Fig. 23.12a–d Schematics of a commercial multimode AFM/FFM made by Bruker Instruments Inc. **(a)** Front view, **(b)** optical head, **(c)** base, and **(d)** cantilever substrate mounted on cantilever mount (not to scale)

The scanner fits into the scanner support ring mounted on the base of the microscope, Fig. 23.12c. The stepper motor is controlled manually with the switch on the upper surface of the base and automatically by the computer during the tip-engage and tip-withdraw processes.

The scan sizes available for these instruments are 0.7 μm, 12 μm, and 125 μm. The scan rate must be decreased as the scan size is increased. A maximum scan rate of 122 Hz can be used. Scan rates of about 60 Hz should be used for small scan lengths (0.7 μm). Scan rates of 0.5–2.5 Hz should be used for large scans on samples with tall features. High scan rates help reduce drift, but they can only be used on flat samples with small scan sizes. Scan rate, or scanning speed in length/time in the fast scan direction, is equal to twice the scan length times the scan rate in Hz, and in the slow direction, it is equal to scan length times the scan rate in Hz divided by the number of data points in the transverse direction. For example, for $10\,\mu m \times 10\,\mu m$ scan size scanned at 0.5 Hz, the scan rates in the fast and slow scan directions are 10 μm/s and 20 nm/s, respectively. Normally 256×256 data points are taken for each image. The lateral resolution at larger scans is approximately equal to scan length divided by 256. The piezo tube requires *x*–*y* calibration, which is carried out by imaging an appropriate calibration standard.

Cleaved graphite is used for small scan heads while two-dimensional grids (with a gold plating rule) can be used for longer range heads.

Examples of AFM images of freshly cleaved highly oriented pyrolytic (HOP) graphite and mica surfaces are shown in Fig. 23.13 [23.59, 131, 135]. Images with near atomic resolution are obtained.

Force calibration mode is used to study interaction between the cantilever and the sample surface. In the force calibration mode, the x and y voltages applied to the piezo tube are held at zero, and a sawtooth voltage is applied to the z-electrode of the piezo tube, Fig. 23.14a. The force measurement starts with the sample far away and the cantilever in its rest position. As a result of the applied voltage, the sample is moved up and down relative to the stationary cantilever tip. As the piezo moves the sample up and down, the cantilever deflection signal from the photodiode is monitored. The force–distance curve, a plot of the cantilever tip deflection signal as a function of the voltage applied to the piezo tube, is obtained. Figure 23.14b shows a typical force–distance curve showing the various features of the curve. The arrow heads reveal the direction of piezo travel. As the piezo extends, it approaches the tip, which is at this point in free air and hence shows no deflection. This is indicated by the flat portion of the curve. As the tip approaches the sample within a few nanometers (point A), an attractive force exists between the atoms of the tip surface and the atoms of the sample surface. The tip is pulled towards the sample and contact occurs at point B on the graph. From this point on, the tip is in contact with the surface and as the piezo further extends, the tip gets further deflected. This is represented by the sloped portion of the curve. As the piezo retracts, the tip goes beyond the zero deflection (flat) line into the adhesive regime because of attractive forces (van der Waals forces and long-range meniscus forces). At point C in the graph, the tip snaps free of the adhesive forces, and is in free air again. The horizontal distance between points B and C along the retrace line gives the distance moved by the tip in the adhesive regime. This distance multiplied by the stiffness of the cantilever gives the adhesive force. Incidentally, the horizontal shift between the loading and unloading curves results from the hysteresis in the PZT tube [23.10].

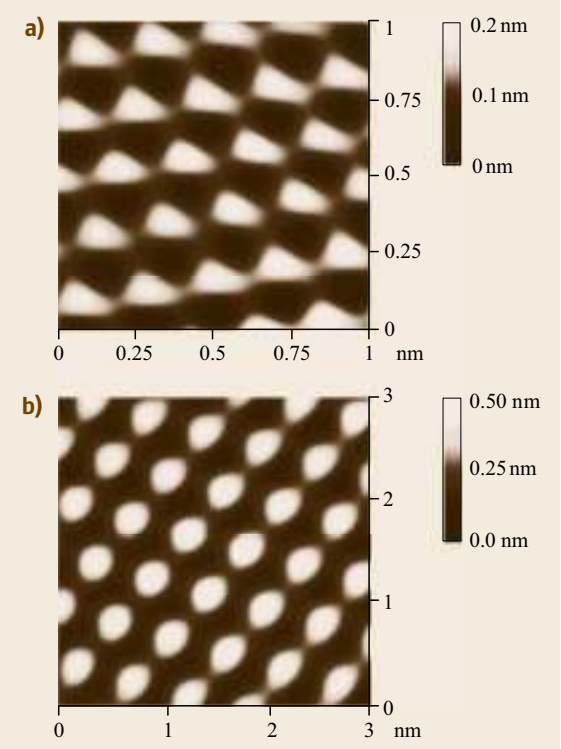


Fig. 23.13a,b Typical AFM images of **(a)** freshly cleaved highly oriented pyrolytic graphite and **(b)** mica surfaces taken using a square pyramidal Si_3N_4 tip

Multimode Capabilities

The multimode AFM can be used for topography measurements in the contact mode and AM or tapping mode as described earlier. It can also be used for measurements of lateral (friction) force, electric force gradients, and magnetic force gradients.

The multimode AFM using a grounded conducting tip can measure electric field gradients by oscillating the tip near its resonant frequency. When the lever encoun-

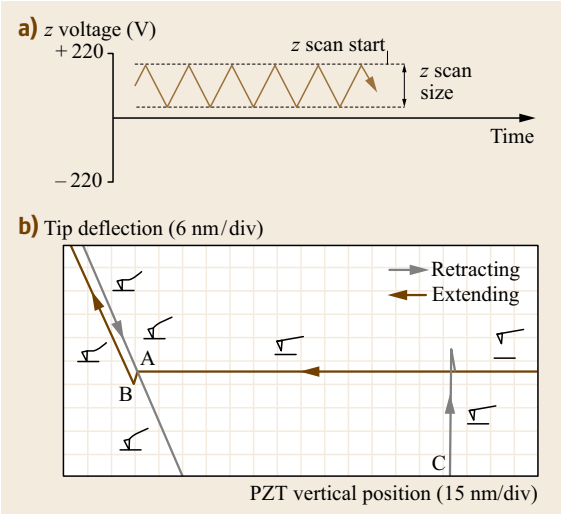


Fig. 23.14 **(a)** Force calibration z-waveform, and **(b)** a typical force–distance curve for a tip in contact with a sample. Contact occurs at point B; tip breaks free of adhesive forces at point C as the sample moves away from the tip

ters a force gradient from the electric field, the effective spring constant of the cantilever is altered, changing its resonant frequency. Depending on which side of the resonance curve is chosen, the oscillation amplitude of the cantilever increases or decreases due to the shift in the resonant frequency. By recording the amplitude of the cantilever, an image revealing the strength of the electric field gradient is obtained.

In a magnetic force microscope (MFM) used with a magnetically coated tip, static cantilever deflection is detected that occurs when a magnetic field exerts a force on the tip and the MFM images of magnetic materials can be produced. MFM sensitivity can be enhanced by oscillating the cantilever near its resonant frequency. When the tip encounters a magnetic force gradient, the effective spring constant, and hence the resonant frequency, is shifted. By driving the cantilever above or below the resonant frequency, the oscillation amplitude varies as the resonance shifts. An image of magnetic field gradients is obtained by recording the oscillation amplitude as the tip is scanned over the sample.

Topographic information is separated from the electric field gradients and magnetic field images by using a so-called lift mode. Measurements in lift mode are taken in two passes over each scan line. On the first pass, topographical information is recorded in the standard tapping mode where the oscillating cantilever lightly taps the surface. On the second pass, the tip is lifted to a user-selected separation (typically 20–200 nm) between the tip and local surface topography. By using the stored topographical data instead of the standard feedback, the separation remains constant without sensing the surface. At this height, cantilever amplitudes are sensitive to electric field force gradients or relatively weak but long-range magnetic forces without being influenced by topographic features. Two-pass measurements are taken for every scan line, producing separate topographic and magnetic force images.

Electrochemical AFM

This option allows study of electrochemical reactions in the AFM. It includes a potentiostat, a fluid cell with a transparent cantilever holder and electrodes, and the software required to operate the potentiostat and display the results of the electrochemical reaction.

23.2.3 AFM Probe Construction

Various probes (cantilevers and tips) are used for AFM studies. The cantilever stylus used in the AFM should meet the following criteria:

1. Low normal spring constant (stiffness)
2. A high resonant frequency
3. A high quality factor of the cantilever Q
4. High lateral spring constant (stiffness)
5. Short cantilever length
6. Incorporation of components (such as mirror) for deflection sensing
7. A sharp protruding tip [23.158].

In order to register a measurable deflection with small forces, the cantilever must flex with a relatively low force (on the order of a few nN) requiring vertical spring constants of 10^{-2}–$10^{2}\,\mathrm{N/m}$ for atomic resolution in the contact profiling mode. The data rate or imaging rate in the AFM is limited by the mechanical resonant frequency of the cantilever. To achieve a large imaging bandwidth, the AFM cantilever should have resonant frequency greater than about 10 kHz (preferably 30–100 kHz) in order to make the cantilever the least sensitive part of the system. Fast imaging rates are not just a matter of convenience, since the effects of thermal drifts are more pronounced with slow scanning speeds. The combined requirements of a low spring constant and a high resonant frequency is met by reducing the mass of the cantilever. The quality factor Q ($= \omega_\mathrm{R}/(c/m)$ where ω_R is the resonant frequency of the damped oscillator, c is the damping constant, and m is the mass of the oscillator) should have a high value for some applications. For example, resonance curve detection is a sensitive modulation technique for measuring small force gradients in noncontact imaging. Increasing the Q increases the sensitivity of the measurements. Mechanical Q values of 100–1000 are typical. In contact modes, Q is less important. A high lateral spring constant in the cantilever is desirable to reduce the effect of lateral forces in the AFM as frictional forces can cause appreciable lateral bending of the cantilever. Lateral bending results in error in the topography measurements. For friction measurements, cantilevers with less lateral rigidity are preferred. A sharp protruding tip must be formed at the end of the cantilever to provide a well-defined interaction with the sample over a small area. The tip radius should be much smaller than the radii of corrugations in the sample in order for these to be measured accurately. The lateral spring constant depends critically on the tip length. Additionally, the tip should be centered at the free end.

In the past, cantilevers have been cut by hand from thin metal foils or formed from fine wires. Tips for these cantilevers were prepared by attaching diamond fragments to the ends of the cantilevers by hand, or, in the case of wire cantilevers, electrochemically etching the wire to a sharp point. Several cantilever geometries for wire cantilevers have been used. The simplest geometry is the L-shaped cantilever, usually made by bending a wire at a 90° angle. Other geometries include single-

Table 23.2 Relevant properties of commonly used materials for cantilevers and diamond used in the some tips

Property	Young's modulus (E) (GPa)	Density (ρg) (kg/m^3)	Microhardness (GPa)	Speed of sound $\left(\sqrt{E/\rho}\right)$ (m/s)
Diamond (for tips)	900–1050	3515	78.4–102	17 000
Si_3N_4	310	3180	19.6	9900
Si	130–188	2330	9–10	8200
Stainless steel	210	8050	2–4	5080
W	350	19 310	3.2	4250
Ir	530	–	≈ 3	5300

V and double-V geometries with a sharp tip attached at the apex of V, and double-X configuration with a sharp tip attached at the intersection [23.34, 159]. These cantilevers can be constructed with high vertical spring constants. For example, a double-cross cantilever with an effective spring constant of 250 N/m was used by *Burnham* and *Colton* [23.34]. The small size and low mass needed in the AFM make hand fabrication of the cantilever a difficult process with poor reproducibility. Conventional microfabrication techniques are ideal for constructing planar thin-film structures which have submicron lateral dimensions. Triangular (V-shaped) cantilevers have improved (higher) lateral spring constants in comparison to rectangular cantilevers. In terms of spring constants, triangular cantilevers are approximately equivalent to two rectangular cantilevers in parallel [23.158]. Although the macroscopic radius of a photolithographically patterned corner is seldom much less than about 50 nm, microscopic asperities on the etched surface provide tips with near atomic dimensions.

Cantilevers and tips have been used from a whole range of materials. Most common are cantilevers/tips made of Si_3N_4, Si, and stainless steel/diamond. Besides the geometry, the material parameters that determine the resonant frequency of the cantilever are Young's modulus and density. Table 23.2 shows the relevant properties of commonly used materials and the speed of sound, indicative of the resonant frequency for a cantilever beam. Hardness is important to judge the durability of the tips, and is also listed in the table. Materials used for STM cantilevers are also included.

Silicon nitride cantilevers with integrated tips are less expensive than those made of other materials. They are very rugged and well suited to imaging in almost all environments. They are especially compatible to organic and biological materials. Microfabricated silicon nitride triangular beams with integrated square pyramidal tips made using plasma-enhanced chemical vapor deposition (PECVD) are the most commonly used [23.158]. Figure 23.15a shows a schematic example of four cantilever beams, each with different sizes and spring constants, arranged in pairs, attached to a borosilicate glass substrate (Pyrex) and marketed by various manufacturers including Bruker Instruments, Inc. Each pair of triangular cantilevers measures 115 and 193 μm from the substrate to the apex, with base widths of 122 and 205 μm, respectively. These cantilevers are commercially available with wide (top) and narrow (bottom) legs, all in the same thickness of 0.6 μm. Calculated spring constant and measured natural frequencies for each of the configurations are listed in Table 23.3. The most commonly used cantilever beam is the 115 μm long, wide-legged cantilever (vertical spring constant = 0.58 N/m). Cantilevers with smaller spring constants should be used on softer samples. Pyramidal tips are highly symmetric with ends having a radius of about 20–50 nm. The tip side walls have a slope of 35°, and the length of the edges of the tip at the cantilever base is about 4 μm.

An alternative to silicon nitride cantilevers with integrated tips are microfabricated single-crystal silicon cantilevers with integrated tips. Si tips are sharper than Si_3N_4 tips because they are directly formed by the anisotropic etch in single-crystal Si rather than using an etch pit as a mask for deposited materials [23.160]. Etched single-crystal n-type silicon rectangular cantilevers with square pyramidal tips with a lower radius of less than 10 nm for contact and tapping mode (tapping mode etched silicon probe or TESP) AFMs are commercially available from various manufacturers in-

Table 23.3 Measured vertical spring constants and natural frequencies of triangular (V-shaped) cantilevers made of PECVD Si_3N_4 (Data provided by Bruker Instruments, Inc.)

Cantilever dimension	Spring constant k_z (N/m)	Natural frequency ω_0 (kHz)
115 μm long, narrow leg	0.38	40
115 μm long, wide leg	0.58	40
193 μm long, narrow leg	0.06	13–22
193 μm long, wide leg	0.12	13–22

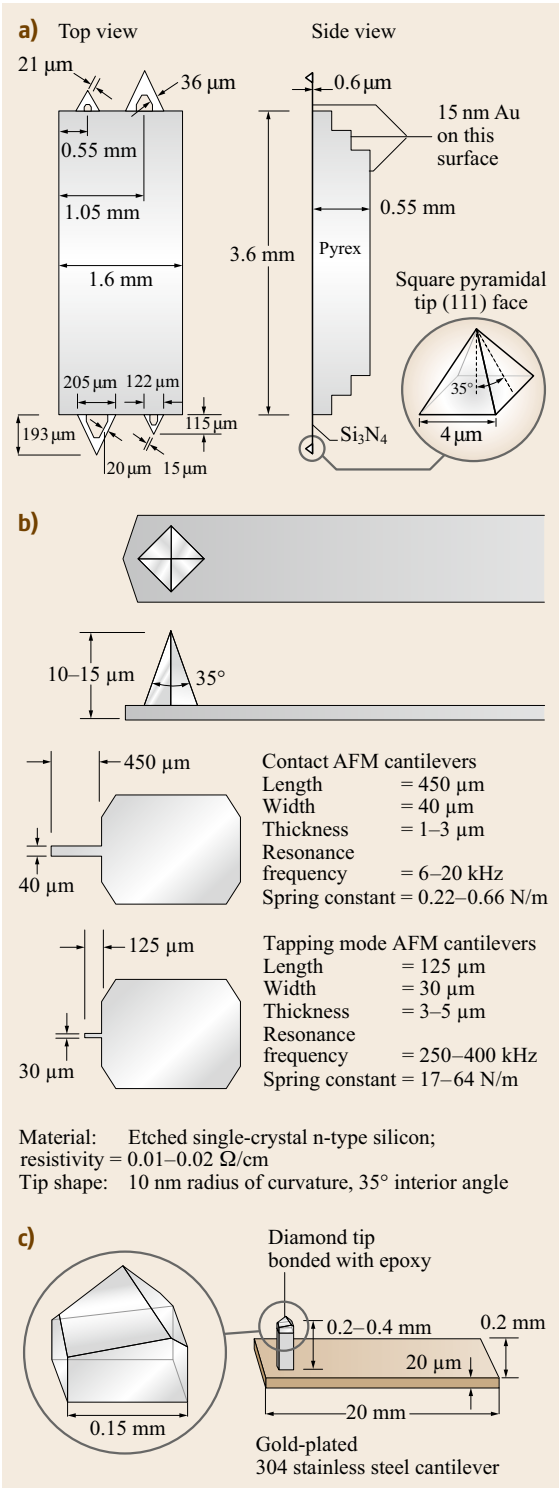


Fig. 23.15a–c Schematics of (a) triangular cantilever beam with square pyramidal tips made of PECVD Si_3N_4, (b) rectangular cantilever beams with square pyramidal tips made of etched single-crystal silicon, and (c) rectangular cantilever stainless steel beam with three-sided pyramidal natural diamond tip ◄

cluding Bruker Instruments, Inc., Fig. 23.15b. Spring constants and resonant frequencies are also presented in Fig. 23.15b.

Commercial triangular Si_3N_4 cantilevers have a typical width-thickness ratio of 10–30, which results in 100 to 1000 times stiffer spring constants in the lateral direction compared to the normal direction. Therefore, these cantilevers are not well suited for torsion. For friction measurements, the torsional spring constant should be minimized in order to be sensitive to lateral forces. Rather long cantilevers with small thicknesses and large tip lengths are most suitable. Rectangular beams have lower torsional spring constants in comparison to the triangular (V-shaped) cantilevers. Table 23.4 lists the spring constants (with full length of the beam used) in three directions of typical rectangular beams. We note that lateral and torsional spring constants are about two orders of magnitude larger than the normal spring constants. A cantilever beam required for the tapping mode is quite stiff and may not be sensitive enough for friction measurements. *Meyer* et al. [23.161] used a specially designed rectangular silicon cantilever with length $= 200\,\mu m$, width $= 21\,\mu m$, thickness $=$

Table 23.4 Vertical (k_z), lateral (k_y), and torsional (k_{yT}) spring constants of rectangular cantilevers made of Si (IBM) and PECVD Si_3N_4 (Bruker Instruments, Inc.)

Dimensions/stiffness	Si cantilever	Si_3N_4 cantilever
Length L (µm)	100	100
Width b (µm)	10	20
Thickness h (µm)	1	0.6
Tip length ℓ (µm)	5	3
k_z (N/m)	0.4	0.15
k_y (N/m)	40	175
k_{yT} (N/m)	120	116
ω_0 (kHz)	≈ 90	≈ 65

Note: $k_z = Ebh^3/4L^3$, $k_y = Eb^3h/4\ell^3$, $k_{yT} = Gbh^3/3L\ell^2$, and $\omega_0 = [k_z/(m_c + 0.24\,bhL\rho)]^{1/2}$ where E is Young's modulus, G is the modulus of rigidity $[= E/2(1+\nu)$, where ν is the Poisson's ratio], ρ is the mass density of the cantilever, and m_c is the concentrated mass of the tip (≈ 4 ng) [23.113]. For Si, $E = 130$ GPa, $\rho g = 2300\,kg/m^3$, and $\nu = 0.3$. For Si_3N_4, $E = 150$ GPa, $\rho g = 3100\,kg/m^3$, and $\nu = 0.3$.

0.4 μm, tip length = 12.5 μm, and shear modulus = 50 GPa, giving a normal spring constant of 0.007 N/m and torsional spring constant of 0.72 N/m which gives a lateral force sensitivity of 10 pN and an angle of resolution of 10^{-7} rad. With this particular geometry, sensitivity to lateral forces could be improved by about a factor of 100 compared with commercial V-shaped Si_3N_4 or rectangular Si or Si_3N_4 cantilevers used by *Meyer* and *Amer* [23.7] with torsional spring constant of ≈ 100 N/m. *Ruan* and *Bhushan* [23.157] and *Bhushan* and *Ruan* [23.8] used 115 μm long, wide-legged V-shaped cantilevers made of Si_3N_4 for friction measurements.

For scratching, wear, and indentation studies, single-crystal natural diamond tips ground to the shape of a three-sided pyramid with an apex angle of either 60 or 80° whose point is sharpened to a radius of about 100 nm are commonly used [23.10, 36] (Fig. 23.15c). The tips are bonded with conductive epoxy to a gold-plated 304 stainless steel spring sheet (length = 20 mm, width = 0.2 mm, thickness = 20–60 μm) which acts as a cantilever. Free length of the spring is varied to change the beam stiffness. The normal spring constant of the beam ranges from about 5–600 N/m for a 20 μm thick beam. The tips are produced by various manufacturers including R-DEC Co., Tsukuba, Japan. Scanning electron microscopy (SEM) micrographs of silicon nitride, single-crystal silicon, and natural diamond tips are shown in Fig. 23.16.

For imaging within trenches by AFM, high aspect ratio tips are used. Examples of the two probes are shown in Fig. 23.17. The high aspect ratio tip (HART) probes are produced by starting with a conventional Si_3N_4 pyramidal probe. Through a combination of focused ion beam (FIB) and high-resolution SEM techniques, a thin filament is grown at the apex of the pyramid. The probe filament is approximately 1 μm long and 0.1 μm in diameter. It tapers to an extremely sharp point (radius better than the resolution of most SEMs). The long thin shape and sharp radius make it ideal for imaging features of microstructures and trenches (> 0.25 μm). Because of flexing of the probe, it is unsuitable for imaging structures at the atomic level since the flexing of the probe can create image artifacts. For atomic-scale imaging, a FIB-milled probe is used which is relatively stiff yet allows for closely spaced topography. These probes start out as conventional Si_3N_4 pyramidal probes but the pyramid is FIB milled until a small cone shape is formed which has a high aspect ratio with 0.2–0.3 μm length. The milled probes allow nanostructure resolution without sacrificing rigidity.

Carbon nanotube tips having small diameter and high aspect ratio are used for high-resolution imaging of surfaces and deep trenches, in the tapping mode or noncontact mode. Single-walled carbon nanotubes (SWNTs) are microscopic graphitic cylinders that are

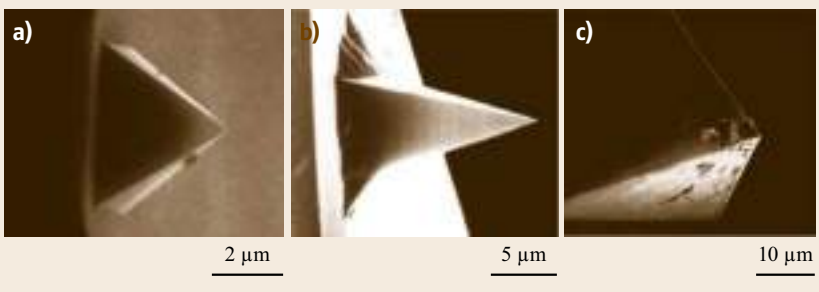


Fig. 23.16a–c SEM micrographs of a square-pyramidal PECVD Si_3N_4 tip **(a)**, a square pyramidal etched single-crystal silicon tip **(b)**, and a three-sided pyramidal natural diamond tip **(c)**

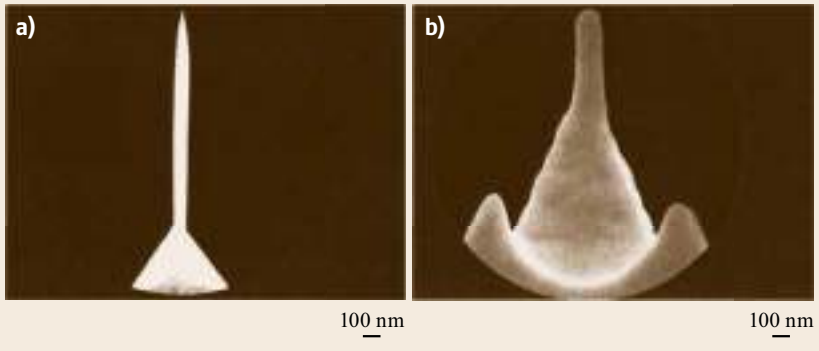


Fig. 23.17a,b Schematics of **(a)** HART Si_3N_4 probe, and **(b)** FIB-milled Si_3N_4 probe

0.7–3 nm in diameter and up to many microns in length. Larger structures called multiwalled carbon nanotubes (MWNTs) consist of nested, concentrically arranged SWNT and have diameters ranging from 3 to 50 nm. MWNT carbon nanotube AFM tips are produced by manual assembly [23.162], chemical vapor deposition (CVD) synthesis, and hybrid fabrication processes [23.163]. Figure 23.18 shows a TEM micrograph of a carbon nanotube tip, ProbeMax, commercially produced using mechanical assembly by Piezomax Technologies, Inc., Middleton, Wisconsin. For production of these tips, MWNT nanotubes are produced by the carbon arc method. They are physically attached on single-crystal silicon square-pyramidal tips in the SEM using a manipulator and the SEM stage to control the nanotubes and the tip independently. Once the nanotube is attached to the tip, it is usually too long to image with. It is shortened using an AFM applying voltage between the tip and the sample. Nanotube tips are also commercially produced by CVD synthesis by NanoDevices, Santa Barbara, California.

23.2.4 Friction Measurement Methods

On the basis of the work by *Ruan* and *Bhushan* [23.157], the two methods for friction measurements are now described in more detail (also see [23.7, 164]). A scanning angle is defined as the angle relative to the y-axis in Fig. 23.19a. This is also the long axis of the cantilever. A zero degree scanning angle corresponds to the sample scanning in the y-direction, and a 90° scanning angle corresponds to the sample scanning perpendicular to this axis in the xy-plane (along the x-axis). If the scanning direction is in both y- and $-y$-directions, we call this a *parallel scan*. Similarly, a *perpendicular scan* means the scanning direction is in x- and $-x$-directions. The sample traveling direction for each of these two methods is illustrated in Fig. 23.19b.



Fig. 23.18 SEM micrograph of a multiwalled carbon nanotube (MWNT) tip physically attached on the single-crystal silicon, square-pyramidal tip (Courtesy Piezomax Technologies, Inc.)

In method 1 (using the *height* mode with parallel scans) in addition to topographic imaging, it is also possible to measure friction force when the scanning direction of the sample is parallel to the y-direction (parallel scan). If there were no friction force between the tip and the moving sample, the topographic feature would be the only factor to cause the cantilever to deflect vertically. However, friction force does exist on all contact surfaces where one object is moving relative to another. The friction force between the sample and the tip will also cause a cantilever deflection. We assume that the normal force between the sample and the tip is W_0 when the sample is stationary (W_0

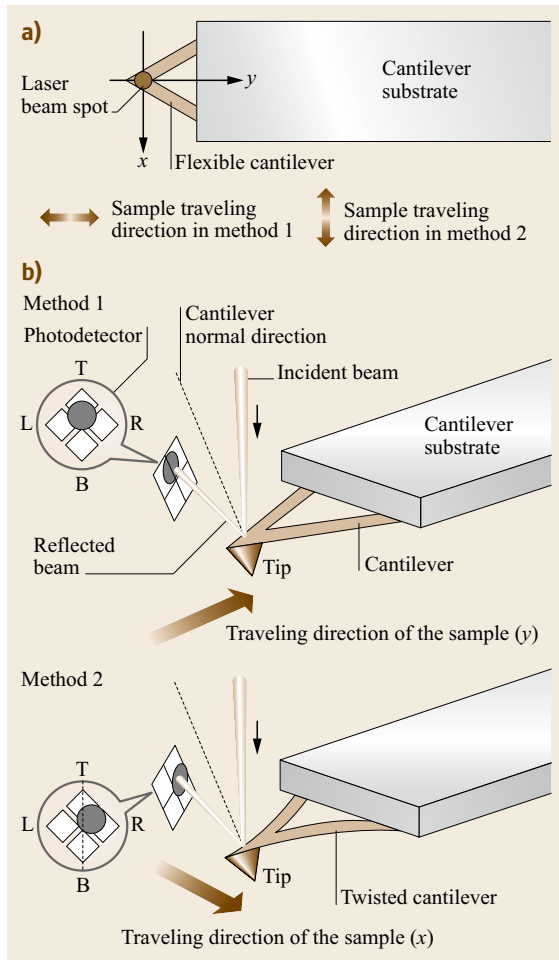


Fig. 23.19 **(a)** Schematic defining the x- and y-directions relative to the cantilever, and showing the sample traveling direction in the two different measurement methods discussed in the text. **(b)** Schematic of deformation of the tip and cantilever shown as a result of sliding in the x- and y-directions. A twist is introduced to the cantilever if the scanning is in the x-direction (Method 2) (after [23.157])

is typically in the range 10–200 nN), and the friction force between the sample and the tip is W_f as the sample scans against the tip. The direction of friction force (W_f) is reversed as the scanning direction of the sample is reversed from the positive (y) to the negative ($-y$) direction $\vec{W}_{f(y)} = -\vec{W}_{f(-y)}$.

When the vertical cantilever deflection is set at a constant level, it is the total force (normal force and friction force) applied to the cantilever that keeps the cantilever deflection at this level. Since the friction force is in opposite directions as the traveling direction of the sample is reversed, the normal force will have to be adjusted accordingly when the sample reverses its traveling direction, so that the total deflection of the cantilever will remain the same. We can calculate the difference of the normal force between the two traveling directions for a given friction force W_f. First, by means of a constant deflection, the total moment applied to the cantilever is constant. If we take the reference point to be the point where the cantilever joins the cantilever holder (substrate), point P in Fig. 23.20, we have the following relationship

$$(W_0 - \Delta W_1)\, L + W_f \ell = (W_0 + \Delta W_2)\, L - W_f \ell \tag{23.1}$$

or

$$(\Delta W_1 + \Delta W_2)\, L = 2 W_f \ell\,. \tag{23.2}$$

Thus,

$$W_f = \frac{(\Delta W_1 + \Delta W_2)\, L}{2\ell}\,, \tag{23.3}$$

where ΔW_1 and ΔW_2 are the absolute value of the changes of normal force when the sample is traveling in $-y$- and y-directions, respectively, as shown in Fig. 23.20; L is the length of the cantilever; and ℓ is the vertical distance between the end of the tip and point P. The coefficient of friction (μ) between the tip and the sample is then given as

$$\mu = \frac{W_f}{W_0} = \left(\frac{\Delta W_1 + \Delta W_2}{W_0}\right)\left(\frac{L}{2\ell}\right)\,. \tag{23.4}$$

In all circumstances, there are adhesive and interatomic attractive forces between the cantilever tip and the sample. The adhesive force can be due to water from the capillary condensation and other contaminants present at the surface which form meniscus bridges [23.10, 165, 166] and the interatomic attractive force includes van der Waals attraction [23.18]. If these forces (and the indentation effect as well, which is usually small for rigid samples) can be neglected, the normal force W_0 is then equal to the initial cantilever deflection H_0 multiplied by the spring constant of the cantilever. $(\Delta W_1 + \Delta W_2)$ can be measured by multiplying the same spring constant by the height difference of the piezo tube between the two traveling directions (y- and $-y$-directions) of the sample. This height difference is denoted as $(\Delta H_1 + \Delta H_2)$, shown schematically in Fig. 23.21. Thus, (23.4) can be rewritten as

$$\mu = \frac{W_f}{W_0} = \left(\frac{\Delta H_1 + \Delta H_2}{H_0}\right)\left(\frac{L}{2\ell}\right)\,. \tag{23.5}$$

Since the vertical position of the piezo tube is affected by the surface topographic profile of the sample in addition to the friction force being applied at the tip, this difference has to be taken point-by-point at the same location on the sample surface as shown in Fig. 23.21. Subtraction of point-by-point measurements may introduce errors, particularly for rough samples. We will come back to this point later. In addition, precise measurement of L and ℓ (which should include the cantilever angle) are also required.

If the adhesive forces between the tip and the sample are large enough that it can not be neglected, one should include it in the calculation. However, there could be a large uncertainty in determining this force, thus an uncertainty in using (23.5). An alternative approach is to make the measurements at different normal loads and to use $\Delta(H_0)$ and $\Delta(\Delta H_1 + \Delta H_2)$ from the measurements in (23.5). Another comment on (23.5) is that, since only the ratio between $(\Delta H_1 + \Delta H_2)$ and H_0 comes into this equation, the piezo tube vertical position H_0 and its position difference $(\Delta H_1 + \Delta H_2)$ can be

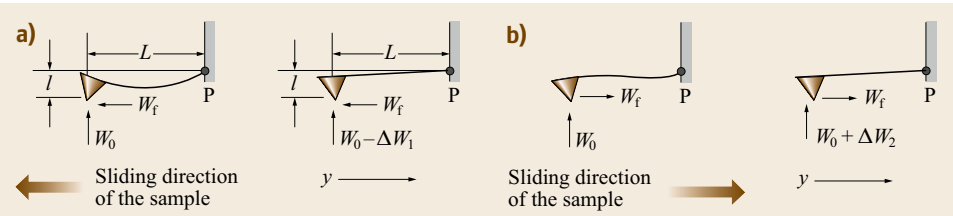


Fig. 23.20 **(a)** Schematic showing an additional bending of the cantilever – due to friction force when the sample is scanned in the y- or $-y$-direction (*left*). **(b)** This effect will be canceled by adjusting the piezo height by a feedback circuit (*right*) (after [23.157])

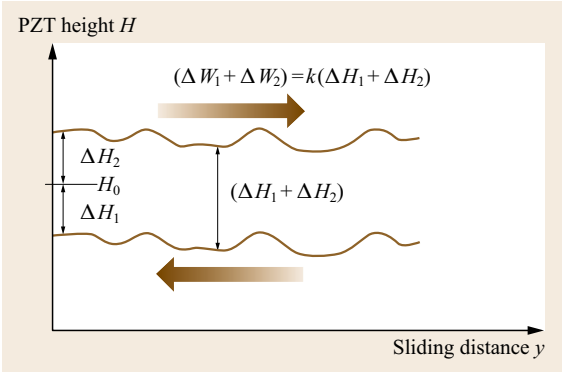


Fig. 23.21 Schematic illustration of the height difference of the piezoelectric tube scanner as the sample is scanned in y- and $-y$-directions

in the units of volts as long as the vertical traveling distance of the piezo tube and the voltage applied to it has a linear relationship. However, if there is a large nonlinearity between the piezo tube traveling distance and the applied voltage, this nonlinearity must be included in the calculation.

It should also be pointed out that (23.4) and (23.5) are derived under the assumption that the friction force W_f is the same for the two scanning directions of the sample. This is an approximation since the normal force is slightly different for the two scans and there may also be a directionality effect in friction. However, this difference is much smaller than W_0 itself. We can ignore the second-order correction.

Method 2 (*aux* mode with perpendicular scan) to measure friction was suggested by *Meyer* and *Amer* [23.7]. The sample is scanned perpendicular to the long axis of the cantilever beam (i. e., to scan along the x- or $-x$-direction in Fig. 23.19a) and the output of the two horizontal quadrants of the photodiode detector is measured. In this arrangement, as the sample moves under the tip, the friction force will cause the cantilever to twist. Therefore, the light intensity between the left and right (L and R in Fig. 23.19b, right) detectors will be different. The differential signal between the left and right detectors is denoted as FFM signal $[(L-R)/(L+R)]$. This signal can be related to the degree of twisting, hence to the magnitude of friction force. Again, because of a possible error in determining normal force due to the presence of an adhesive force at the tip–sample interface, the slope of the friction data (FFM signal versus normal load) needs to be taken for an accurate value of coefficient of friction.

While friction force contributes to the FFM signal, friction force may not be the only contributing factor in commercial FFM instruments. One can notice this fact by simply engaging the cantilever tip with the sample. Before engaging, the left and right detectors can be balanced by adjusting the position of the detectors so that the intensity difference between these two detectors is zero (FFM signal is zero). Once the tip is engaged with the sample, this signal is no longer zero even if the sample is not moving in the xy-plane with no friction force applied. This would be a detrimental effect. It has to be understood and eliminated from the data acquisition before any quantitative measurement of friction force becomes possible.

One of the fundamental reasons for this observation is the following. The detectors may not have been properly aligned with respect to the laser beam. To be precise, the vertical axis of the detector assembly (the line joining T–B in Fig. 23.22) is not in the plane defined by the incident laser beam and the beam reflected from an untwisted cantilever (we call this the *beam plane*). When the cantilever vertical deflection changes due to a change of applied normal force (without having the sample scanned in the xy-plane), the laser beam will be reflected up and down and form a projected trajectory on the detector. (Note that this trajectory is in the defined beam plane). If this trajectory is not coincident with the vertical axis of the detector, the laser beam will not evenly bisect the left and right quadrants of the detectors, even under the condition of no torsional motion of the cantilever, see Fig. 23.22. Thus, when the laser beam is reflected up and down due to a change of the normal force, the intensity difference between the left and right detectors will also change. In other words, the FFM signal will change as the normal force applied to the tip is changed, even if the tip is not experiencing any friction force. This (FFM) signal is unrelated to friction force or to the actual twisting of the cantilever. We will call this part of the FFM signal $\mathrm{FFM_F}$, and the part which is truly related to friction force $\mathrm{FFM_T}$.

The $\mathrm{FFM_F}$ signal can be eliminated. One way of doing this is as follows. First the sample is scanned in both x- and $-x$-directions and the FFM signal for scans in each direction is recorded. Since friction force reverses its direction when the scanning direction is reversed from x- to $-x$-direction, the $\mathrm{FFM_T}$ signal will have opposite signs as the scanning direction of the sample is reversed ($\mathrm{FFM_T}(x) = -\mathrm{FFM_T}(-x)$). Hence, the $\mathrm{FFM_T}$ signal will be canceled out if we take the sum of the FFM signals for the two scans. The average value of the two scans will be related to $\mathrm{FFM_F}$ due to the misalignment,

$$\mathrm{FFM}(x) + \mathrm{FFM}(-x) = 2\mathrm{FFM_F}\,. \tag{23.6}$$

This value can therefore be subtracted from the original FFM signals of each of these two scans to obtain the true FFM signal ($\mathrm{FFM_T}$). Or, alternately, by

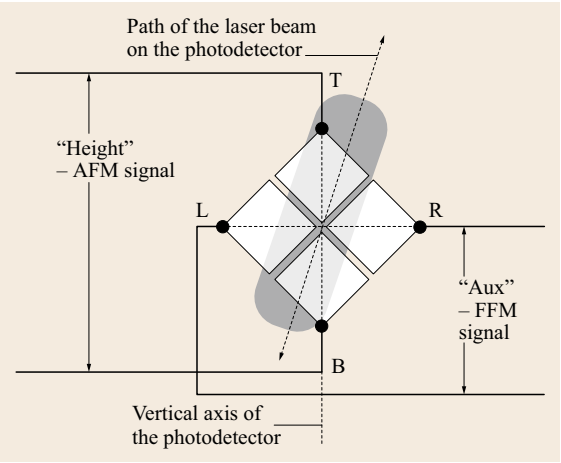


Fig. 23.22 The trajectory of the laser beam on the photodetectors as the cantilever is vertically deflected (with no torsional motion) for a misaligned photodetector with respect to the laser beam. For a change of normal force (vertical deflection of the cantilever), the laser beam is projected at a different position on the detector. Because of a misalignment, the projected trajectory of the laser beam on the detector is not parallel with the detector's vertical axis (the line T–B) (after [23.157])

taking the difference of the two FFM signals, one directly gets the $\mathrm{FFM_T}$ value

$$\begin{aligned}\mathrm{FFM}(x) - \mathrm{FFM}(-x) &= \mathrm{FFM_T}(x) - \mathrm{FFM_T}(-x) \\ &= 2\mathrm{FFM_T}(x)\,.\end{aligned} \tag{23.7}$$

Ruan and *Bhushan* [23.157] have shown that the error signal ($\mathrm{FFM_F}$) can be very large compared to friction signal $\mathrm{FFM_T}$, thus correction is required.

Now we compare the two methods. The method using the *height* mode with a parallel scan (method 1) is very simple to use. Technically, this method can provide 3-D friction profiles and the corresponding topographic profiles. However, there are some problems with this method. Under most circumstances, the piezo scanner displays a hysteresis when the traveling direction of the sample is reversed. Therefore, the measured surface topographic profiles will be shifted relative to each other along the y-axis for the two opposite (y and $-y$) scans. This would make it difficult to measure the local height difference of the piezo tube for the two scans. However, the average height difference between the two scans and hence the average friction can still be measured. The measurement of average friction can serve as an internal means of friction force calibration. Method 2 is the more desirable approach. The subtraction of the $\mathrm{FFM_F}$ signal from FFM for the two scans does not introduce error to local friction force data. An ideal approach in using this method would be to add the average value of the two profiles in order to get the error component ($\mathrm{FFM_F}$) and then subtract this component from either profile to get true friction profiles in either direction. By making measurements at various loads, we can obtain the average value of the coefficient of friction which then can be used to convert the friction profile to the coefficient of friction profile. Thus, any directionality and local variations in friction can be easily measured. In this method, since topography data are not affected by friction, accurate topography data can be measured simultaneously with friction data and a better localized relationship between the two can be established.

23.2.5 Normal Force and Friction Force Calibrations of Cantilever Beams

On the basis of [23.157], we now discuss normal force and friction force calibrations. In order to calculate the absolute value of normal and friction forces in Newtons using the measured AFM and $\mathrm{FFM_T}$ voltage signals, it is necessary to first have an accurate value of the spring constant of the cantilever (k_c). The spring constant can be calculated using the geometry and the physical properties of the cantilever material [23.7, 113, 158]. However, the properties of the PECVD Si_3N_4 (used in fabricating cantilevers) could be different from those of the bulk material. For example, by using an ultrasonic measurement, we found the Young's modulus of the cantilever beam to be about 238 ± 18 GPa, which is less than that of bulk Si_3N_4 (310 GPa). Furthermore, the thickness of the beam is nonuniform and difficult to measure precisely. Since the stiffness of a beam is proportional to the cube of thickness, minor errors in precise measurements of thickness can introduce substantial stiffness errors. Thus, one should experimentally measure the spring constant of the cantilever. *Cleveland* et al. [23.167] measured the normal spring constant by measuring resonant frequencies of the beams.

For normal spring constant measurement, *Ruan* and *Bhushan* [23.157] used a stainless steel spring sheet of known stiffness (width = 1.35 mm, thickness = 15 μm, free hanging length = 5.2 mm). One end of the spring was attached to the sample holder and the other end was made to contact with the cantilever tip during the measurement, see Fig. 23.23. They measured the piezo traveling distance for a given cantilever deflection. For a rigid sample (such as diamond), the piezo traveling distance Z_t (measured from the point where the tip touches the sample) should equal the cantilever deflection. To keep the cantilever deflection at the same level using a flexible spring sheet, the new piezo traveling

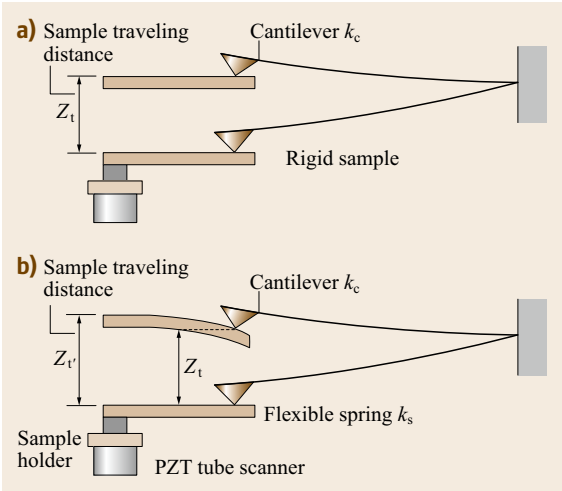


Fig. 23.23a,b Illustration showing the deflection of a cantilever as it is pushed by **(a)** a rigid sample or by **(b)** a flexible spring sheet (after [23.157])

distance $Z_{t'}$ would be different from Z_t. The difference between $Z_{t'}$ and Z_t corresponds to the deflection of the spring sheet. If the spring constant of the spring sheet is k_c, the spring constant of the cantilever k_c can be calculated by

$$(Z_{t'} - Z_t)\, k_s = Z_t k_c$$

or

$$k_c = \frac{k_S\,(Z_{t'} - Z_t)}{Z_t}\,. \tag{23.8}$$

The spring constant of the spring sheet (k_s) used in this study is calculated to be 1.54 N/m. For the wide-legged cantilever used in our study (length = 115 μm, base width = 122 μm, leg width = 21 μm, and thickness = 0.6 μm), k_c was measured to be 0.40 N/m instead of 0.58 N/m reported by its manufacturer – Digital Instruments Inc. To relate photodiode detector output to the cantilever deflection in nm, they used the same rigid sample to push against the AFM tip. Since for a rigid sample the cantilever vertical deflection equals the sample traveling distance measured from the point where the tip touches the sample, the photodiode output as the tip is pushed by the sample can be converted directly to cantilever deflection. For these measurements, they found the conversion factor to be 20 nm/V.

The normal force applied to the tip can be calculated by multiplying the cantilever vertical deflection by the cantilever spring constant for samples which have very small adhesive force with the tip. If the adhesive force between the sample and the tip is large, it should be included in the normal force calculation. This is particularly important in atomic-scale force measurement because in this region, the typical normal force that is measured is in the range of a few hundreds of nN to a few mN. The adhesive force could be comparable to the applied force.

The conversion of friction signal (from FFM_T to friction force) is not as straightforward. For example, one can calculate the degree of twisting for a given friction force using the geometry and the physical properties of the cantilever [23.62, 166]. One would need information on the detectors such as the quantum efficiency of the detector, the laser power, the instrument's gain, etc. in order to be able to convert the signal into the degree of twisting. Generally speaking, this procedure cannot be accomplished without having some detailed information about the instrument. This information is not usually provided by the manufacturers. Even if this information is readily available, error may still occur in using this approach because there will always be variations as a result of the instrumental set up. For example, it has been noticed that the measured FFM_T signal could be different for the same sample when different AFM microscopes of the same kind are used. The essence is that, one cannot calibrate the instrument experimentally using this calculation. *O'Shea* et al. [23.166] did perform a calibration procedure in which the torsional signal was measured as the sample is displaced a known distance laterally while ensuring that the tip does not slide over the surface. However, it is difficult to verify if the tip sliding does not occur.

Apparently, a new method of calibration is required. There is a more direct and simpler way of doing this. The first method described (method 1) to measure friction can directly provide an absolute value of coefficient of friction. It can therefore be used just as an internal means of calibration for the data obtained using method 2. Or, if for method 1 one uses a polished sample which introduces the least error in friction measurements, the method can then be used to obtain a calibration for friction force for method 2. This calibration can in turn be used for measurement of all samples using method 2. In method 1, the length of the cantilever required can be measured using an optical microscope; the length of the tip can be measured using a scanning electron microscope. The relative angle between the cantilever and the horizontal sample surface can be measured directly. Thus, the coefficient of friction can be measured with few unknown parameters. The friction force can then be calculated by multiplying the coefficient of friction by the normal load. The FFM_T signal obtained using method 2 can then be converted into friction force. For their instrument, they found the conversion to be 8.6 nN/V.

23.3 AFM Instrumentation and Analyses

The performance of AFMs and the quality of AFM images greatly depend on the instrument available and the probes (cantilever and tips) in use. This section describes the mechanics of cantilevers, instrumentation and analysis of force detection systems for cantilever deflections, and scanning and control systems.

23.3.1 The Mechanics of Cantilevers

Stiffness and Resonances of Lumped Mass Systems

Any one of the building blocks of an AFM, be it the body of the microscope itself or the force-measuring cantilevers, are mechanical resonators. These resonances can be excited either by the surrounding or by the rapid movement of the tip or the sample. To avoid problems due to building- or air-induced oscillations, it is of paramount importance to optimize the design of the AFMs for high resonant frequencies. This usually means to decrease the size of the microscope [23.168]. By using cube-like or sphere-like structures for the microscope, one can considerably increase the lowest eigen-frequency. The fundamental natural frequency, ω_0, of any spring is given by

$$\omega_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m_{\text{eff}}}}\,, \tag{23.9}$$

where k is the spring constant (stiffness) in the normal direction and m_{eff} is the effective mass. The spring constant k of a cantilever beam with uniform cross section (Fig. 23.24) is given by [23.169]

$$k = \frac{3EI}{L^3}\,, \tag{23.10}$$

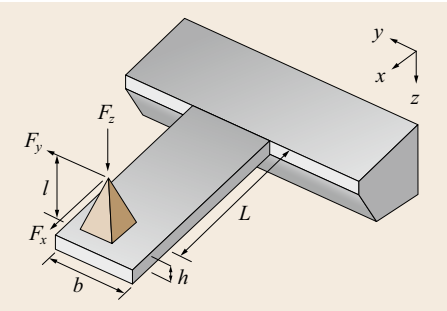


Fig. 23.24 A typical AFM cantilever with length L, width b, and height h. The height of the tip is l. The material is characterized by Young's modulus E, the shear modulus G, and a mass density ρ. Normal (F_z), axial (F_x), and lateral (F_y) forces exist at the end of the tip

where E is the Young's modulus of the material, L is the length of the beam, and I is the moment of inertia of the cross section. For a rectangular cross section with width b (perpendicular to the deflection) and height h one obtains an expression for I

$$I = \frac{bh^3}{12}\,. \tag{23.11}$$

Combining (23.9)–(23.11) we get an expression for ω_0

$$\omega_0 = \sqrt{\frac{Ebh^3}{4L^3 m_{\text{eff}}}}\,. \tag{23.12}$$

The effective mass can be calculated using Raleigh's method. The general formula using Raleigh's method for the kinetic energy T of a bar is

$$T = \frac{1}{2}\int_0^L \frac{m}{L}\left(\frac{\partial z(x)}{\partial t}\right)^2 \mathrm{d}x\,. \tag{23.13}$$

For the case of a uniform beam with a constant cross section and length L one obtains for the deflection

$$z(x) = z_{\max}\left[1 - \left(\frac{3x}{2L}\right) + \left(\frac{x^3}{2L^3}\right)\right]\,.$$

Inserting $z_{\max}$ into (23.13) and solving the integral gives

$$\begin{aligned} T &= \frac{1}{2}\int_0^L \frac{m}{L}\left[\frac{\partial z_{\max}(x)}{\partial t}\left(1 - \frac{3x}{2L}\right) + \left(\frac{x^3}{L^3}\right)\right]^2 \mathrm{d}x \\ &= \frac{1}{2} m_{\text{eff}}\,(z_{\max} t)^2 \end{aligned}$$

which gives

$$m_{\text{eff}} = \frac{9}{20}m\,. \tag{23.14}$$

Substituting (23.14) into (23.12) and noting that $m = \rho Lbh$, where ρ is the mass density, one obtains the following expression

$$\omega_0 = \left(\frac{\sqrt{5}}{3}\sqrt{\frac{E}{\rho}}\right)\frac{h}{L^2}\,. \tag{23.15}$$

It is evident from (23.15), that one way to increase the natural frequency is to choose a material with a high ratio E/ρ; see Table 23.2 for typical values of $\sqrt{E/\rho}$

of various commonly used materials. Another way to increase the lowest eigen-frequency is also evident in (23.15). By optimizing the ratio h/L^2, one can increase the resonant frequency. However, it does not help to make the length of the structure smaller than the width or height. Their roles will just be interchanged. Hence, the optimum structure is a cube. This leads to the design rule, that long, thin structures like sheet metal should be avoided. For a given resonant frequency, the quality factor Q should be as low as possible. This means that an inelastic medium such as rubber should be in contact with the structure to convert kinetic energy into heat.

Stiffness and Resonances of Cantilevers

Cantilevers are mechanical devices specially shaped to measure tiny forces. The analysis given in the previous section is applicable. However, to understand better the intricacies of force detection systems we will discuss the example of a cantilever beam with uniform cross section, Fig. 23.24. The bending of a beam due to a normal load on the beam is governed by the Euler equation [23.169]

$$M = EI(x)\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\,, \tag{23.16}$$

where M is the bending moment acting on the beam cross section. $I(x)$, the moment of inertia of the cross section with respect to the neutral axis, is defined by

$$I(x) = \int_z \int_y z^2 \mathrm{d}y\mathrm{d}z\,. \tag{23.17}$$

For a normal force F_z acting at the tip,

$$M(x) = (L-x)F_z \tag{23.18}$$

since the moment must vanish at the endpoint of the cantilever. Integrating (23.16) for a normal force F_z acting at the tip and observing that EI is a constant for beams with a uniform cross section, one gets

$$z(x) = \frac{L^3}{6EI}\left(\frac{x}{L}\right)^2\left(3-\frac{x}{L}\right)F_z\,. \tag{23.19}$$

The slope of the beam is

$$z'(x) = \frac{Lx}{2EI}\left(2-\frac{x}{L}\right)F_z\,. \tag{23.20}$$

From (23.19) and (23.20), at the end of the cantilever, i. e., for $x = L$, for a rectangular beam, and by using an expression for I in (23.11), one gets,

$$z(L) = \frac{4}{Eb}\left(\frac{L}{h}\right)^3 F_z\,, \tag{23.21}$$

$$z'(L) = \frac{3}{2}\left(\frac{z}{L}\right)\,. \tag{23.22}$$

Now the stiffness in the normal (z) direction, k_z, is

$$k_z = \frac{F_z}{z(L)} = \frac{Eb}{4}\left(\frac{h}{L}\right)^3 \tag{23.23}$$

and a change in angular orientation of the end of the cantilever beam is

$$\Delta\alpha = \frac{3}{2}\frac{z}{L} = \frac{6}{Ebh}\left(\frac{L}{h}\right)^2 F_z\,. \tag{23.24}$$

Now we ask what will, to first order, happen if we apply a lateral force F_y to the end of the tip (Fig. 23.24. The cantilever will bend sideways and it will twist. The stiffness in the lateral (y) direction, k_y, can be calculated with (23.23) by exchanging b and h

$$k_y = \frac{Eh}{4}\left(\frac{b}{L}\right)^3\,. \tag{23.25}$$

Therefore, the bending stiffness in the lateral direction is larger than the stiffness for bending in the normal direction by $(b/h)^2$. The twisting or torsion on the other side is more complicated to handle. For a wide, thin cantilever ($b \gg h$) we obtain torsional stiffness along the y-axis, $k_{y\mathrm{T}}$

$$k_{y\mathrm{T}} = \frac{Gbh^3}{3L\ell^2}\,, \tag{23.26}$$

where G is the modulus of rigidity [$= E/2(1+\nu)$, where ν is Poisson's ratio]. The ratio of the torsional stiffness to the lateral bending stiffness is

$$\frac{k_{y\mathrm{T}}}{k_y} = \frac{1}{2}\left(\frac{\ell b}{hL}\right)^2\,, \tag{23.27}$$

where we assume $\nu = 0.333$. We see that thin, wide cantilevers with long tips favor torsion while cantilevers with square cross sections and short tips favor bending. Finally, we calculate the ratio between the torsional stiffness and the normal bending stiffness,

$$\frac{k_{y\mathrm{T}}}{k_z} = 2\left(\frac{L}{\ell}\right)^2\,. \tag{23.28}$$

Equations (23.26) to (23.28) hold in the case where the cantilever tip is exactly in the middle axis of the cantilever. Triangular cantilevers and cantilevers with tips not on the middle axis can be dealt with by finite element methods.

The third possible deflection mode is the one from the force on the end of the tip along the cantilever axis, F_x (Fig. 23.24). The bending moment at the free end of

the cantilever is equal to $F_x\ell$. This leads to the following modification of (23.18) for the case of forces F_z and F_x

$$M(x) = (L-x)F_z + F_x\ell \,. \tag{23.29}$$

Integration of (23.16) now leads to

$$z(x) = \frac{1}{2EI}\left[Lx^2\left(1-\frac{x}{3L}\right)F_z + \ell x^2 F_x\right] \tag{23.30}$$

and

$$z'(x) = \frac{1}{EI}\left[\frac{Lx}{2}\left(2-\frac{x}{L}\right)F_z + \ell x F_x\right] \,. \tag{23.31}$$

Evaluating (23.30) and (23.31) at the end of the cantilever, we get the deflection and the tilt

$$\begin{aligned} z(L) &= \frac{L^2}{EI}\left(\frac{L}{3}F_z - \frac{\ell}{2}F_x\right) \,, \\ z'(x) &= \frac{L}{EI}\left(\frac{L}{2}F_z + \ell F_x\right) \,. \end{aligned} \tag{23.32}$$

From these equations, one gets

$$\begin{aligned} F_z &= \frac{12EI}{L^3}\left(z(L) - \frac{Lz'(L)}{2}\right) \,, \\ F_x &= \frac{2EI}{\ell L^2}\left(2Lz'(L) - 3z(L)\right) \,. \end{aligned} \tag{23.33}$$

A second class of interesting properties of cantilevers is their resonance behavior. For cantilever beams one can calculate the resonant frequencies [23.169, 170]

$$\omega_n^{\text{free}} = \frac{\lambda_n^2}{2\sqrt{3}}\frac{h}{L^2}\sqrt{\frac{E}{\rho}} \tag{23.34}$$

with $\lambda_0 = (0.596864\ldots)\pi$, $\lambda_1 = (1.494175\ldots)\pi$, $\lambda_n \to (n+1/2)\pi$. The subscript n represents the order of the frequency, e.g., fundamental, second mode, and the n-th mode.

A similar equation as (23.34) holds for cantilevers in rigid contact with the surface. Since there is an additional restriction on the movement of the cantilever, namely the location of its end point, the resonant frequency increases. Only the λ_n's terms change to [23.170]

$$\begin{aligned} \lambda_0' &= (1.2498763\ldots)\pi \,, \\ \lambda_1' &= (2.2499997\ldots)\pi \,, \\ \lambda_n' &\to \left(n+\frac{1}{4}\right)\pi \,. \end{aligned} \tag{23.35}$$

The ratio of the fundamental resonant frequency in contact to the fundamental resonant frequency not in contact is 4.3851.

For the torsional mode we can calculate the resonant frequencies as

$$\omega_0^{\text{tors}} = 2\pi\frac{h}{Lb}\sqrt{\frac{G}{\rho}} \,. \tag{23.36}$$

For cantilevers in rigid contact with the surface, we obtain the expression for the fundamental resonant frequency [23.170]

$$\omega_0^{\text{tors,contact}} = \frac{\omega_0^{\text{tors}}}{\sqrt{1+3\left(\frac{2L}{b}\right)^2}} \,. \tag{23.37}$$

The amplitude of the thermally induced vibration can be calculated from the resonant frequency using

$$\Delta z_{\text{therm}} = \sqrt{\frac{k_B T}{k}} \,, \tag{23.38}$$

where k_B is Boltzmann's constant and T is the absolute temperature. Since AFM cantilevers are resonant structures, sometimes with rather high Q, the thermal noise is not evenly distributed as (23.38) suggests. The spectral noise density below the peak of the response curve is [23.170]

$$z_0 = \sqrt{\frac{4k_B T}{k\omega_0 Q}} \quad \left\{\text{in m}/\sqrt{\text{Hz}}\right\} \,, \tag{23.39}$$

where Q is the quality factor of the cantilever, described earlier.

23.3.2 Instrumentation and Analyses of Detection Systems for Cantilever Deflections

A summary of selected detection systems was provided in Fig. 23.8. Here we discuss in detail the pros and cons of various systems.

Optical Interferometer Detection Systems

Interferometers. Soon after the first papers on the AFM [23.2], which used a tunneling sensor, an instrument based on an interferometer was presented [23.171]. The sensitivity of the interferometer depends on the wavelength of the light employed in the apparatus. Figure 23.25 shows the principle of such an interferometric design. The light incident from the left is

focused by a lens on the cantilever. The reflected light is collimated by the same lens and interferes with the light reflected at the flat. To separate the reflected light from the incident light a $\lambda/4$ plate converts the linear polarized incident light to circular polarization. The reflected light is made linearly polarized again by the $\lambda/4$ plate, but with a polarization orthogonal to that of the incident light. The polarizing beam splitter then deflects the reflected light to the photodiode.

Homodyne Interferometer. To improve the signal-to-noise ratio of the interferometer the cantilever is driven by a piezo near its resonant frequency. The amplitude Δz of the cantilever as a function of driving frequency Ω is

$$\Delta z(\Omega) = \Delta z_0 \frac{\Omega_0^2}{\sqrt{\left(\Omega^2 - \Omega_0^2\right)^2 + \frac{\Omega^2 \Omega_0^2}{Q^2}}} , \qquad (23.40)$$

where Δz_0 is the constant drive amplitude and Ω_0 the resonant frequency of the cantilever. The resonant frequency of the cantilever is given by the effective potential

$$\Omega_0 = \sqrt{\left(k + \frac{\partial^2 U}{\partial z^2}\right) \frac{1}{m_{\text{eff}}}} \qquad (23.41)$$

where U is the interaction potential between the tip and the sample. Equation (23.41) shows that an attractive potential decreases Ω_0. The change in Ω_0 in turn results in a change of the Δz (see (23.40)). The movement of the cantilever changes the path difference in the interferometer. The light reflected from the cantilever with the amplitude $A_{\ell,0}$ and the reference light with the amplitude $A_{r,0}$ interfere on the detector. The detected

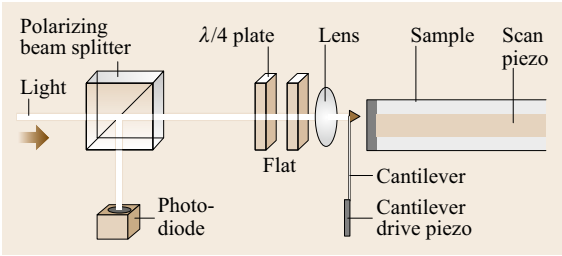


Fig. 23.25 Principle of an interferometric AFM. The light of the laser light source is polarized by the polarizing beam splitter and focused on the back of the cantilever. The light passes twice through a quarter wave plate and is hence orthogonally polarized to the incident light. The second arm of the interferometer is formed by the flat. The interference pattern is modulated by the oscillating cantilever

intensity

$$I(t) = \{A_\ell(t) + A_r(t)\}^2$$

consists of two constant terms and a fluctuating term

$$2A_\ell(t)A_r(t) = A_{\ell,0}A_{r,0} \sin\left[\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda} \sin(\Omega t)\right] \sin(\omega t) . \qquad (23.42)$$

Here ω is the frequency of the light, λ is the wavelength of the light, δ is the path difference in the interferometer, and Δz is the instantaneous amplitude of the cantilever, given according to (23.40) and (23.41) as a function of Ω, k, and U. The time average of (23.42) then becomes

$$\begin{aligned}\langle 2A_\ell(t)A_r(t)\rangle_{\text{T}} &\propto \cos\left[\frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda} \sin(\Omega t)\right] \\ &\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \sin\left[\frac{4\pi\Delta z}{\lambda} \sin(\Omega t)\right] \\ &\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \frac{4\pi\Delta z}{\lambda} \sin(\Omega t) .\end{aligned} \qquad (23.43)$$

Here all small quantities have been omitted and functions with small arguments have been linearized. The amplitude of Δz can be recovered with a lock-in technique. However, (23.43) shows that the measured amplitude is also a function of the path difference δ in the interferometer. Hence this path difference δ must be very stable. The best sensitivity is obtained when $\sin(4\delta/\lambda) \approx 0$.

Heterodyne Interferometer. This influence is not present in the heterodyne detection scheme shown in Fig. 23.26. Light incident from the left with a frequency ω is split into a reference path (upper path in Fig. 23.26) and a measurement path. Light in the measurement path is shifted in frequency to $\omega_1 = \omega + \Delta\omega$ and focused on the cantilever. The cantilever oscillates at the frequency Ω, as in the homodyne detection scheme. The reflected light $A_\ell(t)$ is collimated by the same lens and interferes on the photodiode with the reference light $A_r(t)$. The fluctuating term of the intensity is given by

$$2A_\ell(t)A_r(t) = A_{\ell,0}A_{r,0} \sin\left[(\omega + \Delta\omega)t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda} \sin(\Omega t)\right] \sin(\omega t) , \qquad (23.44)$$

where the variables are defined as in (23.42). Setting the path difference $\sin(4\pi\delta/\lambda) \approx 0$ and taking the time average, omitting small quantities and linearizing functions with small arguments we get

$$
\begin{aligned}
&\langle 2A_\ell(t)A_r(t)\rangle_\mathrm{T} \\
&\propto \cos\left[\Delta\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\cos\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\quad - \sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \sin\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\approx \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\left[1 - \frac{8\pi^2\Delta z^2}{\lambda^2}\sin(\Omega t)\right] \\
&\quad - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) - \frac{8\pi^2\Delta z^2}{\lambda^2} \\
&\quad \times \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&\quad - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) - \frac{4\pi^2\Delta z^2}{\lambda^2} \\
&\quad \times \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) \\
&\quad + \frac{4\pi^2\Delta z^2}{\lambda^2}\cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\cos(2\Omega t) \\
&\quad - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \\
&\quad + \frac{2\pi^2\Delta z^2}{\lambda^2}\left(\cos\left[(\Delta\omega + 2\Omega)\,t + \frac{4\pi\delta}{\lambda}\right]\right. \\
&\quad \left. + \cos\left[(\Delta\omega - 2\Omega)\,t + \frac{4\pi\delta}{\lambda}\right]\right) \\
&\quad + \frac{2\pi\Delta z}{\lambda}\left(\cos\left[(\Delta\omega + \Omega)\,t + \frac{4\pi\delta}{\lambda}\right]\right. \\
&\quad \left. + \cos\left[(\Delta\omega - \Omega)\,t + \frac{4\pi\delta}{\lambda}\right]\right) .
\end{aligned}
\tag{23.45}
$$

Multiplying electronically the components oscillating at $\Delta\omega$ and $\Delta\omega + \Omega$ and rejecting any product except the one oscillating at Ω we obtain

$$
\begin{aligned}
A &= \frac{2\Delta z}{\lambda}\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \\
&\quad \times \cos\left[(\Delta\omega + 2\Omega)\,t + \frac{4\pi\delta}{\lambda}\right] \\
&\quad \times \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) \\
&= \frac{\Delta z}{\lambda}\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \\
&\quad \times \left(\cos\left[(2\Delta\omega + \Omega)\,t + \frac{8\pi\delta}{\lambda}\right] + \cos(\Omega t)\right) \\
&\approx \frac{\pi\Delta z}{\lambda}\cos(\Omega t) \ .
\end{aligned}
\tag{23.46}
$$

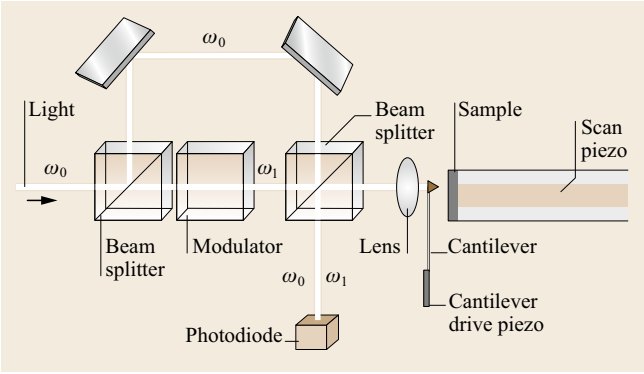


Fig. 23.26 Principle of a heterodyne interferometric AFM. Light with the frequency ω_0 is split into a reference path (upper path) and a measurement path. The light in the measurement path is frequency shifted to ω_1 by an acousto-optical modulator (or an electro-optical modulator). The light reflected from the oscillating cantilever interferes with the reference beam on the detector

Unlike in the homodyne detection scheme the recovered signal is independent from the path difference δ of the interferometer. Furthermore, a lock-in amplifier with the reference set $\sin(\Delta\omega t)$ can measure the path difference δ independent of the cantilever oscillation. If necessary, a feedback circuit can keep $\delta = 0$.

Fiber-optical Interferometer. The fiber-optical interferometer [23.150] is one of the simplest interferometers to build and use. Its principle is sketched in Fig. 23.27. The light of a laser is fed into an optical fiber. Laser diodes with integrated fiber pigtails are convenient light sources. The light is split in a fiber-optic beam splitter into two fibers. One fiber is terminated by index matching oil to avoid any reflections back into the fiber. The end of the other fiber is brought close to the

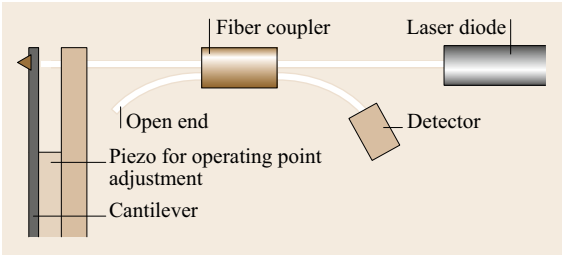


Fig. 23.27 A typical setup for a fiber-optic interferometer readout

cantilever in the AFM. The emerging light is partially reflected back into the fiber by the cantilever. Most of the light, however, is lost. This is not a big problem since only 4% of the light is reflected at the end of the fiber, at the glass-air interface. The two reflected light waves interfere with each other. The product is guided back into the fiber coupler and again split into two parts. One half is analyzed by the photodiode. The other half is fed back into the laser. Communications-grade laser diodes are sufficiently resistant against feedback to be operated in this environment. They have, however, a bad coherence length, which in this case does not matter, since the optical path difference is in any case no larger than 5 μm. Again the end of the fiber has to be positioned on a piezo drive to set the distance between the fiber and the cantilever to $\lambda(n+1/4)$.

Nomarski Interferometer. Another solution to minimize the optical path difference is to use the Nomarski interferometer [23.151]. Figure 23.28 shows a schematic of the microscope. The light of a laser is focused on the cantilever by a lens. A birefringent crystal (for instance calcite) between the cantilever and the lens with its optical axis 45° off the polarization direction of the light splits the light beam into two paths, offset by a distance given by the length of the crystal. Birefringent crystals have varying indexes of refraction. In calcite, one crystal axis has a lower index than the other two. This means, that certain light rays will propagate at a different speed through the crystal than the others. By choosing a correct polarization, one can select the ordinary ray, the extraordinary ray or one can get any distribution of the intensity amongst those two rays. A detailed description of birefringence can be found in textbooks (e.g., [23.172]). A calcite crystal deflects the extraordinary ray at an angle of 6° within the crystal. By choosing a suitable length of the calcite crystal, any separation can be set.

The focus of one light ray is positioned near the free end of the cantilever while the other is placed close to the clamped end. Both arms of the interferometer pass

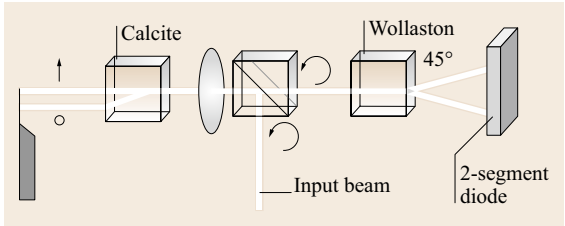


Fig. 23.28 Principle of the Nomarski AFM. The circular polarized input beam is deflected to the left by a nonpolarizing beam splitter. The light is focused onto a cantilever. The calcite crystal between the lens and the cantilever splits the circular polarized light into two spatially separated beams with orthogonal polarizations. The two light beams reflected from the lever are superimposed by the calcite crystal and collected by the lens. The resulting beam is again circular polarized. A Wollaston prism produces two interfering beams with $\pi/2$ phase shift between them. The minimal path difference accounts for the excellent stability of this microscope

through the same space, except for the distance between the calcite crystal and the lever. The closer the calcite crystal is placed to the lever, the less influence disturbances like air currents have.

Sensitivity. *Sarid* [23.137] has given values for the sensitivity of the different interferometric detection systems. Table 23.5 presents a summary of his results.

Optical Lever

The most common cantilever deflection detection system is the optical lever [23.62, 132]. This method, depicted in Fig. 23.29, employs the same technique as light beam deflection galvanometers. A fairly well collimated light beam is reflected off a mirror and projected to a receiving target. Any change in the angular position of the mirror will change the position at which the light ray hits the target. Galvanometers use optical path lengths of several meters and scales projected to the target wall as a read-out aid.

For the AFM using the optical lever method a photodiode segmented into two (or four) closely spaced devices detects the orientation of the end of the cantilever. Initially, the light ray is set to hit the photodiodes in the middle of the two subdiodes. Any deflection of the cantilever will cause an imbalance of the number of photons reaching the two halves. Hence, the electrical currents in the photodiodes will be unbalanced too. The difference signal is further amplified and is the input signal to the feedback loop. Unlike the interferometric AFMs, where often a modulation technique is necessary to get a sufficient signal-to-noise ratio, most AFMs employing the optical lever

Table 23.5 Noise in interferometers. F is the finesse of the cavity in the homodyne interferometer, P_i the incident power, P_d is the power on the detector, η is the sensitivity of the photodetector, and RIN is the relative intensity noise of the laser. P_R and P_S are the power in the reference and sample beam in the heterodyne interferometer. P is the power in the Nomarski interferometer, $\delta\theta$ is the phase difference between the reference and the probe beam in the Nomarski interferometer. B is the bandwidth, e is the electron charge, λ is the wavelength of the laser, k the cantilever stiffness, ω_0 is the resonant frequency of the cantilever, Q is the quality factor of the cantilever, T is the temperature, and δi is the variation of current i

	Homodyne interferometer, fiber-optic interferometer	Heterodyne interferometer	Nomarski interferometer
Laser noise $\langle \delta i^2 \rangle_\mathrm{L}$	$\frac{1}{4}\eta^2 F^2 P_i^2 \mathrm{RIN}$	$\eta^2 \left(P_\mathrm{R}^2 + P_\mathrm{S}^2\right) \mathrm{RIN}$	$\frac{1}{16}\eta^2 P^2 \delta\theta$
Thermal noise $\langle \delta i^2 \rangle_\mathrm{T}$	$\frac{16\pi^2}{\lambda^2}\eta^2 F^2 P_i^2 \frac{4k_\mathrm{B}TBQ}{\omega_0 k}$	$\frac{4\pi^2}{\lambda^2}\eta^2 P_d^2 \frac{4k_\mathrm{B}TBQ}{\omega_0 k}$	$\frac{\pi^2}{\lambda^2}\eta^2 P^2 \frac{4k_\mathrm{B}TBQ}{\omega_0 k}$
Shot Noise $\langle \delta i^2 \rangle_\mathrm{S}$	$4e\eta P_\mathrm{d} B$	$2e\eta \left(P_\mathrm{R} + P_\mathrm{S}\right) B$	$\frac{1}{2}e\eta P B$

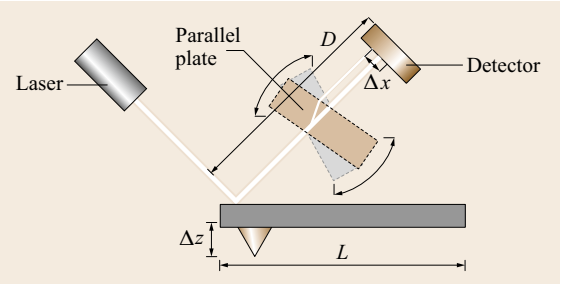


Fig. 23.29 The setup of the optical lever detection microscope

method are operated in a static mode. AFMs based on the optical lever method are universally used. It is the simplest method to construct an optical readout and it can be confined in volumes smaller than 5 cm on the side.

The optical lever detection system is a simple yet elegant way to detect normal and lateral force signals simultaneously [23.6, 7, 62, 132]. It has the additional advantage that it is a remote detection system.

Implementations. Light from a laser diode or from a super luminescent diode is focused on the end of the cantilever. The reflected light is directed onto a quadrant diode that measures the direction of the light beam. A Gaussian light beam far from its waist is characterized by an opening angle β. The deflection of the light beam by the cantilever surface tilted by an angle α is 2α. The intensity on the detector then shifts to the side by the product of 2α and the separation between the detector and the cantilever. The readout electronics calculates the difference of the photocurrents. The photocurrents in turn are proportional to the intensity incident on the diode.

The output signal is hence proportional to the change in intensity on the segments

$$I_\mathrm{sig} \propto 4\frac{\alpha}{\beta} I_\mathrm{tot} \,. \tag{23.47}$$

For the sake of simplicity, we assume that the light beam is of uniform intensity with its cross section increasing proportional to the distance between the cantilever and the quadrant detector. The movement of the center of the light beam is then given by

$$\Delta x_\mathrm{Det} = \Delta z \frac{D}{L} \,. \tag{23.48}$$

The photocurrent generated in a photodiode is proportional to the number of incoming photons hitting it. If the light beam contains a total number of N_0 photons then the change in difference current becomes

$$\Delta \left(I_\mathrm{R} - I_\mathrm{L}\right) = \Delta I = \mathrm{const}\, \Delta z D N_0 \,. \tag{23.49}$$

Combining (23.48) and (23.49), one obtains that the difference current ΔI is independent of the separation of the quadrant detector and the cantilever. This relation is true, if the light spot is smaller than the quadrant detector. If it is greater, the difference current ΔI becomes smaller with increasing distance. In reality the light beam has a Gaussian intensity profile. For small movements Δx (compared to the diameter of the light spot at the quadrant detector), (23.49) still holds. Larger movements Δx, however, will introduce a nonlinear response. If the AFM is operated in a constant force mode, only small movements Δx of the light spot will occur. The feedback loop will cancel out all other movements.

The scanning of a sample with an AFM can twist the microfabricated cantilevers because of lateral

forces [23.4, 6, 7] and affect the images [23.141]. When the tip is subjected to lateral forces, it will twist the cantilever and the light beam reflected from the end of the cantilever will be deflected perpendicular to the ordinary deflection direction. For many investigations this influence of lateral forces is unwanted. The design of the triangular cantilever stems from the desire to minimize the torsion effects. However, lateral forces open up a new dimension in force measurements. They allow, for instance, a distinction between two materials because of the different friction coefficient, or the determination of adhesion energies. To measure lateral forces the original optical lever AFM has to be modified. The only modification compared with Fig. 23.29 is the use of a quadrant detector photodiode instead of a two-segment photodiode and the necessary readout electronics, see Fig. 23.9a. The electronics calculates the following signals

$$\begin{aligned} U_{\text{NormalForce}} &= \alpha\left[\left(I_{\text{UpperLeft}} + I_{\text{UpperRight}}\right) - \right. \\ &\quad \left.\left(I_{\text{LowerLeft}} + I_{\text{LowerRight}}\right)\right] , \\ U_{\text{LateralForce}} &= \beta\left[\left(I_{\text{UpperLeft}} + I_{\text{LowerLeft}}\right) - \right. \\ &\quad \left.\left(I_{\text{UpperRight}} + I_{\text{LowerRight}}\right)\right] . \end{aligned} \tag{23.50}$$

The calculation of the lateral force as a function of the deflection angle does not have a simple solution for cross sections other than circles. An approximate formula for the angle of twist for rectangular beams is [23.173]

$$\theta = \frac{M_t L}{\beta G b^3 h} , \tag{23.51}$$

where $M_t = F_y \ell$ is the external twisting moment due to lateral force, F_y, and β a constant determined by the value of h/b. For the equation to hold, h has to be larger than b.

Inserting the values for a typical microfabricated cantilever with integrated tips

$$\begin{aligned} b &= 6\times10^{-7}\,\text{m} \\ h &= 10^{-5}\,\text{m} \\ L &= 10^{-4}\,\text{m} \\ \ell &= 3.3\times10^{-6}\,\text{m} \\ G &= 5\times10^{10}\,\text{Pa} \\ \beta &= 0.333 \end{aligned} \tag{23.52}$$

into (23.51) we obtain the relation

$$F_y = 1.1\times10^{-4} N\times\theta . \tag{23.53}$$

Typical lateral forces are of order 10^{-10} N.

Sensitivity. The sensitivity of this setup has been calculated in various papers [23.137, 170, 174]. Assuming a Gaussian beam the resulting output signal as a function of the deflection angle is dispersion-like. (23.47) shows that the sensitivity can be increased by increasing the intensity of the light beam I_{tot} or by decreasing the divergence of the laser beam. The upper bound of the intensity of the light I_{tot} is given by saturation effects on the photodiode. If we decrease the divergence of a laser beam we automatically increase the beam waist. If the beam waist becomes larger than the width of the cantilever we start to get diffraction. Diffraction sets a lower bound on the divergence angle. Hence, one can calculate the optimal beam waist w_{opt} and the optimal divergence angle β [23.170, 174]

$$\begin{aligned} w_{\text{opt}} &\approx 0.36b \\ \theta_{\text{opt}} &\approx 0.89\frac{\lambda}{b} . \end{aligned} \tag{23.54}$$

The optimal sensitivity of the optical lever then becomes

$$\varepsilon\left[\frac{\text{mW}}{\text{rad}}\right] = 1.8\frac{b}{\lambda} I_{\text{tot}}\,[\text{mW}] . \tag{23.55}$$

The angular sensitivity of the optical lever can be measured by introducing a parallel plate into the beam. A tilt of the parallel plate results in a displacement of the beam, mimicking an angular deflection.

An additional noise source can be considered. Of little importance is the quantum mechanical uncertainty of the position [23.170, 174], which is for typical cantilevers at room temperature

$$\Delta z = \sqrt{\frac{\hbar}{2m\omega_0}} = 0.05\,\text{fm} , \tag{23.56}$$

where $\hbar$ is the Planck constant ($= 6.626\times10^{-34}$ Js). At very low temperatures and for high-frequency cantilevers this could become the dominant noise source. A second noise source is the shot noise of the light. The shot noise is related to the particle number. We can calculate the number of photons incident on the detector

$$n = \frac{I\tau}{\hbar\omega} = \frac{I\lambda}{2\pi B\hbar c} = 1.8\times10^{9}\frac{I\,[\text{W}]}{B\,[\text{Hz}]} , \tag{23.57}$$

where I is the intensity of the light, τ the measurement time, $B = 1/\tau$ the bandwidth, and c the speed of light. The shot noise is proportional to the square root of the number of particles. Equating the shot noise signal with the signal resulting for the deflection of the cantilever

one obtains

$$\Delta z_{\text{shot}} = 68\frac{L}{w}\sqrt{\frac{B\,[\text{kHz}]}{I\,[\text{mW}]}}[\text{fm}]\,, \tag{23.58}$$

where w is the diameter of the focal spot. Typical AFM setups have a shot noise of 2 pm. The thermal noise can be calculated from the equipartition principle. The amplitude at the resonant frequency is

$$\Delta z_{\text{therm}} = 129\sqrt{\frac{B}{k\,[\text{N/m}]\omega_0 Q}}[\text{pm}]\,. \tag{23.59}$$

A typical value is 16 pm. Upon touching the surface, the cantilever increases its resonant frequency by a factor of 4.39. This results in a new thermal noise amplitude of 3.2 pm for the cantilever in contact with the sample.

Piezoresistive Detection

Implementations. An alternative detection system which is not as widely spread as the optical detection schemes are piezoresistive cantilevers [23.146, 147, 153]. These cantilevers are based on the fact that the resistivity of certain materials, in particular of Si, changes with the applied stress. Figure 23.30 shows a typical implementation of a piezoresistive cantilever. Four resistances are integrated on the chip, forming a Wheatstone bridge. Two of the resistors are in unstrained parts of the cantilever, the other two measure the bending at the point of maximal deflection. For instance, when an AC voltage is applied between terminals a and c one can measure the detuning of the bridge between terminals b and d. With such a connection the output signal varies only due to bending, but not due to changing the ambient temperature and thus the coefficient of piezoresistance.

Sensitivity. The resistance change is [23.147]

$$\frac{\Delta R}{R_0} = \Pi\delta\,, \tag{23.60}$$

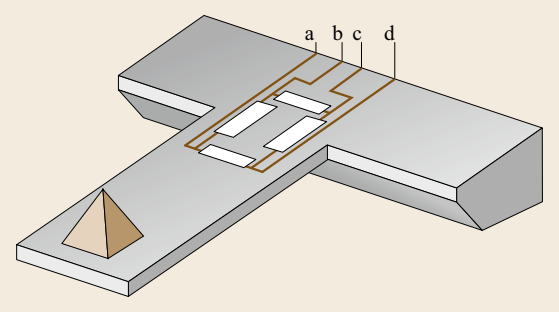


Fig. 23.30 A typical setup for a piezoresistive readout

where Π is the tensor element of the piezoresistive coefficients, δ the mechanical stress tensor element, and R_0 the equilibrium resistance. For a single resistor they separate the mechanical stress and the tensor element in longitudinal and transversal components

$$\frac{\Delta R}{R_0} = \Pi_{\text{t}}\delta_{\text{t}} + \Pi_{\text{l}}\delta_{\text{l}}\,. \tag{23.61}$$

The maximum value of the stress components are $\Pi_{\text{t}} = -64.0\times10^{-11}\,\text{m}^2/\text{N}$ and $\Pi_{\text{l}} = -71.4\times10^{-11}\,\text{m}^2/\text{N}$ for a resistor oriented along the (110)-direction in silicon [23.147]. In the resistor arrangement of Fig. 23.30, two of the resistors are subject to the longitudinal piezoresistive effect and two of them are subject to the transverse piezoresistive effect. The sensitivity of that setup is about four times that of a single resistor, with the advantage that temperature effects cancel to first order. The resistance change is then calculated as

$$\frac{\Delta R}{R_0} = \Pi\frac{3Eh}{2L^2}\Delta z = \Pi\frac{6L}{bh^2}F_z\,, \tag{23.62}$$

where $\Pi = 67.7\times10^{-11}\,\text{m}^2/\text{N}$ is the averaged piezoresistive coefficient. Plugging in typical values for the dimensions (Fig. 23.24) ($L = 100\,\mu\text{m}$, $b = 10\,\mu\text{m}$, $h = 1\,\mu\text{m}$) one obtains

$$\frac{\Delta R}{R_0} = \frac{4\times10^{-5}}{nN}F_z\,. \tag{23.63}$$

The sensitivity can be tailored by optimizing the dimensions of the cantilever.

Capacitance Detection

The capacitance of an arrangement of conductors depends on the geometry. Generally speaking, the capacitance increases for decreasing separations. Two parallel plates form a simple capacitor (see Fig. 23.31, upper left), with the capacitance

$$C = \frac{\varepsilon\varepsilon_0 A}{x} \tag{23.64}$$

where A is the area of the plates, assumed equal, and x is the separation. Alternatively, one can consider a sphere versus an infinite plane (see Fig. 23.31, lower left). Here the capacitance is [23.137]

$$C = 4\pi\varepsilon_0 R\sum_{n=2}^{\infty}\frac{\sinh(\alpha)}{\sinh(n\alpha)}\,, \tag{23.65}$$

where R is the radius of the sphere, and α is defined by

$$\alpha = \ln\left(1+\frac{z}{R}+\sqrt{\frac{z^2}{R^2}+2\frac{z}{R}}\right)\,. \tag{23.66}$$

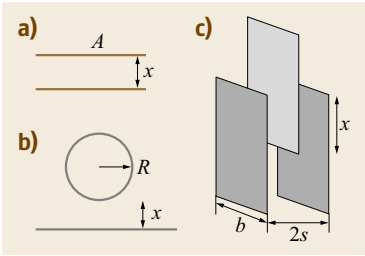


Fig. 23.31a–c Three possible arrangements of a capacitive readout. The **(a)** shows the cross section through a parallel plate capacitor. The **(b)** shows the geometry sphere versus plane. The **(c)** shows the more complicated, but linear capacitive readout

One has to keep in mind that capacitance of a parallel plate capacitor is a nonlinear function of the separation. Using a voltage divider one can circumvent this problem. Figure 23.32a shows a low-pass filter. The output voltage is given by

$$U_{\text{out}} = U_{\approx}\frac{\frac{1}{\mathrm{i}\omega C}}{R+\frac{1}{\mathrm{i}\omega C}} = U_{\approx}\frac{1}{J\omega CR+1} \cong \frac{U_{\approx}}{J\omega CR}\,. \tag{23.67}$$

Here, $U_{\approx}$ is the driving voltage, C is given by (23.64), ω is the excitation frequency, and i is the imaginary unit. The approximate relation in the end is true when $\omega CR \gg 1$. This is equivalent to the statement that C is fed by a current source, since R must be large in this setup. Plugging (23.64) into (23.67) and neglecting the phase information one obtains

$$U_{\text{out}} = \frac{U_{\approx}x}{\omega R\varepsilon\varepsilon_0 A} \tag{23.68}$$

which is linear in the displacement x.

Figure 23.32b shows a capacitive divider. Again the output voltage U_{out} is given by

$$U_{\text{out}} = U_{\approx}\frac{C_1}{C_2+C_1} = U_{\approx}\frac{C_1}{\frac{\varepsilon\varepsilon_0 A}{x}+C_1}\,. \tag{23.69}$$

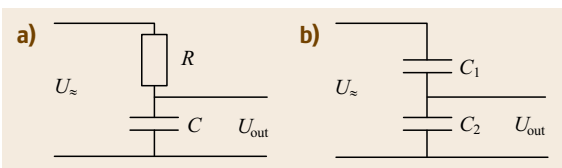


Fig. 23.32a,b Measuring the capacitance. The *left side*, **(a)**, shows a low-pass filter, the *right side*, **(b)**, shows a capacitive divider. C (*left*) or C_2 (*right*) are the capacitances under test

If there is a stray capacitance C_{s} then (23.69) is modified as

$$U_{\text{out}} = U_{\approx}\frac{C_1}{\frac{\varepsilon\varepsilon_0 A}{x}+C_{\text{s}}+C_1}\,. \tag{23.70}$$

Provided $C_{\text{s}}+C_1 \ll C_2$ one has a system which is linear in x. The driving voltage $U_{\approx}$ has to be large (more than 100 V) to have the output voltage in the range of 1 V. The linearity of the readout depends on the capacitance C_1 (Fig. 23.33).

Another idea is to keep the distance constant and to change the relative overlap of the plates (see Fig. 23.31, right side). The capacitance of the moving center plate versus the stationary outer plates becomes

$$C = C_{\text{s}} + 2\frac{\varepsilon\varepsilon_0 bx}{S}\,, \tag{23.71}$$

where the variables are defined in Fig. 23.31. The stray capacitance comprises all effects, including the capacitance of the fringe fields. When length x is comparable to the width b of the plates one can safely assume that the stray capacitance is constant, independent of x. The main disadvantage of this setup is that it is not as easily incorporated in a microfabricated device as the others.

Sensitivity. The capacitance itself is not a measure of the sensitivity, but its derivative is indicative of the signals one can expect. Using the situation described in Fig. 23.31, upper left, and in (23.64) one obtains for the

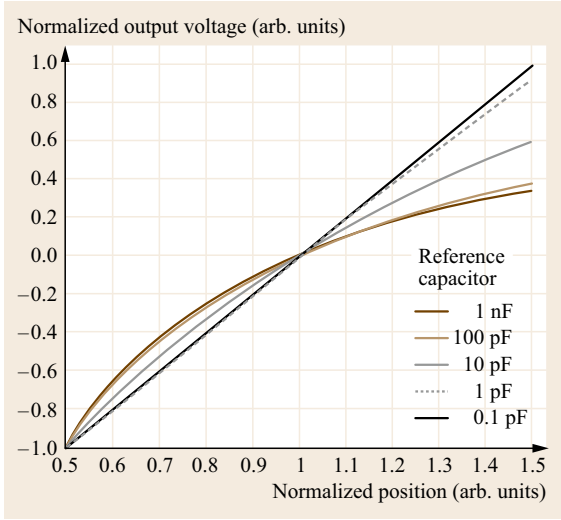


Fig. 23.33 Linearity of the capacitance readout as a function of the reference capacitor

parallel plate capacitor

$$\frac{\mathrm{d}C}{\mathrm{d}x} = -\frac{\varepsilon\varepsilon_0 A}{x^2} \,. \tag{23.72}$$

Assuming a plate area A of 20 µm by 40 µm and a separation of 1 µm one obtains a capacitance of 31 fF (neglecting stray capacitance and the capacitance of the connection leads) and a $\mathrm{d}C/\mathrm{d}x$ of $3.1\times10^{-8}\,\mathrm{F/m} = 31\,\mathrm{fF/\mu m}$. Hence, it is of paramount importance to maximize the area between the two contacts and to minimize the distance x. The latter however is far from being trivial. One has to go to the limits of microfabrication to achieve a decent sensitivity.

If the capacitance is measured by the circuit shown in Fig. 23.32 one obtains for the sensitivity

$$\frac{\mathrm{d}U_{\mathrm{out}}}{U_{\approx}} = \frac{\mathrm{d}x}{\omega R\varepsilon\varepsilon_0 A} \,. \tag{23.73}$$

Using the same value for A as above, setting the reference frequency to 100 kHz, and selecting $R = 1\,\mathrm{G\Omega}$, we get the relative change of the output voltage U_{out} to

$$\frac{\mathrm{d}U_{\mathrm{out}}}{U_{\approx}} = \frac{22.5\times10^{-6}}{\text{Å}}\mathrm{d}x \,. \tag{23.74}$$

A driving voltage of 45 V then translates to a sensitivity of 1 mV/Å. A problem in this setup is the stray capacitances. They are in parallel to the original capacitance and decrease the sensitivity considerably.

Alternatively one could build an oscillator with this capacitance and measure the frequency. RC-oscillators typically have an oscillation frequency of

$$f_{\mathrm{res}} \propto \frac{1}{RC} = \frac{x}{R\varepsilon\varepsilon_0 A} \,. \tag{23.75}$$

Again the resistance R must be of the order of 1 GΩ, when stray capacitances C_{s} are neglected. However C_{s} is of the order of 1 pF. Therefore, one gets $R = 10\,\mathrm{M\Omega}$. Using these values the sensitivity becomes

$$\mathrm{d}f_{\mathrm{res}} = \frac{C\mathrm{d}x}{R\,(C+C_{\mathrm{s}})^2\,x} \approx \frac{0.1\,\mathrm{Hz}}{\text{Å}}\mathrm{d}x \,. \tag{23.76}$$

The bad thing is that the stray capacitances have made the signal nonlinear again. The linearized setup in Fig. 23.31 has a sensitivity of

$$\frac{\mathrm{d}C}{\mathrm{d}x} = 2\frac{\varepsilon\varepsilon_0 b}{s} \,. \tag{23.77}$$

Substituting typical values, $b = 10\,\mu\mathrm{m}$, $s = 1\,\mu\mathrm{m}$ one gets $\mathrm{d}C/\mathrm{d}x = 1.8\times10^{-10}\,\mathrm{F/m}$. It is noteworthy that the sensitivity remains constant for scaled devices.

Implementations. The readout of the capacitance can be done in different ways [23.144, 145]. All include an alternating current or voltage with frequencies in the 100 kHz to 100 MHz range. One possibility is to build a tuned circuit with the capacitance of the cantilever determining the frequency. The resonance frequency of a high-quality Q-tuned circuit is

$$\omega_0 = (LC)^{-\frac{1}{2}} \,, \tag{23.78}$$

where L is the inductance of the circuit. The capacitance C includes not only the sensor capacitance but also the capacitance of the leads. The precision of a frequency measurement is mainly determined by the ratio of L and C

$$Q = \left(\frac{L}{C}\right)^{\frac{1}{2}} \frac{1}{R} \,. \tag{23.79}$$

Here R symbolizes the losses in the circuit. The higher the quality the more precise the frequency measurement. For instance, a frequency of 100 MHz and a capacitance of 1 pF gives an inductance of 250 µH. The quality then becomes 2.5×10^8. This value is an upper limit, since losses are usually too high.

Using a value of $\mathrm{d}C/\mathrm{d}x = 31\,\mathrm{fF/\mu m}$ one gets $\Delta C/\text{Å} = 3.1\,\mathrm{aF/Å}$. With a capacitance of 1 pF one gets

$$\frac{\Delta\omega}{\omega} = \frac{1}{2}\frac{\Delta C}{C} \,,$$
$$\Delta\omega = 100\,\mathrm{MHz}\times\frac{1}{2}\frac{3.1\,\mathrm{aF}}{1\,\mathrm{pF}} = 155\,\mathrm{Hz} \,. \tag{23.80}$$

This is the frequency shift for 1 Å deflection. The calculation shows, that this is a measurable quantity. The quality also indicates that there is no physical reason why this scheme should not work.

23.3.3 Combinations for 3-D-Force Measurements

Three-dimensional force measurements are essential if one wants to know all the details of the interaction between the tip and the cantilever. The straightforward attempt to measure three forces is complicated, since force sensors such as interferometers or capacitive sensors need a minimal detection volume, which often is too large. The second problem is that the force-sensing tip has to be held by some means. This implies that one of the three Cartesian axes is stiffer than the others.

However, through the combination of different sensors one can achieve this goal. Straight cantilevers are employed for these measurements, because they can be

handled analytically. The key observation is that the optical lever method does not determine the position of the end of the cantilever; it measures the orientation. In the previous sections, one has always made use of the fact that for a force along one of the orthogonal symmetry directions at the end of the cantilever (normal force, lateral force, force along the cantilever beam axis) there is a one-to-one correspondence of the tilt angle and the deflection. The problem is that the force along the cantilever beam axis and the normal force create a deflection in the same direction. Hence, what is called the normal force component is actually a mixture of two forces. The deflection of the cantilever is the third quantity, which is not considered in most of the AFMs. A fiber-optic interferometer in parallel with the optical lever measures the deflection. Three measured quantities then allow the separation of the three orthonormal force directions, as is evident from (23.27) and (23.33) [23.12–16].

Alternatively one can put the fast scanning direction along the axis of the cantilever. Forward and backward scans then exert opposite forces F_x. If the piezo movement is linearized, both force components in AFM based on the optical lever detection can be determined. In this case, the normal force is simply the average of the forces in the forward and backward direction. The force from the front, F_x, is the difference of the forces measured in forward and backward direction.

23.3.4 Scanning and Control Systems

Almost all SPMs use piezo translators to scan the tip or the sample. Even the first STM [23.1, 122] and some of the predecessor instruments [23.175, 176] used them. Other materials or setups for nanopositioning have been proposed, but were not successful [23.177, 178].

Piezo Tubes

A popular solution is tube scanners (Fig. 23.34). They are now widely used in SPMs due to their simplicity and their small size [23.154, 179]. The outer electrode is segmented in four equal sectors of 90°. Opposite sectors are driven by signals of the same magnitude, but opposite sign. This gives, through bending, a two-dimensional movement on, approximately, a sphere. The inner electrode is normally driven by the z-signal. It is possible, however, to use only the outer electrodes for scanning and for the z-movement. The main drawback of applying the z-signal to the outer electrodes is that the applied voltage is the sum of both the x- or y-movement and the z-movement. Hence, a larger scan size effectively reduces the available range for the z-control.

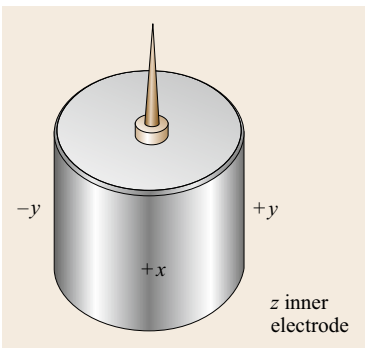


Fig. 23.34 Schematic drawing of a piezoelectric tube scanner. The piezo ceramic is molded into a tube form. The outer electrode is separated into four segments and connected to the scanning voltages. The z-voltage is applied to the inner electrode

Piezo Effect

An electric field applied across a piezoelectric material causes a change in the crystal structure, with expansion in some directions and contraction in others. Also, a net volume change occurs [23.153]. Many SPMs use the transverse piezoelectric effect, where the applied electric field $\boldsymbol{E}$ is perpendicular to the expansion/contraction direction.

$$\Delta L = L\,(\boldsymbol{E}\cdot\boldsymbol{n})\,d_{31} = L\frac{V}{t}d_{31}\,, \tag{23.81}$$

where d_{31} is the transverse piezoelectric constant, V is the applied voltage, t is the thickness of the piezo slab or the distance between the electrodes where the voltage is applied, L is the free length of the piezo slab, and $\boldsymbol{n}$ is the direction of polarization. Piezo translators based on the transverse piezoelectric effect have a wide range of sensitivities, limited mainly by mechanical stability and breakdown voltage.

Scan Range

The calculation of the scanning range of a piezo tube is difficult [23.179–181]. The bending of the tube depends on the electric fields and the nonuniform strain induced. A finite element calculation where the piezo tube was divided into 218 identical elements was used [23.180] to calculate the deflection. On each node the mechanical stress, stiffness, strain, and piezoelectric stress was calculated when a voltage was applied on one electrode. The results were found to be linear on the first iteration and higher order corrections were very small even for large electrode voltages. It was found that to first order the x- and z-movement of the tube could be reasonably well approximated by assuming that the piezo tube is a segment of a torus. Using this model one obtains

$$dx = (V_+ - V_-)\,|d_{31}|\,\frac{L^2}{2td}\,, \tag{23.82}$$

$$dz = (V_+ + V_- - 2V_z)\,|d_{31}|\,\frac{L}{2t}\,, \tag{23.83}$$

where $|d_{31}|$ is the coefficient of the transversal piezoelectric effect, L is the tube's free length, t is the tube's

wall thickness, d is the tube's diameter, V_+ is the voltage on the positive outer electrode while V_- is the voltage of the opposite quadrant negative electrode, and V_z is voltage of the inner electrode. The cantilever or sample mounted on the piezo tube has an additional lateral movement because the point of measurement is not in the end plane of the piezo tube. The additional lateral displacement of the end of the tip is $\ell \sin\varphi \approx \ell\varphi$, where ℓ is the tip length and φ is the deflection angle of the end surface. Assuming that the sample or cantilever are always perpendicular to the end of the walls of the tube and calculating with the torus model one gets for the angle

$$\varphi = \frac{L}{R} = \frac{2dx}{L} \tag{23.84}$$

where R is the radius of curvature of the piezo tube. Using the result of (23.84) one obtains for the additional x-movement

$$dx_{\text{add}} = \ell\varphi = \frac{2dx\ell}{L} = (V_+ - V_-)\,|d_{31}|\,\frac{\ell L}{td} \tag{23.85}$$

and for the additional z-movement due to the x-movement

$$dz_{\text{add}} = \ell - \ell\cos\varphi = \frac{\ell\varphi^2}{2} = \frac{2\ell\,(dx)^2}{L^2} = (V_+ - V_-)^2\,|d_{31}|^2\,\frac{\ell L^2}{2t^2d^2}\,. \tag{23.86}$$

Carr [23.180] assumed for his finite element calculations that the top of the tube was completely free to move and, as a consequence, the top surface was distorted, leading to a deflection angle about half that of the geometrical model. Depending on the attachment of the sample or the cantilever this distortion may be smaller, leading to a deflection angle in-between that of the geometrical model and the one of the finite element calculation.

Nonlinearities and Creep

Piezo materials with a high conversion ratio, i. e., a large d_{31} or small electrode separations, with large scanning ranges are hampered by substantial hysteresis resulting in a deviation from linearity by more than 10%. The sensitivity of the piezo ceramic material (mechanical displacement divided by driving voltage) decreases with reduced scanning range, whereas the hysteresis is reduced. A careful selection of the material for the piezo scanners, the design of the scanners, and of the operating conditions is necessary to get optimum performance.

Linearization Strategies

Passive Linearization: Calculation. The analysis of images affected by piezo nonlinearities [23.182–185] shows that the dominant term is

$$x = AV + BV^2\,, \tag{23.87}$$

where x is the excursion of the piezo, V the applied voltage, and A and B two coefficients describing the sensitivity of the material. Equation (23.87) holds for scanning from $V = 0$ to large V. For the reverse direction the equation becomes

$$x = \tilde{A}V - \tilde{B}\,(V - V_{\max})^2\,, \tag{23.88}$$

where $\tilde{A}$ and $\tilde{B}$ are the coefficients for the back scan, and $V_{\max}$ is the applied voltage at the turning point. Both equations demonstrate the true x-travel is small at the beginning of the scan and becomes larger towards the end. Therefore, images are stretched at the beginning and compressed at the end.

Similar equations hold for the slow scan direction. The coefficients, however, are different. The combined action causes a greatly distorted image. This distortion can be calculated. The data acquisition systems record the signal as a function of V. However, the data is measured as a function of x. Therefore, we have to distribute the x-values evenly across the image and this can be done by inverting an approximation of (23.87). First we write

$$x = AV\left(1 - \frac{B}{A}V\right)\,. \tag{23.89}$$

For $B \ll A$ we can approximate

$$V = \frac{x}{A}\,. \tag{23.90}$$

We now substitute (23.90) into the nonlinear term of (23.89). This gives

$$x = AV\left(1 + \frac{Bx}{A^2}\right)\,,$$
$$V = \frac{x}{A}\,\frac{1}{(1 + Bx/A^2)} \approx \frac{x}{A}\left(1 - \frac{Bx}{A^2}\right)\,. \tag{23.91}$$

Hence an equation of the type

$$x_{\text{true}} = x\left(\alpha - \frac{\beta x}{x_{\max}}\right) \quad \text{with } 1 = \alpha - \beta \tag{23.92}$$

takes out the distortion of an image. α and β are dependent on the scan range, scan speed, and scan history and have to be determined with exactly the same settings

as for the measurement. x_{max} is the maximal scanning range. The condition for α and β guarantees that the image is transformed onto itself.

Similar equations as the empirical one shown above (23.92) can be derived by analyzing the movements of domain walls in piezo ceramics.

Passive Linearization: Measuring the Position. An alternative strategy is to measure the position of the piezo translators. Several possibilities exist:

1. The interferometers described above can be used to measure the elongation of the piezo elongation. Especially the fiber-optic interferometer is easy to implement. The coherence length of the laser only limits the measurement range. However, the signal is of periodic nature. Hence, a direct use of the signal in a feedback circuit for the position is not possible. However, as a measurement tool and, especially, as a calibration tool the interferometer is without competition. The wavelength of the light, for instance in a HeNe laser is so well defined that the precision of the other components determines the error of the calibration or measurement.
2. The movement of the light spot on the quadrant detector can be used to measure the position of a piezo [23.186]. The output current changes by $0.5\,\mathrm{A/cm} \times (P\,[\mathrm{W}]/R\,[\mathrm{cm}])$. Typical values ($P = 1\,\mathrm{mW}$, $R = 0.001\,\mathrm{cm}$) give $0.5\,\mathrm{A/cm}$. The noise limit is typically

$$0.15\,\mathrm{nm} \times \sqrt{(\Delta f\,[Hz])/(H\,[\mathrm{W/cm^2}])}.$$

   Again this means that the laser beam above would have a 0.1 nm noise limitation for a bandwidth of 21 Hz. The advantage of this method is that, in principle, one can linearize two axes with only one detector.
3. A knife-edge blocking part of a light beam incident on a photodiode can be used to measure the position of the piezo. This technique, commonly used in optical shear force detection [23.86, 187], has a sensitivity of better than 0.1 nm.
4. The capacitive detection [23.188, 189] of the cantilever deflection can be applied to the measurement of the piezo elongation. Equations (23.64)–(23.79) apply to the problem. This technique is used in some commercial instruments. The difficulties lie in the avoidance of fringe effects at the borders of the two plates. While conceptually simple, one needs the latest technology in surface preparation to obtain a decent linearity. The electronic circuits used for the readout are often proprietary.
5. Linear variable differential transformers (LVDTs) are a convenient means to measure positions down to 1 nm. They can be used together with a solid-state joint setup, as often used for large scan-range stages. Unlike the capacitive detection there are few difficulties to implementation. The sensors and the detection circuits for LVDTs are available commercially.
6. A popular measurement technique is the use of strain gauges. They are especially sensitive when mounted on a solid-state joint where the curvature is maximal. The resolution depends mainly on the induced curvature. A precision of 1 nm is attainable. The signals are low – a Wheatstone bridge is needed for the readout.

Active Linearization. Active linearization is done with feedback systems. Sensors need to be monotonic. Hence all the systems described above, with the exception of the interferometers are suitable. The most common solutions include the strain gauge approach, the capacitance measurement or the LVDT, which are all electronic solutions. Optical detection systems have the disadvantage that the intensity enters into the calibration.

Alternative Scanning Systems

The first STMs were based on piezo tripods [23.1]. The piezo tripod (Fig. 23.35) is an intuitive way to generate the three-dimensional movement of a tip attached to its center. However, to obtain a suitable stability and scanning range, the tripod needs to be fairly large (about 50 mm). Some instruments use piezo stacks instead of monolithic piezoactuators. They have the tripod arrangement. Piezo stacks are thin layers of piezoactive materials glued together to form a device with up to 200 µm of actuation range. Preloading with a suitable metal casing reduces the nonlinearity.

If one tries to construct a home-built scanning system, the use of linearized scanning tables is recommended. They are built around solid state joints and actuated by piezo stacks. The joints guarantee that the movement is parallel with little deviation from the predefined scanning plane. Because of the construction it

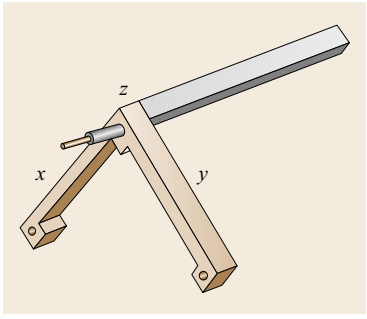


Fig. 23.35 An alternative type of piezo scanner: the tripod

is easy to add measurement devices such as capacitive sensors, LVDTs, or strain gauges which are essential for a closed-loop linearization. Two-dimensional tables can be bought from several manufacturers. They have a linearity of better than 0.1% and a noise level of 10^{-4}–10^{-5} of the maximal scanning range.

Control Systems

Basics. The electronics and software play an important role in the optimal performance of an SPM. Control electronics and software are supplied with commercial SPMs. Control electronic systems can use either analog or digital feedback. While digital feedback offers greater flexibility and the ease of configuration, analog feedback circuits might be better suited for ultralow noise operation. We will describe here the basic setups for AFM.

Figure 23.36 shows a block schematic of a typical AFM feedback loop. The signal from the force transducer is fed into the feedback loop consisting mainly of a subtraction stage to get an error signal and an integrator. The gain of the integrator (high gain corresponds to short integration times) is set as high as possible without generating more than 1% overshoot. High gain minimizes the error margin of the current and forces the tip to follow the contours of constant density of states as precisely as possible. This operating mode is known as constant force mode. A high voltage amplifier amplifies the outputs of the integrator. As AFMs using piezo tubes usually require ± 150 V at the output, the output of the integrator needs to be amplified by a high-voltage amplifier.

In order to scan the sample, additional voltages at high tension are required to drive the piezo. For example, with a tube scanner, four scanning voltages are required, namely $+V_x$, $-V_x$, $+V_y$, and $-V_y$. The x- and y-scanning voltages are generated in a scan generator (analog or computer-controlled). Both voltages are input to the two respective power amplifiers. Two inverting amplifiers generate the input voltages for the other two power amplifiers. The topography of the sample surface is determined by recording the input voltage to the high voltage amplifier for the z-channel as a function of x and y (Constant Force Mode).

Another operating mode is the variable force mode. The gain in the feedback loop is lowered and the scanning speed increased such that the force on the cantilever is no longer constant. Here the force is recorded as a function of x and y.

Force Spectroscopy. Four modes of spectroscopic imaging are in common use with force microscopes: measuring lateral forces, $\partial F/\partial z$, $\partial F/\partial x$ spatially resolved, and measuring force versus distance curves.

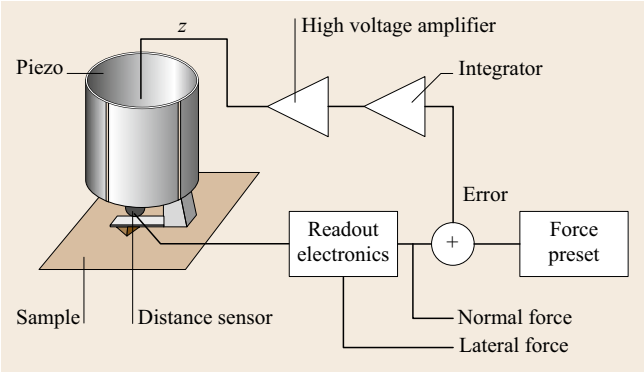


Fig. 23.36 Block schematic of the feedback control loop of an AFM

Lateral forces can be measured by detecting the deflection of a cantilever in a direction orthogonal to the normal direction. The optical lever deflection method most easily does this. Lateral force measurements give indications of adhesion forces between the tip and the sample.

$\partial F/\partial z$ measurements probe the local elasticity of the sample surface. In many cases the measured quantity originates from a volume of a few cubic nanometers. The $\partial F/\partial z$ or local stiffness signal is proportional to Young's modulus, as far as one can define this quantity. Local stiffness is measured by vibrating the cantilever by a small amount in the z-direction. The expected signal for very stiff samples is zero: for very soft samples one obtains, independent of the stiffness, also a constant signal. This signal is again zero for the optical lever deflection and equal to the driving amplitude for interferometric measurements. The best sensitivity is obtained when the compliance of the cantilever matches the stiffness of the sample.

A third spectroscopic quantity is the lateral stiffness. It is measured by applying a small modulation in the x-direction on the cantilever. The signal is again optimal when the lateral compliance of the cantilever matches the lateral stiffness of the sample. The lateral stiffness is, in turn, related to the shear modulus of the sample.

Detailed information on the interaction of the tip and the sample can be gained by measuring force versus distance curves. It is necessary to have cantilevers with high enough compliance to avoid instabilities due to the attractive forces on the sample.

Using the Control Electronics as a 2-D-Measurement Tool. Usually the control electronics of an AFM is used to control the x- and y-piezo signals while several data acquisition channels record the position-dependent signals. The control electronics can be used

in another way: it can be viewed as a two-dimensional function generator. What is normally the x- and y-signal can be used to control two independent variables of an experiment. The control logic of the AFM then ensures that the available parameter space is systematically probed at equally spaced points. An example is friction force curves measured along a line across a step on graphite.

Figure 23.37 shows the connections. The z-piezo is connected as usual, like the x-piezo. However, the y-output is used to command the desired input parameter. The offset of the y-channel determines the position of the tip on the sample surface, together with the x-channel.

Some Imaging Processing Methods

The visualization and interpretation of images from AFMs is intimately connected to the processing of these images. An ideal AFM is a noise-free device that images a sample with perfect tips of known shape and has perfect linear scanning piezos. In reality, AFMs are not that ideal. The scanning device in AFMs is affected by distortions. The distortions are both linear and nonlinear. Linear distortions mainly result from imperfections in the machining of the piezo translators causing crosstalk from the z-piezo to the x- and y-piezos, and vice versa. Among the linear distortions, there are two kinds that are very important. First, scanning piezos invariably have different sensitivities along the different scan axes due to the variation of the piezo material and uneven sizes of the electrode areas. Second, the same reasons might cause the scanning axes not to be orthogonal. Furthermore, the plane in which the piezo scanner moves for constant height z is hardly ever coincident with the sample plane. Hence, a linear ramp is added to the sample data. This ramp is especially bothersome when the height z is displayed as an intensity map.

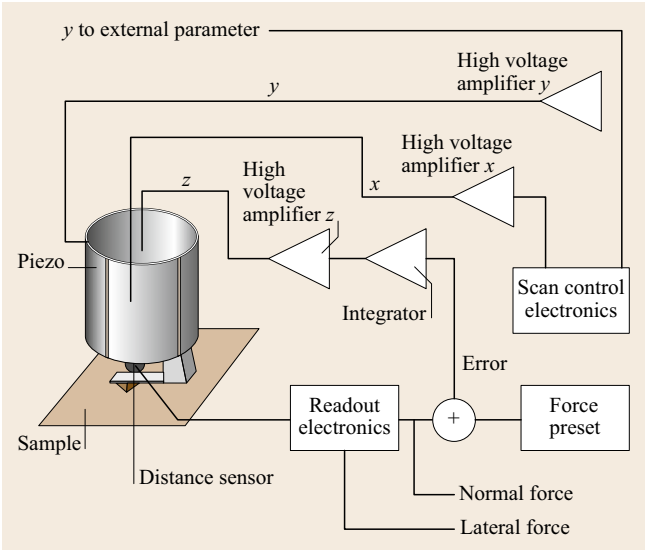


Fig. 23.37 Wiring of an AFM to measure friction force curves along a line

The nonlinear distortions are harder to deal with. They can affect AFM data for a variety of reasons. First, piezoelectric ceramics do have a hysteresis loop, much like ferromagnetic materials. The deviations of piezoceramic materials from linearity increase with increasing amplitude of the driving voltage. The mechanical position for one voltage depends on the previously applied voltages to the piezo. Hence, to obtain the best position accuracy, one should always approach a point on the sample from the same direction. Another type of nonlinear distortion of the images occurs when the scan frequency approaches the upper frequency limit of the x- and y-drive amplifiers or the upper frequency limit of the feedback loop (z-component). This distortion, due to the feedback loop, can only be minimized by reducing the scan frequency. Conversely, there is a simple way to reduce distortions due to the x- and y-piezo drive amplifiers. To keep the system as simple as possible, one normally uses a triangular waveform for driving the scanning piezos. However, triangular waves contain frequency components as multiples of the scan frequency. If the cutoff frequency of the x- and y-drive electronics or of the feedback loop is too close to the scanning frequency (two or three times the scanning frequency), the triangular drive voltage is rounded off at the turning points. This rounding error causes, first, a distortion of the scan linearity and, second, through phase lags, the projection of part of the backward scan onto the forward scan. This type of distortion can be minimized by carefully selecting the scanning frequency and by using driving voltages for the x- and y-piezos with waveforms like trapezoidal waves, which are closer to a sine wave. The values measured for x-, y-, or z-piezos are affected by noise. The origin of this noise can be either electronic, some disturbances, or a property of the sample surface due to adsorbates. In addition to this incoherent noise, interference with the main and other equipment nearby might be present. Depending on the type of noise, one can filter it in the real space or in Fourier space. The most important part of image processing is to visualize the measured data. Typical AFM data sets can consist of many thousands to over a million points per plane. There may be more than one image plane present. The AFM data represents a topography in various data spaces.

Most commercial data acquisition systems use implicitly some kind of data processing. Since the original data is commonly subject to slopes on the surface,

most programs use some kind of slope correction. The least disturbing way is to substrate a plane $z(x, y) = Ax + By + C$ from the data. The coefficients are determined by fitting $z(x, y)$ to the data. Another operation is to subtract a second-order function such as $z(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey + F$. Again, the parameters are determined with a fit. This function is appropriate for almost plane data, where the nonlinearity of the piezos caused such a distortion.

In the image processing software from Digital Instruments, up to three operations are performed on the raw data. First, a zero-order flatten is applied. The flatten operation is used to eliminate image bow in the slow scan direction (caused by a physical bow in the instrument itself), slope in the slow scan direction, and bands in the image (caused by differences in the scan height from one scan line to the next). The flatten operation takes each scan line and subtracts the average value of the height along each scan line from each point in that scan line. This brings each scan line to the same height. Next, a first-order planefit is applied in the fast scan direction. The planefit operation is used to eliminate bow and slope in the fast scan direction. The planefit operation calculates a best-fit plane for the image and subtracts it from the image. This plane has a constant nonzero slope in the fast scan direction. In some cases, a higher order polynomial *plane* may be required. Depending upon the quality of the raw data, the flatten operation and/or the planefit operation may not be required at all.

23.4 Conclusion

Since introduction of the STM in 1981 and AFM in 1985, many variations of probe-based microscopies, referred to as SPMs, have been developed. While the pure imaging capabilities of SPM techniques are dominated by the application of these methods at their early development stages, the physics of probe-sample interactions and the quantitative analyses of tribological, electronic, magnetic, biological, and chemical surfaces have now become of increasing interest. Nanoscale science and technology are strongly driven by SPMs which allow investigation and manipulation of surfaces down to the atomic scale. With growing understanding of the underlying interaction mechanisms, SPMs have found application in many fields outside basic research fields. In addition, various derivatives of all these methods have been developed for special applications, some of them with targets far beyond microscopy.

References

23.1 G. Binnig, H. Rohrer, C. Gerber, E. Weibel: Surface studies by scanning tunneling microscopy, Phys. Rev. Lett. **49**, 57–61 (1982)

23.2 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56**, 930–933 (1986)

23.3 G. Binnig, C. Gerber, E. Stoll, T.R. Albrecht, C.F. Quate: Atomic resolution with atomic force microscope, Europhys. Lett. **3**, 1281–1286 (1987)

23.4 C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59**, 1942–1945 (1987)

23.5 R. Erlandsson, G.M. McClelland, C.M. Mate, S. Chiang: Atomic force microscopy using optical interferometry, J. Vac. Sci. Technol. A **6**, 266–270 (1988)

23.6 O. Marti, J. Colchero, J. Mlynek: Combined scanning force and friction microscopy of mica, Nanotechnology **1**, 141–144 (1990)

23.7 G. Meyer, N.M. Amer: Simultaneous measurement of lateral and normal forces with an optical-beam-deflection atomic force microscope, Appl. Phys. Lett. **57**, 2089–2091 (1990)

23.8 B. Bhushan, J. Ruan: Atomic-scale friction measurements using friction force microscopy: Part II – Application to magnetic media, ASME J. Tribol. **116**, 389–396 (1994)

23.9 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology: Friction, wear, and lubrication at the atomic scale, Nature **374**, 607–616 (1995)

23.10 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

23.11 B. Bhushan: *Nanotribology and Nanomechanics I & II*, 3rd edn. (Springer, Berlin, Heidelberg 2011)

23.12 S. Fujisawa, M. Ohta, T. Konishi, Y. Sugawara, S. Morita: Difference between the forces measured by an optical lever deflection and by an optical interferometer in an atomic force microscope, Rev. Sci. Instrum. **65**, 644–647 (1994)

23.13 S. Fujisawa, E. Kishi, Y. Sugawara, S. Morita: Fluctuation in 2-dimensional stick-slip phenomenon observed with 2-dimensional frictional force microscope, Jpn. J. Appl. Phys. **33**, 3752–3755 (1994)

23.14 S. Grafstrom, J. Ackermann, T. Hagen, R. Neumann, O. Probst: Analysis of lateral force effects on the topography in scanning force microscopy,

J. Vac. Sci. Technol. B **12**, 1559–1564 (1994)
23.15 R.M. Overney, H. Takano, M. Fujihira, W. Paulus, H. Ringsdorf: Anisotropy in friction and molecular stick-slip motion, Phys. Rev. Lett. **72**, 3546–3549 (1994)
23.16 R.J. Warmack, X.Y. Zheng, T. Thundat, D.P. Allison: Friction effects in the deflection of atomic force microscope cantilevers, Rev. Sci. Instrum. **65**, 394–399 (1994)
23.17 N.A. Burnham, D.D. Domiguez, R.L. Mowery, R.J. Colton: Probing the surface forces of monolayer films with an atomic force microscope, Phys. Rev. Lett. **64**, 1931–1934 (1990)
23.18 N.A. Burham, R.J. Colton, H.M. Pollock: Interpretation issues in force microscopy, J. Vac. Sci. Technol. A **9**, 2548–2556 (1991)
23.19 C.D. Frisbie, L.F. Rozsnyai, A. Noy, M.S. Wrighton, C.M. Lieber: Functional group imaging by chemical force microscopy, Science **265**, 2071–2074 (1994)
23.20 V.N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14**, 2378–2391 (1996)
23.21 V. Scherer, B. Bhushan, U. Rabe, W. Arnold: Local elasticity and lubrication measurements using atomic force and friction force microscopy at ultrasonic frequencies, IEEE Trans. Magn. **33**, 4077–4079 (1997)
23.22 V. Scherer, W. Arnold, B. Bhushan: Lateral force microscopy using acoustic friction force microscopy, Surf. Interface Anal. **27**, 578–587 (1999)
23.23 B. Bhushan, S. Sundararajan: Micro-/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46**, 3793–3804 (1998)
23.24 U. Krotil, T. Stifter, H. Waschipky, K. Weishaupt, S. Hild, O. Marti: Pulse force mode: A new method for the investigation of surface properties, Surf. Interface Anal. **27**, 336–340 (1999)
23.25 B. Bhushan, C. Dandavate: Thin-film friction and adhesion studies using atomic force microscopy, J. Appl. Phys. **87**, 1201–1210 (2000)
23.26 B. Bhushan: *Micro/Nanotribology and its Applications*, Vol. E330 (Kluwer, Dordrecht 1997)
23.27 B. Bhushan: Principles of tribology. In: *Modern Tribology Handbook*, Vol. 1 (CRC, Boca Raton 2001)
23.28 B. Bhushan: *Fundamentals of Tribology and Bridging the Gap between the Macro- and Micro/Nanoscales*, NATO Science Series II, Vol. 10 (Kluwer, Dordrecht 2001)
23.29 B. Bhushan: Nanotribology and nanomechanics, Wear **259**, 1507–1531 (2005)
23.30 B. Bhushan: Nanotribology, nanomechanics and nanomaterials characterization, Philos. Trans. R. Soc. A **366**, 1351–1381 (2008)
23.31 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)
23.32 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)
23.33 M. Reinstädtler, U. Rabe, V. Scherer, U. Hartmann, A. Goldade, B. Bhushan, W. Arnold: On the nanoscale measurement of friction using atomic force microscope cantilever torsional resonances, Appl. Phys. Lett. **82**, 2604–2606 (2003)
23.34 N.A. Burnham, R.J. Colton: Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope, J. Vac. Sci. Technol. A **7**, 2906–2913 (1989)
23.35 P. Maivald, H.J. Butt, S.A.C. Gould, C.B. Prater, B. Drake, J.A. Gurley, V.B. Elings, P.K. Hansma: Using force modulation to image surface elasticities with the atomic force microscope, Nanotechnology **2**, 103–106 (1991)
23.36 B. Bhushan, V.N. Koinkar, J. Ruan: Microtribology of magnetic media, Proc. Inst. Mech. Eng. Part J **208**, 17–29 (1994)
23.37 B. Bhushan, A.V. Kulkarni, W. Bonin, J.T. Wyrobek: Nano/picoindentation measurements using capacitive transducer in atomic force microscopy, Philos. Mag. A **74**, 1117–1128 (1996)
23.38 B. Bhushan, V.N. Koinkar: Nanoindentation hardness measurements using atomic force microscopy, Appl. Phys. Lett. **75**, 5741–5746 (1994)
23.39 D. DeVecchio, B. Bhushan: Localized surface elasticity measurements using an atomic force microscope, Rev. Sci. Instrum. **68**, 4498–4505 (1997)
23.40 B. Bhushan: Wear and mechanical characterisation on micro-to picoscales using AFM, Int. Mat. Rev. **44**, 105–117 (1999)
23.41 B. Bhushan: Nano- to microscale wear and mechanical characterization studies using scanning probe microscopy, Wear **251**, 1105–1123 (2001)
23.42 S. Amelio, A.V. Goldade, U. Rabe, V. Scherer, B. Bhushan, W. Arnold: Measurements of mechanical properties of ultra-thin diamond-like carbon coatings using atomic force acoustic microscopy, Thin Solid Films **392**, 75–84 (2001)
23.43 B. Bhushan, J. Qi: Phase contrast imaging of nanocomposites and molecularly-thick lubricant films in magnetic media, Nanotechnology **14**, 886–895 (2003)
23.44 T. Kasai, B. Bhushan, L. Huang, C. Su: Topography and phase imaging using the torsional resonance mode, Nanotechnology **15**, 731–742 (2004)
23.45 M. Reinstaedtler, T. Kasai, U. Rabe, B. Bhushan, W. Arnold: Imaging and measurement of elasticity and friction using the TR mode, J. Phys. D: Appl. Phys. **38**, R269–R282 (2005)
23.46 D.M. Eigler, E.K. Schweizer: Positioning single atoms with a scanning tunnelling microscope, Nature **344**, 524–528 (1990)
23.47 A.L. Weisenhorn, J.E. MacDougall, J.A.C. Gould, S.D. Cox, W.S. Wise, J. Massie, P. Maivald, V.B. Elings, G.D. Stucky, P.K. Hansma: Imaging and manipulating of molecules on a zeolite surface with an atomic force microscope, Science **247**, 1330–1333 (1990)
23.48 I.W. Lyo, P. Avouris: Field-induced nanometer-to-atomic-scale manipulation of silicon surfaces with the STM, Science **253**, 173–176 (1991)
23.49 O.M. Leung, M.C. Goh: Orientation ordering of polymers by atomic force microscope tip-surface interactions, Science **225**, 64–66 (1992)

23.50 D.W. Abraham, H.J. Mamin, E. Ganz, J. Clark: Surface modification with the scanning tunneling microscope, IBM J. Res. Dev. **30**, 492–499 (1986)

23.51 R.M. Silver, E.E. Ehrichs, A.L. de Lozanne: Direct writing of submicron metallic features with a scanning tunnelling microscope, Appl. Phys. Lett. **51**, 247–249 (1987)

23.52 A. Kobayashi, F. Grey, R.S. Williams, M. Ano: Formation of nanometer-scale grooves in silicon with a scanning tunneling microscope, Science **259**, 1724–1726 (1993)

23.53 B. Parkinson: Layer-by-layer nanometer scale etching of two-dimensional substrates using the scanning tunneling microscopy, J. Am. Chem. Soc. **112**, 7498–7502 (1990)

23.54 A. Majumdar, P.I. Oden, J.P. Carrejo, L.A. Nagahara, J.J. Graham, J. Alexander: Nanometer-scale lithography using the atomic force microscope, Appl. Phys. Lett. **61**, 2293–2295 (1992)

23.55 B. Bhushan: Micro-/nanotribology and its applications to magnetic storage devices and MEMS, Tribol. Int. **28**, 85–96 (1995)

23.56 L. Tsau, D. Wang, K.L. Wang: Nanometer scale patterning of silicon(100) surface by an atomic force microscope operating in air, Appl. Phys. Lett. **64**, 2133–2135 (1994)

23.57 E. Delawski, B.A. Parkinson: Layer-by-layer etching of two-dimensional metal chalcogenides with the atomic force microscope, J. Am. Chem. Soc. **114**, 1661–1667 (1992)

23.58 B. Bhushan, G.S. Blackman: Atomic force microscopy of magnetic rigid disks and sliders and its applications to tribology, ASME J. Tribol. **113**, 452–458 (1991)

23.59 O. Marti, B. Drake, P.K. Hansma: Atomic force microscopy of liquid-covered surfaces: atomic resolution images, Appl. Phys. Lett. **51**, 484–486 (1987)

23.60 B. Drake, C.B. Prater, A.L. Weisenhorn, S.A.C. Gould, T.R. Albrecht, C.F. Quate, D.S. Cannell, H.G. Hansma, P.K. Hansma: Imaging crystals, polymers and processes in water with the atomic force microscope, Science **243**, 1586–1589 (1989)

23.61 M. Binggeli, R. Christoph, H.E. Hintermann, J. Colchero, O. Marti: Friction force measurements on potential controlled graphite in an electrolytic environment, Nanotechnology **4**, 59–63 (1993)

23.62 G. Meyer, N.M. Amer: Novel optical approach to atomic force microscopy, Appl. Phys. Lett. **53**, 1045–1047 (1988)

23.63 J.H. Coombs, J.B. Pethica: Properties of vacuum tunneling currents: Anomalous barrier heights, IBM J. Res. Dev. **30**, 455–459 (1986)

23.64 M.D. Kirk, T. Albrecht, C.F. Quate: Low-temperature atomic force microscopy, Rev. Sci. Instrum. **59**, 833–835 (1988)

23.65 F.J. Giessibl, C. Gerber, G. Binnig: A low-temperature atomic force/scanning tunneling microscope for ultrahigh vacuum, J. Vac. Sci. Technol. B **9**, 984–988 (1991)

23.66 T.R. Albrecht, P. Grutter, D. Rugar, D.P.E. Smith: Low temperature force microscope with all-fiber interferometer, Ultramicroscopy **42–44**, 1638–1646 (1992)

23.67 H.J. Hug, A. Moser, T. Jung, O. Fritz, A. Wadas, I. Parashikor, H.J. Güntherodt: Low temperature magnetic force microscopy, Rev. Sci. Instrum. **64**, 2920–2925 (1993)

23.68 C. Basire, D.A. Ivanov: Evolution of the lamellar structure during crystallization of a semicrystalline-amorphous polymer blend: Time-resolved hot-stage SPM study, Phys. Rev. Lett. **85**, 5587–5590 (2000)

23.69 H. Liu, B. Bhushan: Investigation of nanotribological properties of self-assembled monolayers with alkyl and biphenyl spacer chains, Ultramicroscopy **91**, 185–202 (2002)

23.70 J. Foster, J. Frommer: Imaging of liquid crystal using a tunneling microscope, Nature **333**, 542–547 (1988)

23.71 D. Smith, H. Horber, C. Gerber, G. Binnig: Smectic liquid crystal monolayers on graphite observed by scanning tunneling microscopy, Science **245**, 43–45 (1989)

23.72 D. Smith, J. Horber, G. Binnig, H. Nejoh: Structure, registry and imaging mechanism of alkylcyanobiphenyl molecules by tunnelling microscopy, Nature **344**, 641–644 (1990)

23.73 Y. Andoh, S. Oguchi, R. Kaneko, T. Miyamoto: Evaluation of very thin lubricant films, J. Phys. D **25**, A71–A75 (1992)

23.74 Y. Martin, C.C. Williams, H.K. Wickramasinghe: Atomic force microscope-force mapping and profiling on a sub 100 Å scale, J. Appl. Phys. **61**, 4723–4729 (1987)

23.75 J.E. Stern, B.D. Terris, H.J. Mamin, D. Rugar: Deposition and imaging of localized charge on insulator surfaces using a force microscope, Appl. Phys. Lett. **53**, 2717–2719 (1988)

23.76 K. Yamanaka, H. Ogisco, O. Kolosov: Ultrasonic force microscopy for nanometer resolution subsurface imaging, Appl. Phys. Lett. **64**, 178–180 (1994)

23.77 K. Yamanaka, E. Tomita: Lateral force modulation atomic force microscope for selective imaging of friction forces, Jpn. J. Appl. Phys. **34**, 2879–2882 (1995)

23.78 U. Rabe, K. Janser, W. Arnold: Vibrations of free and surface-coupled atomic force microscope: Theory and experiment, Rev. Sci. Instrum. **67**, 3281–3293 (1996)

23.79 Y. Martin, H.K. Wickramasinghe: Magnetic imaging by force microscopy with 1000 Å resolution, Appl. Phys. Lett. **50**, 1455–1457 (1987)

23.80 D. Rugar, H.J. Mamin, P. Güthner, S.E. Lambert, J.E. Stern, I. McFadyen, T. Yogi: Magnetic force microscopy – General principles and application to longitudinal recording media, J. Appl. Phys. **63**, 1169–1183 (1990)

23.81 C. Schönenberger, S.F. Alvarado: Understanding magnetic force microscopy, Z. Phys. B **80**, 373–383 (1990)

23.82 U. Hartmann: Magnetic force microscopy, Annu. Rev. Mater. Sci. **29**, 53–87 (1999)

23.83 A. Avila, B. Bhushan: Electrical measurement techniques in atomic force microscopy, Crit. Rev. Solid State Mater. Sci. **35**, 38–51 (2010)

23.84 D.W. Pohl, W. Denk, M. Lanz: Optical stethoscopy-image recording with resolution $\lambda/20$, Appl. Phys. Lett. **44**, 651–653 (1984)

23.85 E. Betzig, J.K. Troutman, T.D. Harris, J.S. Weiner, R.L. Kostelak: Breaking the diffraction barrier – optical microscopy on a nanometric scale, Science **251**, 1468–1470 (1991)

23.86 E. Betzig, P.L. Finn, J.S. Weiner: Combined shear force and near-field scanning optical microscopy, Appl. Phys. Lett. **60**, 2484 (1992)

23.87 P.F. Barbara, D.M. Adams, D.B. O'Connor: Characterization of organic thin film materials with near-field scanning optical microscopy (NSOM), Annu. Rev. Mater. Sci. **29**, 433–469 (1999)

23.88 C.C. Williams, H.K. Wickramasinghe: Scanning thermal profiler, Appl. Phys. Lett. **49**, 1587–1589 (1986)

23.89 C.C. Williams, H.K. Wickramasinghe: Microscopy of chemical-potential variations on an atomic scale, Nature **344**, 317–319 (1990)

23.90 A. Majumdar: Scanning thermal microscopy, Annu. Rev. Mater. Sci. **29**, 505–585 (1999)

23.91 O.E. Husser, D.H. Craston, A.J. Bard: Scanning electrochemical microscopy – High resolution deposition and etching of materials, J. Electrochem. Soc. **136**, 3222–3229 (1989)

23.92 Y. Martin, D.W. Abraham, H.K. Wickramasinghe: High-resolution capacitance measurement and potentiometry by force microscopy, Appl. Phys. Lett. **52**, 1103–1105 (1988)

23.93 M. Nonnenmacher, M.P. O'Boyle, H.K. Wickramasinghe: Kelvin probe force microscopy, Appl. Phys. Lett. **58**, 2921–2923 (1991)

23.94 J.M.R. Weaver, D.W. Abraham: High resolution atomic force microscopy potentiometry, J. Vac. Sci. Technol. B **9**, 1559–1561 (1991)

23.95 D. DeVecchio, B. Bhushan: Use of a nanoscale Kelvin probe for detecting wear precursors, Rev. Sci. Instrum. **69**, 3618–3624 (1998)

23.96 B. Bhushan, A.V. Goldade: Measurements and analysis of surface potential change during wear of single-crystal silicon (100) at ultralow loads using Kelvin probe microscopy, Appl. Surf. Sci. **157**, 373–381 (2000)

23.97 P.K. Hansma, B. Drake, O. Marti, S.A.C. Gould, C.B. Prater: The scanning ion-conductance microscope, Science **243**, 641–643 (1989)

23.98 C.B. Prater, P.K. Hansma, M. Tortonese, C.F. Quate: Improved scanning ion-conductance microscope using microfabricated probes, Rev. Sci. Instrum. **62**, 2634–2638 (1991)

23.99 J. Matey, J. Blanc: Scanning capacitance microscopy, J. Appl. Phys. **57**, 1437–1444 (1985)

23.100 C.C. Williams: Two-dimensional dopant profiling by scanning capacitance microscopy, Annu. Rev. Mater. Sci. **29**, 471–504 (1999)

23.101 D.T. Lee, J.P. Pelz, B. Bhushan: Instrumentation for direct, low frequency scanning capacitance microscopy, and analysis of position dependent stray capacitance, Rev. Sci. Instrum. **73**, 3523–3533 (2002)

23.102 B. Bhushan, H. Fuchs, S. Hosaka: *Applied Scanning Probe Methods* (Springer, Berlin, Heidelberg 2004)

23.103 B. Bhushan, H. Fuchs: *Applied Scanning Probe Methods II – Scanning Probe Microscopy Techniques; III – Characterization; IV – Industrial Applications* (Springer, Berlin, Heidelberg 2006)

23.104 B. Bhushan, H. Fuchs, S. Kawata: *Applied Scanning Probe Methods V – Scanning Probe Microscopy Techniques* (Springer, Berlin, Heidelberg 2007)

23.105 B. Bhushan, S. Kawata: *Applied Scanning Probe Methods VI – Characterization* (Springer, Berlin, Heidelberg 2007)

23.106 B. Bhushan, H. Fuchs: *Applied Scanning Probe Methods VII – Biomimetics and Industrial Applications* (Springer, Berlin, Heidelberg 2007)

23.107 B. Bhushan, H. Fuchs, M. Tomitori: *Applied Scanning Probe Methods VIII – Scanning Probe Microscopy Techniques; IX – Characterization; X – Biomimetics and Industrial Applications* (Springer, Berlin, Heidelberg 2008)

23.108 B. Bhushan, H. Fuchs: *Applied Scanning Probe Methods XI – Scanning Probe Microscopy Techniques; XII – Characterization; XIII – Biomimetics and Industrial Applications* (Springer, Berlin, Heidelberg 2009)

23.109 H. Fuchs, B. Bhushan: *Biosystems – Investigated by Scanning Probe Microscopy* (Springer, Berlin, Heidelberg 2010)

23.110 B. Bhushan: *Scanning Probe Microscopy in Nanoscience and Nanotechnology*. Vols. 1–3 (Springer, Berlin, Heidelberg 2010–2013)

23.111 I. Giaever: Energy gap in superconductors measured by electron tunneling, Phys. Rev. Lett. **5**, 147–148 (1960)

23.112 P.K. Hansma, J. Tersoff: Scanning tunneling microscopy, J. Appl. Phys. **61**, R1–R23 (1987)

23.113 D. Sarid, V. Elings: Review of scanning force microscopy, J. Vac. Sci. Technol. B **9**, 431–437 (1991)

23.114 U. Durig, O. Zuger, A. Stalder: Interaction force detection in scanning probe microscopy: Methods and applications, J. Appl. Phys. **72**, 1778–1797 (1992)

23.115 J. Frommer: Scanning tunneling microscopy and atomic force microscopy in organic chemistry, Angew. Chem. Int. Ed. **31**, 1298–1328 (1992)

23.116 H.J. Güntherodt, R. Wiesendanger (Eds.): *Scanning Tunneling Microscopy I: General Principles and Applications to Clean and Adsorbate-Covered Surfaces* (Springer, Berlin, Heidelberg 1992)

23.117 R. Wiesendanger, H.J. Güntherodt (Eds.): *Scanning Tunneling Microscopy II: Further Applications and Related Scanning Techniques* (Springer, Berlin, Heidelberg 1992)

23.118 D.A. Bonnell (Ed.): *Scanning Tunneling Microscopy and Spectroscopy – Theory, Techniques, and Applications* (VCH, New York 1993)

23.119 O. Marti, M. Amrein (Eds.): *STM and SFM in Biology* (Academic, San Diego 1993)

23.120 J.A. Stroscio, W.J. Kaiser (Eds.): *Scanning Tunneling Microscopy* (Academic, Boston 1993)

23.121 H.J. Güntherodt, D. Anselmetti, E. Meyer (Eds.): *Forces in Scanning Probe Methods* (Kluwer, Dordrecht 1995)

23.122 G. Binnig, H. Rohrer: Scanning tunnelling microscopy, Surf. Sci. **126**, 236–244 (1983)

23.123 B. Bhushan, J. Ruan, B.K. Gupta: A scanning tunnelling microscopy study of fullerene films, J. Phys. D **26**, 1319–1322 (1993)

23.124 R.L. Nicolaides, W.E. Yong, W.F. Packard, H.A. Zhou: Scanning tunneling microscope tip structures, J. Vac. Sci. Technol. A **6**, 445–447 (1988)

23.125 J.P. Ibe, P.P. Bey, S.L. Brandon, R.A. Brizzolara, N.A. Burnham, D.P. DiLella, K.P. Lee, C.R.K. Marrian, R.J. Colton: On the electrochemical etching of tips for scanning tunneling microscopy, J. Vac. Sci. Technol. A **8**, 3570–3575 (1990)

23.126 R. Kaneko, S. Oguchi: Ion-implanted diamond tip for a scanning tunneling microscope, Jpn. J. Appl. Phys. **28**, 1854–1855 (1990)

23.127 F.J. Giessibl, Y. Sugawara, S. Morita, H. Hosoi, K. Sueoka, K. Mukasa, A. Sasahara, H. Onishi: Noncontact atomic force microscopy and related topics. In: *Nanotribology and Nanomechanics I*, 3rd edn., ed. by B. Bhushan (Springer, Berlin, Heidelberg 2011) pp. 195–237

23.128 A. Schirmeisen, B. Anczykowski, H. Hoelscher, F. Fuchs: Dynamic Modes of Atomic Force Microscopy. In: *Nanotribology and Nanomechanics I*, 3rd edn., ed. by B. Bhushan (Springer, Berlin, Heidelberg 2011) pp. 307–353

23.129 F.J. Giessibl: Atomic resolution of the silicon(111)–(7×7) surface by atomic force microscopy, Science **267**, 68–71 (1995)

23.130 B. Anczykowski, D. Krüger, K.L. Babcock, H. Fuchs: Basic properties of dynamic force spectroscopy with the scanning force microscope in experiment and simulation, Ultramicroscopy **66**, 251–259 (1996)

23.131 T.R. Albrecht, C.F. Quate: Atomic resolution imaging of a nonconductor by atomic force microscopy, J. Appl. Phys. **62**, 2599–2602 (1987)

23.132 S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P.K. Hansma: An atomic-resolution atomic-force microscope implemented using an optical lever, J. Appl. Phys. **65**, 164–167 (1989)

23.133 G. Meyer, N.M. Amer: Optical-beam-deflection atomic force microscopy: The NaCl(001) surface, Appl. Phys. Lett. **56**, 2100–2101 (1990)

23.134 A.L. Weisenhorn, M. Egger, F. Ohnesorge, S.A.C. Gould, S.P. Heyn, H.G. Hansma, R.L. Sinsheimer, H.E. Gaub, P.K. Hansma: Molecular resolution images of Langmuir–Blodgett films and DNA by atomic force microscopy, Langmuir **7**, 8–12 (1991)

23.135 J. Ruan, B. Bhushan: Atomic-scale and microscale friction of graphite and diamond using friction force microscopy, J. Appl. Phys. **76**, 5022–5035 (1994)

23.136 D. Rugar, P.K. Hansma: Atomic force microscopy, Phys. Today **43**, 23–30 (1990)

23.137 D. Sarid: *Scanning Force Microscopy* (Oxford Univ. Press, Oxford 1991)

23.138 G. Binnig: Force microscopy, Ultramicroscopy **42–44**, 7–15 (1992)

23.139 E. Meyer: Atomic force microscopy, Surf. Sci. **41**, 3–49 (1992)

23.140 H.K. Wickramasinghe: Progress in scanning probe microscopy, Acta Mater. **48**, 347–358 (2000)

23.141 A.J. den Boef: The influence of lateral forces in scanning force microscopy, Rev. Sci. Instrum. **62**, 88–92 (1991)

23.142 M. Radmacher, R.W. Tillman, M. Fritz, H.E. Gaub: From molecules to cells: Imaging soft samples with the atomic force microscope, Science **257**, 1900–1905 (1992)

23.143 F. Ohnesorge, G. Binnig: True atomic resolution by atomic force microscopy through repulsive and attractive forces, Science **260**, 1451–1456 (1993)

23.144 G. Neubauer, S.R. Coben, G.M. McClelland, D. Horne, C.M. Mate: Force microscopy with a bidirectional capacitance sensor, Rev. Sci. Instrum. **61**, 2296–2308 (1990)

23.145 T. Goddenhenrich, H. Lemke, U. Hartmann, C. Heiden: Force microscope with capacitive displacement detection, J. Vac. Sci. Technol. A **8**, 383–387 (1990)

23.146 U. Stahl, C.W. Yuan, A.L. Delozanne, M. Tortonese: Atomic force microscope using piezoresistive cantilevers and combined with a scanning electron microscope, Appl. Phys. Lett. **65**, 2878–2880 (1994)

23.147 R. Kassing, E. Oesterschulze: Sensors for scanning probe microscopy. In: *Micro-/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 35–54

23.148 C.M. Mate: Atomic-force-microscope study of polymer lubricants on silicon surfaces, Phys. Rev. Lett. **68**, 3323–3326 (1992)

23.149 S.P. Jarvis, A. Oral, T.P. Weihs, J.B. Pethica: A novel force microscope and point contact probe, Rev. Sci. Instrum. **64**, 3515–3520 (1993)

23.150 D. Rugar, H.J. Mamin, P. Güthner: Improved fiber-optical interferometer for atomic force microscopy, Appl. Phys. Lett. **55**, 2588–2590 (1989)

23.151 C. Schönenberger, S.F. Alvarado: A differential interferometer for force microscopy, Rev. Sci. Instrum. **60**, 3131–3135 (1989)

23.152 D. Sarid, D. Iams, V. Weissenberger, L.S. Bell: Compact scanning-force microscope using laser diode, Opt. Lett. **13**, 1057–1059 (1988)

23.153 N.W. Ashcroft, N.D. Mermin: *Solid State Physics* (Holt Reinhart and Winston, New York 1976)

23.154 G. Binnig, D.P.E. Smith: Single-tube three-dimensional scanner for scanning tunneling microscopy, Rev. Sci. Instrum. **57**, 1688 (1986)

23.155 S.I. Park, C.F. Quate: Digital filtering of STM images, J. Appl. Phys. **62**, 312 (1987)

23.156 J.W. Cooley, J.W. Tukey: An algorithm for machine calculation of complex Fourier series, Math. Comput. **19**, 297 (1965)

23.157 J. Ruan, B. Bhushan: Atomic-scale friction measurements using friction force microscopy: Part I – General principles and new measurement techniques, ASME J. Tribol. **116**, 378–388 (1994)

23.158 T.R. Albrecht, S. Akamine, T.E. Carver, C.F. Quate: Microfabrication of cantilever styli for the atomic force microscope, J. Vac. Sci. Technol. A **8**, 3386–3396 (1990)

23.159 O. Marti, S. Gould, P.K. Hansma: Control electronics for atomic force microscopy, Rev. Sci. Instrum. **59**, 836–839 (1988)

23.160 O. Wolter, T. Bayer, J. Greschner: Micromachined silicon sensors for scanning force microscopy, J. Vac. Sci. Technol. B **9**, 1353–1357 (1991)

23.161 E. Meyer, R. Overney, R. Luthi, D. Brodbeck: Friction force microscopy of mixed Langmuir–Blodgett films, Thin Solid Films **220**, 132–137 (1992)

23.162 H.J. Dai, J.H. Hafner, A.G. Rinzler, D.T. Colbert, R.E. Smalley: Nanotubes as nanoprobes in scanning probe microscopy, Nature **384**, 147–150 (1996)

23.163 J.H. Hafner, C.L. Cheung, A.T. Woolley, C.M. Lieber: Structural and functional imaging with carbon nanotube AFM probes, Prog. Biophys. Mol. Biol. **77**, 73–110 (2001)

23.164 M. Palacio, B. Bhushan: Normal and lateral force calibration techniques for AFM cantilevers, Crit. Rev. Solid State **35**, 73–104 (2010)

23.165 G.S. Blackman, C.M. Mate, M.R. Philpott: Interaction forces of a sharp tungsten tip with molecular films on silicon surface, Phys. Rev. Lett. **65**, 2270–2273 (1990)

23.166 S.J. O'Shea, M.E. Welland, T. Rayment: Atomic force microscope study of boundary layer lubrication, Appl. Phys. Lett. **61**, 2240–2242 (1992)

23.167 J.P. Cleveland, S. Manne, D. Bocek, P.K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64**, 403–405 (1993)

23.168 D.W. Pohl: Some design criteria in STM, IBM J. Res. Dev. **30**, 417 (1986)

23.169 W.T. Thomson, M.D. Dahleh: *Theory of Vibration with Applications*, 5th edn. (Prentice Hall, Upper Saddle River 1998)

23.170 J. Colchero: *Reibungskraftmikroskopie*, Ph.D. Thesis (University of Konstanz, Konstanz 1993), in German

23.171 G.M. McClelland, R. Erlandsson, S. Chiang: Atomic force microscopy: General principles and a new implementation. In: *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 6B, ed. by D.O. Thompson, D.E. Chimenti (Plenum, New York 1987) pp. 1307–1314

23.172 Y.R. Shen: *The Principles of Nonlinear Optics* (Wiley, New York 1984)

23.173 T. Baumeister, S.L. Marks: *Standard Handbook for Mechanical Engineers*, 7th edn. (McGraw-Hill, New York 1967)

23.174 J. Colchero, O. Marti, H. Bielefeldt, J. Mlynek: Scanning force and friction microscopy, Phys. Status Solidi (a) **131**, 73–75 (1991)

23.175 R. Young, J. Ward, F. Scire: Observation of metal-vacuum-metal tunneling, field emission, and the transition region, Phys. Rev. Lett. **27**, 922 (1971)

23.176 R. Young, J. Ward, F. Scire: The topographiner: An instrument for measuring surface microtopography, Rev. Sci. Instrum. **43**, 999 (1972)

23.177 C. Gerber, O. Marti: Magnetostrictive positioner, IBM Tech. Discl. Bull. **27**, 6373 (1985)

23.178 R. Garcìa Cantù, M.A. Huerta Garnica: Long-scan imaging by STM, J. Vac. Sci. Technol. A **8**, 354 (1990)

23.179 C.J. Chen: In situ testing and calibration of tube piezoelectric scanners, Ultramicroscopy **42–44**, 1653–1658 (1992)

23.180 R.G. Carr: Finite element analysis of PZT tube scanner motion for scanning tunnelling microscopy, J. Microsc. **152**, 379–385 (1988)

23.181 C.J. Chen: Electromechanical deflections of piezoelectric tubes with quartered electrodes, Appl. Phys. Lett. **60**, 132 (1992)

23.182 N. Libioulle, A. Ronda, M. Taborelli, J.M. Gilles: Deformations and nonlinearity in scanning tunneling microscope images, J. Vac. Sci. Technol. B **9**, 655–658 (1991)

23.183 E.P. Stoll: Restoration of STM images distorted by time-dependent piezo driver aftereffects, Ultramicroscopy **42–44**, 1585–1589 (1991)

23.184 R. Durselen, U. Grunewald, W. Preuss: Calibration and applications of a high precision piezo scanner for nanometrology, Scanning **17**, 91–96 (1995)

23.185 J. Fu: In situ testing and calibrating of Z-piezo of an atomic force microscope, Rev. Sci. Instrum. **66**, 3785–3788 (1995)

23.186 R.C. Barrett, C.F. Quate: Optical scan-correction system applied to atomic force microscopy, Rev. Sci. Instrum. **62**, 1393 (1991)

23.187 R. Toledo-Crow, P.C. Yang, Y. Chen, M. Vaez-Iravani: Near-field differential scanning optical microscope with atomic force regulation, Appl. Phys. Lett. **60**, 2957–2959 (1992)

23.188 J.E. Griffith, G.L. Miller, C.A. Green: A scanning tunneling microscope with a capacitance-based position monitor, J. Vac. Sci. Technol. B **8**, 2023–2027 (1990)

23.189 A.E. Holman, C.D. Laman, P.M.L.O. Scholte, W.C. Heerens, F. Tuinstra: A calibrated scanning tunneling microscope equipped with capacitive sensors, Rev. Sci. Instrum. **67**, 2274–2280 (1996)

24. Low-Temperature Scanning Probe Microscopy

Mehmet Z. Baykara, Markus Morgenstern, Alexander Schwarz, Udo D. Schwarz

This chapter is dedicated to scanning probe microscopy (SPM) operated at cryogenic temperatures, where the more fundamental aspects of phenomena important in the fields of nanoscience and nanotechnology can be investigated with high sensitivity under well-defined conditions. In general, scanning probe techniques allow the measurement of physical properties down to the nanometer scale. Some techniques, such as scanning tunneling microscopy (STM) and scanning force microscopy (SFM), even go down to the atomic scale. Various properties are accessible. Most importantly, one can image the arrangement of atoms on conducting surfaces by STM and on insulating samples by SFM. However, electronic states (scanning tunneling spectroscopy), force interaction between different atoms (scanning force spectroscopy), magnetic domains (magnetic force microscopy), magnetic exchange interactions (magnetic exchange force microscopy and spectroscopy), local capacitance (scanning capacitance microscopy), local contact potential differences (Kelvin probe force microscopy), local temperature (scanning thermal microscopy), and local light-induced excitations (scanning near-field microscopy) can also be measured with high spatial resolution, among others. In addition, some modern techniques even allow the controlled manipulation of individual atoms/molecules and the visualization of the internal structure of individual molecules. Moreover, combined STM/SFM experiments are now possible, mainly thanks to the advent of tuning forks as sensing elements in low-temperature (LT) SPM systems.

Probably the most important advantage associated with the low-temperature operation of scanning probes is that it leads to a significantly better signal-to-noise ratio than measuring at room temperature. This is why many researchers work below 100 K. However, there are also physical reasons to use low-temperature equipment. For example, visualizing the internal structure of molecules with SFM or the utilization of scanning tunneling spectroscopy with high energy resolution can only be realized at low temperatures. Moreover, some physical effects such as superconductivity or the Kondo effect are restricted to low temperatures. Here, we describe the advantages of low-temperature scanning probe operation and the basics of related instrumentation. Additionally, some of the important results achieved by low-temperature scanning probe microscopy are summarized. We first focus on the STM, giving examples of atomic manipulation and the analysis of electronic properties in different material arrangements, among others. Afterwards, we describe results obtained by SFM, reporting on atomic-scale and submolecular imaging, as well as three-dimensional (3-D) force spectroscopy. Results obtained with the method of Kelvin probe force microscopy (KPFM) that is used to study variations in local contact potential difference (LCPD) are briefly discussed. Magnetic force microscopy (MFM), magnetic exchange force microscopy (MExFM), and magnetic resonance force microscopy (MRFM) are also introduced. Although the presented selection of results is far from complete, we feel that it gives an adequate impression of the fascinating possibilities of low-temperature scanning probe instruments.

In this chapter low temperatures are defined as lower than about 100 K and are normally achieved by cooling with liquid nitrogen or liquid helium. Applications in which SPMs are operated close to 0 °C are not covered in this chapter.

24.1 **Microscope Operation at Low Temperatures** 771
24.1.1 Drift 771
24.1.2 Noise 771
24.1.3 Stability 771
24.1.4 Piezo Relaxation and Hysteresis 771
24.2 **Instrumentation** 772
24.3 **Scanning Tunneling Microscopy and Spectroscopy** 773
24.3.1 Atomic Manipulation 774
24.3.2 High-Resolution Spectroscopy 775
24.3.3 Imaging Electronic Wave Functions 779

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_24

24.3.4 Imaging Spin Polarization: Nanomagnetism 786
24.4 **Scanning Force Microscopy and Spectroscopy** 788
24.4.1 Atomic-Scale and Intramolecular Imaging 789
24.4.2 Force Spectroscopy 790
24.4.3 Atomic and Molecular Manipulation 793
24.4.4 Kelvin Probe Force Microscopy 794
24.4.5 Magnetic Force Microscopy 795
24.4.6 Magnetic Exchange Force Microscopy and Spectroscopy 797
24.4.7 Magnetic Resonance Force Microscopy 798
24.5 **Summary** 799
References 799

Three decades ago, the first design of an experimental setup was presented where a sharp tip was systematically scanned over a sample surface in order to acquire local information on the tip–sample interaction down to the atomic scale. This original instrument used the tunneling current between a conducting tip and a conducting sample as a feedback signal and was thus named the *scanning tunneling microscope* [24.1]. Soon after this historic breakthrough, it became widely recognized that virtually any type of tip–sample interaction could be used to obtain local information on the sample by applying the same general principle, provided that the selected interaction was reasonably short-ranged. Thus, a whole variety of new methods has been introduced, which are denoted collectively as *scanning probe methods*. Overviews are provided in [24.2] and more recently [24.3].

The various methods, especially the above-mentioned scanning tunneling microscopy (STM) and scanning force microscopy (SFM) – which is often further classified into subdisciplines such as topography-reflecting atomic force microscopy (AFM), Kelvin probe force microscopy (KPFM), or magnetic force microscopy (MFM) – have been established as standard methods for surface characterization on the nanometer scale. The reason is that they feature extremely high resolution (often down to the atomic scale for STM and AFM), despite a principally simple, compact, and comparatively inexpensive design.

An important advantage of the simple working principle and the compact design of many scanning probe microscopes (SPMs) is that they can be adapted to different environments such as air, all kinds of gaseous atmospheres, liquids, or vacuum with reasonable effort. Another advantage is their ability to work within a wide temperature range. Microscope operation at higher temperatures is chosen to study surface diffusion, surface reactivity, surface reconstructions that only manifest at elevated temperatures, high-temperature phase transitions, or to simulate conditions as they occur, e.g., in engines, catalytic converters or reactors. Ultimately, the upper limit for the operation of an SPM is determined by the stability of the sample. However, thermal drift, which limits the ability to move the tip in a controlled manner with respect to the sample, as well as the depolarization temperature of the piezoelectric positioning elements might further restrict successful measurements.

Conversely, low-temperature (LT) application of SPMs is much more widespread than operation at high temperatures. Essentially five reasons make researchers adapt their experimental setups to low-temperature compatibility. These are:

1. Reduced thermal drift
2. Lower noise levels
3. Enhanced stability of tip and sample
4. Reduction in piezo hysteresis/creep
5. Probably the most obvious, the fact that many physical effects are restricted to low temperature.

Reasons 1–4 only apply unconditionally if the whole microscope body is kept at low temperature (typically by employing a bath cryostat, Sect. 24.2). Conversely, setups in which only the sample or the tip is cooled (typically those that employ flow cryostats) may show considerably less favorable operating characteristics. As a result of 1–4, ultrahigh resolution and long-term stability can be achieved on a level that significantly exceeds what can be accomplished at room temperature even under the most favorable circumstances. Typical examples of (5) are superconductivity [24.4] and the Kondo effect [24.5].

This chapter is organized such that the advantages associated with low-temperature microscope operation are introduced first, followed by an overview of instrumentational aspects. Subsequently, certain exemplary results of low-temperature STM and SFM operation are presented and the chapter is concluded with a summary.

24.1 Microscope Operation at Low Temperatures

24.1.1 Drift

Thermal drift originates from thermally activated movements of individual atoms, which are reflected by the thermal expansion coefficient. At room temperature, typical values for solids are in the order of $(1-50) \times 10^{-6}\,\mathrm{K}^{-1}$. If the temperature could be kept precisely constant, any thermal drift would vanish, regardless of the absolute temperature of the system. The close coupling of the microscope to a large temperature bath that maintains a constant temperature therefore ensures a significant reduction in thermal drift and allows for distortion-free long-term measurements. As such, microscopes that are efficiently attached to sufficiently large bath cryostats show a one- to two-order-of-magnitude increase in thermal stability compared with nonstabilized setups operated at room temperature.

A second effect also helps suppress thermally induced drift of the probing tip relative to a specific location on the sample surface: The thermal expansion coefficients of materials at liquid-helium temperatures are two or more orders of magnitude smaller than at room temperature. Consequently, thermal drift during low-temperature operation decreases significantly.

For some specific scanning probe methods, there may be additional ways in which a change in temperature can affect the quality of the data. In *frequency-modulation SFM* (FM-SFM), for example, the measurement principle relies on the accurate determination of the eigenfrequency of the cantilever, which is determined by its spring constant and its effective mass. However, the spring constant changes with temperature due to both thermal expansion (i. e., the resulting change in cantilever dimensions) and the variation of the Young's modulus with temperature. Assuming temperature drift rates of about 2 mK/min, as is typical for room-temperature measurements, this effect might have a significant influence on the obtained data.

It should be mentioned that software-based correction of thermal drift to a certain extent in SPM experiments is now a possibility [24.6, 7]. Despite this fact, low-temperature operation currently provides the most direct and reliable route to the elimination of thermal-drift-related effects.

24.1.2 Noise

The theoretically achievable resolution in SPM often increases with decreasing temperature due to a decrease in thermally induced noise. An example is the thermal noise in SFM, which is proportional to the square root of the temperature [24.8, 9]. Lowering the temperature from $T = 300$ to 10 K thus results in a reduction of the thermal frequency noise by more than a factor of five, enabling, e.g., three-dimensional force fields to be acquired with atomic resolution on low-corrugation surfaces such as graphite [24.10].

Another, even more striking, example is the spectroscopic resolution in *scanning tunneling spectroscopy* (STS). This depends linearly on the temperature [24.2] and is consequently reduced even more at LT than the thermal noise in AFM. This provides the opportunity to study structures or physical effects not accessible at room temperature such as Landau levels in semiconductors [24.11].

Finally, it might be worth mentioning that the enhanced stiffness of most materials at low temperatures (increased Young's modulus) leads to a reduced coupling to external noise. Even though this effect is considered small [24.9], it should not be ignored, especially when combined with the fact that artifacts in atomic-resolution SPM experiments due to tip elasticity [24.12, 13] are also expected to decrease at low temperatures due to the increase in stiffness.

24.1.3 Stability

There are two major stability issues that considerably improve at low temperature. First, low temperatures close to the temperature of liquid helium inhibit most of the thermally activated diffusion processes. As a consequence, the sample surfaces show a significantly increased long-term stability since defect motion and/or adatom diffusion is massively suppressed. Most strikingly, even single xenon atoms deposited on suitable substrates can be successfully imaged [24.14, 15] or even manipulated [24.16]. In the same way, low temperatures also stabilize the atomic configuration at the tip end by preventing sudden jumps of the most loosely bound, foremost tip atom(s). Moreover, the large cryostat that usually surrounds the microscope acts as an effective cryo-pump. Thus, samples can be kept clean for several weeks, which is a multiple of the corresponding time at room temperature (about 3–4 h).

24.1.4 Piezo Relaxation and Hysteresis

The last important benefit of low-temperature operation of SPMs is that artifacts from the response of the piezoelectric scanners are substantially reduced. After applying a voltage ramp to one electrode of a piezoelectric scanner, its immediate initial deflection, l_0, is followed by a much slower relaxation, Δl, with a loga-

rithmic time dependence. This effect, known as piezo relaxation or *creep*, diminishes substantially at low temperatures, typically by a factor of ten or more. As a consequence, piezo nonlinearities and piezo hysteresis decrease accordingly. Additional information is given in [24.17].

24.2 Instrumentation

The two main design criteria for all ultrahigh vacuum (UHV)-based SPM systems (which we are exclusively focusing on) are: (1) to provide an efficient decoupling of the microscope from the UHV system and other sources of external vibrations, and (2) to avoid most internal noise sources through the high mechanical rigidity of the microscope body itself. In vacuum systems designed for low-temperature applications, a significant degree of complexity is added, since, on the one hand, close thermal contact of the SPM and cryogen is necessary to ensure the (nearly) drift-free conditions described above, while, on the other hand, good vibration isolation (both from the outside world, as well as from the boiling or flowing cryogen) has to be maintained.

During the last couple of decades, a considerable number of both home-built and commercial SPM designs have been presented for low-temperature operation. Because of the variety of different approaches, it is not possible to cover all related aspects here. Instead, we will briefly discuss a home-built combined STM/SFM system that is specifically geared toward atomic-resolution imaging and spectroscopy at low temperatures that relies on a bath cryostat [24.18] rather than a flow cryostat. The main advantage of using bath cryostats instead of flow cryostats for low-temperature operation is the fact that the *whole microscope* is kept at the same cryogenic temperature which thus (nearly) eliminates the occurrence of thermal drift between the tip and the sample.

Drawings of the (i) UHV system that houses the specific microscope to be discussed here, (ii) the cryostat to which the microscope is attached, and (iii) the microscope itself, are provided in Fig. 24.1. The UHV system (Fig. 24.1a) is partially based on a commercial design by *Omicron Nanotechnology* (Germany) and features two individual chambers for sample preparation (via sputtering, annealing, oxygen plasma deposition, etc.) and surface analysis (via conventional surface science techniques such as low-energy electron diffraction (LEED) and Auger spectroscopy) as well as a third chamber (the SPM chamber) that houses the microscope attached to the on-top bath cryostat. Samples are introduced into the UHV system via a fast entry load lock pumped by a turbo pump and sample transfer between different chambers occurs in situ via magnetic transfer arms. All chambers are equipped with ion getter and titanium sublimation pumps that allow base pressures on the order of 10^{-11} mbar to be reached. Additionally, the SPM and preparation chambers are equipped with turbo pumps for the bakeout process and pump-down from ambient pressure. For vibration isolation purposes, the whole UHV system is situated on 12 active vibration isolation units in a sound-proofed room on a basement floor.

The bath cryostat employed in the present design (Fig. 24.1b) has been fabricated by *Cryovac* and consists of a central dewar for liquid helium with a capacity of 8.5 l and a *shielding* dewar that can contain up to 18 l of liquid nitrogen. This setup allows low liquid helium consumption ($\approx 3\,\mathrm{l/day}$). The cryostat features, on the vacuum chamber side, a double set of thermal radiation shields (gold-coated copper) that are cooled by the liquid helium and liquid nitrogen dewars, respectively. The microscope is situated inside the thermal shields and may be suspended on a set of three springs for vibrational isolation from the cryostat during measurements. Thanks to a set of sliding doors on the shields, sample and tip transfer occur in situ (as the microscope is cold and under vacuum) and cool-down times after transfer are on the order of 2 h. During operation, the whole microscope is at a typical temperature of $\approx 5\,\mathrm{K}$.

The microscope itself (Fig. 24.1c) allows combined STM/SFM operation thanks to the fact that it employs quartz tuning forks rather than the conventional micromachined SFM cantilevers. While the particular details associated with using tuning forks as sensing elements for atomic-resolution SFM (often combined with STM) are explained elsewhere [24.19], the main advantages involve (i) *self-sensing* thanks to the piezoelectric character of the quartz and (ii) enhanced sensitivity to short-range forces due to high mechanical stiffness. The main body of the microscope is manufactured out of Macor and is cylindrical with a height of 80 mm and a diameter of 40 mm. To minimize thermal drift, a (mostly) symmetric structural design has been employed and most metallic parts are made out of titanium which has a similar thermal expansion coefficient to Macor. In this particular design, the tip is stationary and for the *coarse* approach, the sample is moved vertically with respect to the tip by using a *pan-style* motor that is based on a sapphire prism in point-contact with several shear-piezo-stacks (see [24.18] for details).

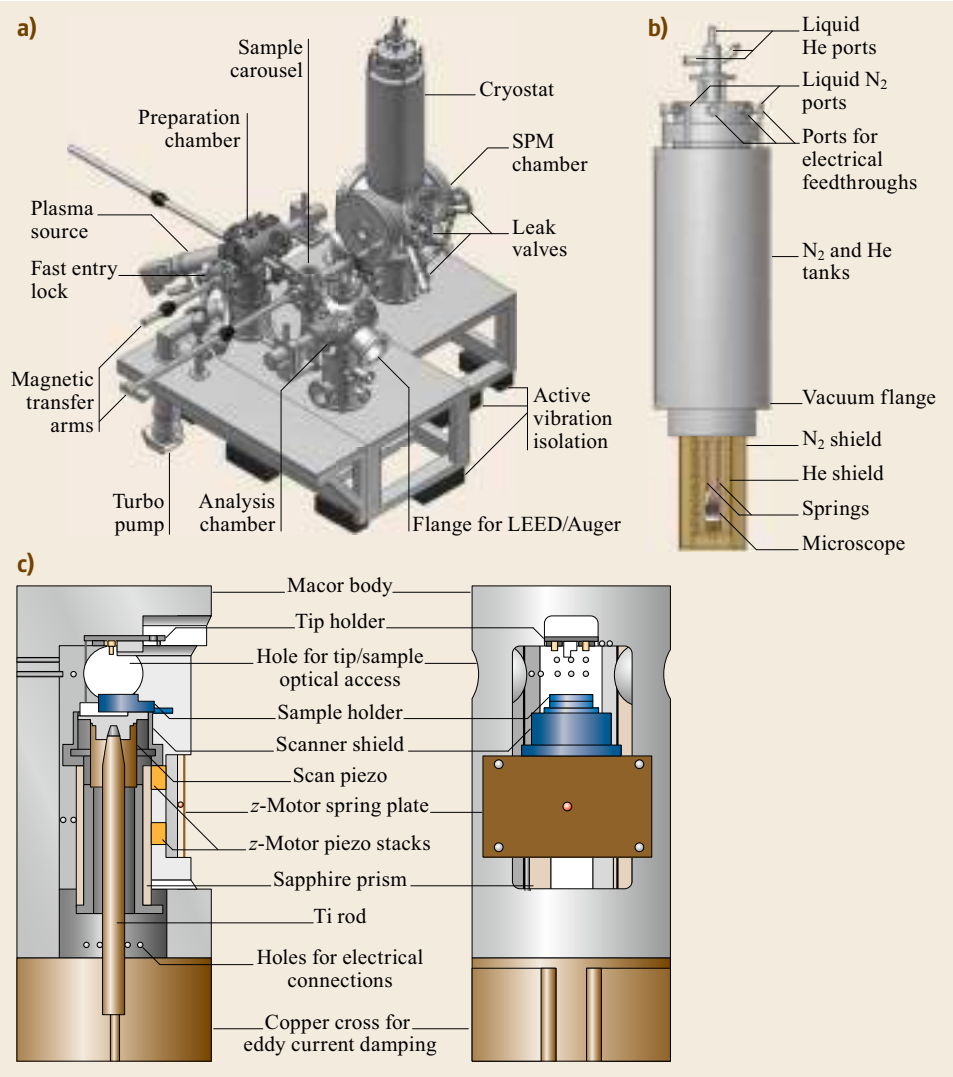


Fig. 24.1 Schematic drawings of the **(a)** UHV system that houses the low-temperature microscope described in this chapter, **(b)** the cryostat to which the microscope is attached and **(c)** the microscope itself (*left*: section view, *right*: front view). Adapted from [24.18]

For precise three-dimensional positioning of the sample with respect to the tip during the *fine* approach and data acquisition, a piezoelectric element in the shape of a cylinder (the *scan piezo*) is employed.

To conclude this section, one should re-iterate that careful instrumentational design for low-temperature scanning probe microscopy allows imaging and spectroscopy experiments with ultrahigh resolution and stability to be successfully performed. In particular, the 3-D force spectroscopy data on surface-oxidized copper and titanium dioxide surfaces discussed in Sect. 24.4 have been obtained with the specific microscope presented here. Finally, rapidly increasing prices for liquid helium have recently led to the development of cryogen-free designs for low-temperature SPM with promising results (e.g., by *RHK technology*). It is projected that this trend in SPM instrumentation will accelerate in the near future.

24.3 Scanning Tunneling Microscopy and Spectroscopy

In this section, we review some of the most important results achieved by low-temperature scanning tunneling microscopy (LT-STM). After summarizing the results, placing emphasis on the necessity of LT equipment, we turn to the details of the different experiments and the physical meaning of the results obtained.

As described in Sect. 24.1, the LT equipment has basically *three* advantages for scanning tunneling microscopy (STM) and spectroscopy (STS): First, the instruments are much more stable with respect to thermal drift and the coupling to external noise, allowing the establishment of new functionalities for the instrument. In particular, LT-STM has been used to move atoms on a surface [24.16], cut molecules into pieces [24.20], reform bonds [24.21], charge individual atoms [24.22], and, consequently, establish new structures on the nanometer scale. Also, the detection of light resulting from tunneling into a particular molecule [24.23, 24], the visualization of thermally induced atomic movements [24.25], and the detection of hysteresis curves of individual atoms [24.26] require LT instrumentation. More recent accomplishments achieved with LT-STM include the measurement of electron spin relaxation times of individual atoms [24.27] and the detection of magnetic remanence in single atoms [24.28].

Second, the spectroscopic resolution in STS depends linearly on temperature and is, therefore, considerably reduced at LT. This provides the opportunity to study physical effects inaccessible at room temperature. Examples are the resolution of spin and Landau levels in semiconductors [24.11], or the investigation of lifetime-broadening effects on the nanometer scale [24.29]. Also the imaging of distinct electronic wavefunctions in real space requires LT-STM [24.30]. More recently, vibrational levels, spin-flip excitations and phonons have been detected with high spatial resolution at LT using the additional inelastic tunneling channel [24.31–33].

Third, many physical effects, in particular effects guided by electronic correlations, are restricted to low temperature. Typical examples are superconductivity [24.4], the Kondo effect [24.5], and many of the electron phases found in semiconductors [24.34]. Here, LT-STM provides the possibility to study electronic effects on a local scale, and intensive work has been done in this field, the most elaborate with respect to high-temperature superconductivity [24.35–37].

24.3.1 Atomic Manipulation

Although manipulation of surfaces on the atomic scale can be achieved at room temperature [24.38, 39], only the use of LT-STM allows the placement of individual atoms at desired atomic positions [24.40]. The main reason is that rotation, diffusion, or charge transfer of entities could be excited at higher temperature, making the intentionally produced configurations unstable.

The usual technique to manipulate atoms is to increase the current above a certain atom, which reduces the tip–atom distance, then to move the tip with the atom to a desired position and finally to reduce the current again in order to decouple atom and tip. The first demonstration of this technique was performed by *Eigler* and *Schweizer* [24.16], who used Xe atoms on a Ni(110) surface to write the three letters *IBM* (their employer) on the atomic scale (Fig. 24.2a). Nowadays, many laboratories are able to move different kinds of atoms and molecules on different surfaces with high precision. An example featuring CO molecules on Cu(110) is shown in Fig. 24.2b–g. Even more complex structures than the "2000" shown in the figure, such as cascades of CO molecules that by mutual repulsive interaction mimic different kinds of logic gates, have been assembled and their functionality tested [24.41]. Although these devices are slow and restricted to low temperature, they nicely demonstrate the high degree of control achieved on the atomic scale.

The basic modes of controlled motion of atoms and molecules by the tip are pushing, pulling, and sliding. The selection of the particular mode depends on the tunneling current, i. e., the distance between tip and molecule, as well as on the particular molecule–substrate combination [24.42]. It has been shown experimentally that the potential landscape for the adsorbate movement is modified by the presence of the tip [24.43, 44] and that excitations induced by the tunneling current can trigger atomic or molecular motion [24.45, 46]. Other sources of motion are the electric field between tip and molecule or electromigration caused by the high current density [24.40]. The required lateral tip force for atomic motion has been measured for typical adsorbate–substrate combinations to be on the order of 0.1 nN [24.47]. Other types of manipulation on the atomic scale are feasible. Some of them require selective inelastic tunneling into vibrational or rotational modes of the molecules [24.48]. This leads to controlled desorption [24.49], diffusion [24.50], molecular rotation [24.51, 52], conformational change [24.53], or even the controlled pick-up of molecules by the tip [24.21]. Dissociation can be achieved by voltage pulses [24.20] inducing local heating, even if the pulse is applied at distances of 100 nm away from the molecule [24.54]. Also, association of individual molecules [24.21, 55–57] can require voltage pulses in order to overcome local energy barriers. The process of controlled bond formation can even be used for doping of single C_{60} molecules by up to four potassium atoms [24.58]. As an example of controlled manipulation, Fig. 24.2h–m shows the production of biphenyl from two iodobenzene molecules [24.59]. The iodine is abstracted by voltage pulses (Fig. 24.2i,j), then the iodine is moved to the terrace by the pulling mode (Fig. 24.2k,l), and finally the two phenyl parts are slid along the step edge until they are close enough to react (Fig. 24.2m). The

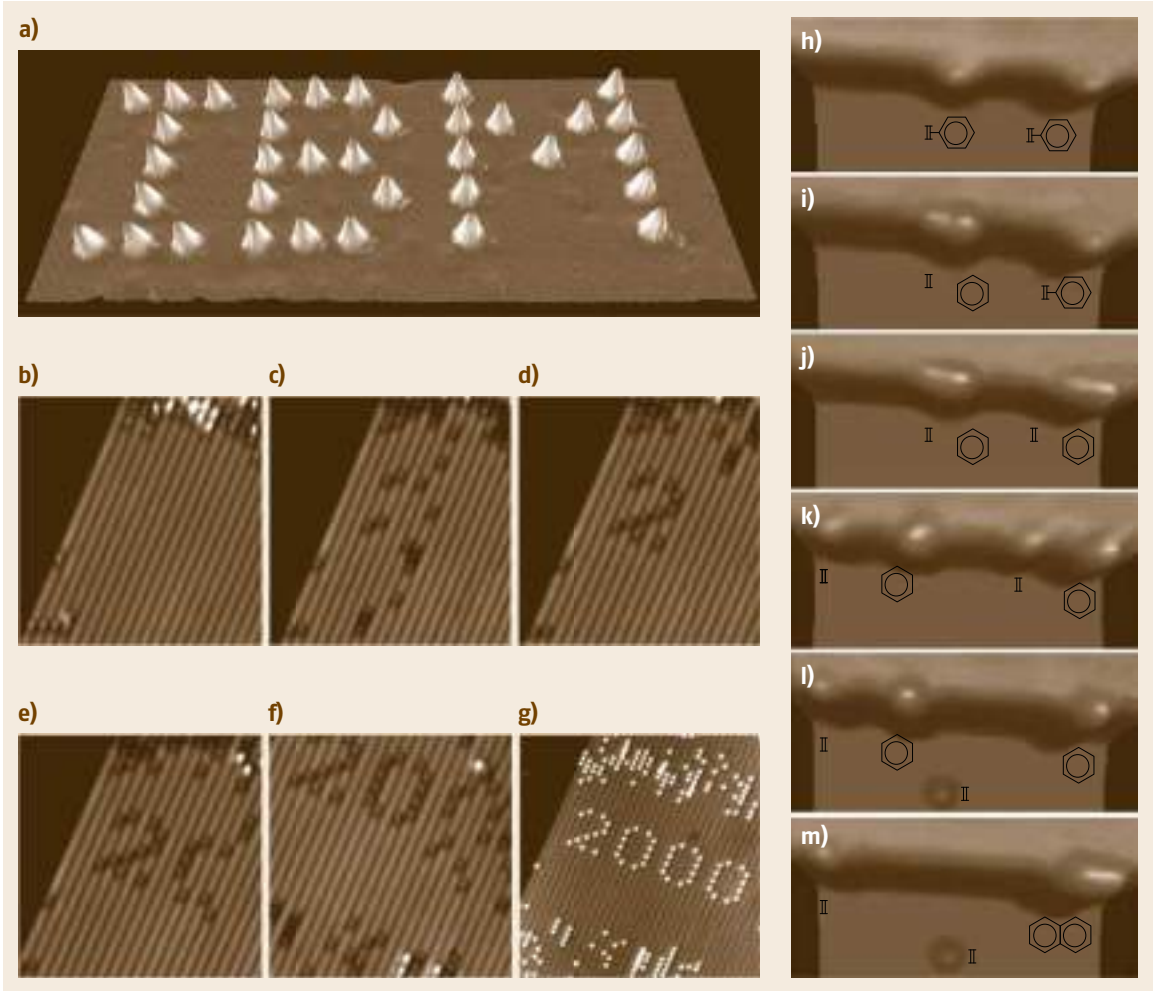


Fig. 24.2 **(a)** STM image of single Xe atoms positioned on a Ni(110) surface in order to realize the letters *IBM* on the atomic scale (© D. Eigler, Almaden); **(b–f)** STM images recorded after different positioning processes of CO molecules on a Cu(110) surface; **(g)** final artwork greeting the new millennium on the atomic scale (**(b–g)** © G. Meyer, Zürich). **(h–m)** Synthesis of biphenyl from two iodobenzene molecules on Cu(111): First, iodine is abstracted from both molecules **(i,j)**; then the iodine between the two phenyl groups is removed from the step **(k)**, and finally one of the phenyls is slid along the Cu step **(l)** until it reacts with the other phenyl **(m)**; the line drawings symbolize the actual status of the molecules (**(h–m)** © S. W. Hla and K. H. Rieder, Berlin)

chemical identification of the product is not straightforward and partly requires detailed vibrational STM spectroscopy [24.60].

Finally, also the charge state of a single atom or molecule can be manipulated, tested, and read out. A Au atom has been switched reversibly between two charge states using an insulating thin film as the substrate [24.22]. In addition, the carrier capture rate of a single impurity level within the bandgap of a semiconductor has been quantified [24.61], and the point conductance of a single atom has been measured and turned out to be a reproducible quantity [24.62]. These promising results might trigger a novel electronic field of manipulation of matter on the atomic scale, which is tightly related to the currently very popular field of molecular electronics.

24.3.2 High-Resolution Spectroscopy

One of the most important modes of LT-STM is STS, which detects the differential conductivity $\mathrm{d}I/\mathrm{d}V$ as a function of the applied voltage V and the position (x, y). The $\mathrm{d}I/\mathrm{d}V$ signal is basically proportional to the local density of states (LDOS) of the sample, the sum

over squared single particle wavefunctions Ψ_i [24.2],

$$\frac{\mathrm{d}I}{\mathrm{d}V}(V,x,y) \propto \mathrm{LDOS}\,(E,x,y) = \sum_{\Delta E} |\Psi_i(E,x,y)|^2\,, \tag{24.1}$$

where ΔE is the energy resolution of the experiment. In simple terms, each state corresponds to a tunneling channel, if it is located between the Fermi levels (E_F) of the tip and the sample. Thus, all states located in this energy interval contribute to I, while $\mathrm{d}I/\mathrm{d}V(V)$ detects only the states at the energy E corresponding to V. The local intensity of each channel depends further on the LDOS of the state at the corresponding surface position and its decay length into vacuum. For s-like tip states, *Tersoff* and *Hamann* have shown that it is simply proportional to the LDOS at the position of the tip [24.63]. Therefore, as long as the decay length is spatially constant, one measures the LDOS at the surface (24.1). Note that the contributing states are not only surface states, but also bulk states. However, surface states usually dominate if present. *Chen* has shown that higher orbital tip states lead to the so-called derivation rule [24.64]: p_z-type tip states detect $\mathrm{d(LDOS)}/\mathrm{d}z$, d_z^2-states detect $\mathrm{d}^2(\mathrm{LDOS})/\mathrm{d}z^2$, and so on. As long as the decay into vacuum is exponential and spatially constant, this leads only to an additional, spatially constant factor in $\mathrm{d}I/\mathrm{d}V$. Thus, it is still the LDOS that is measured (24.1). The requirement of a spatially constant decay is usually fulfilled on larger length scales, but not on the atomic scale [24.64]. There, states located close to the atoms show a stronger decay into vacuum than the less localized states in the interstitial region. This effect can lead to STS corrugations that are larger than the real LDOS corrugations [24.65].

The voltage dependence of $\mathrm{d}I/\mathrm{d}V$ is sensitive to a changing decay length with V, which increases with V. This influence can be reduced at higher V by displaying $\mathrm{d}I/\mathrm{d}V/(I/V)$ [24.66]. Additionally, $\mathrm{d}I/\mathrm{d}V(V)$ curves might be influenced by possible structures in the density of states (DOS) of the tip, which also contribute to the number of tunneling channels [24.67]. However, these structures can usually be identified, and only tips free of characteristic DOS structures are used for quantitative experiments.

Importantly, the energy resolution ΔE is largely determined by temperature. It is defined as the smallest energy distance of two δ-peaks in the LDOS that can still be resolved as two individual peaks in $\mathrm{d}I/\mathrm{d}V(V)$ curves and is $\Delta E = 3.3\,k_\mathrm{B}T$ [24.2]. The temperature dependence is nicely demonstrated in Fig. 24.3, where the tunneling gap of the superconductor Nb is measured at different temperatures [24.68]. The peaks at the rim of the gap get wider at temperatures well below the critical temperature of the superconductor ($T_\mathrm{c} = 9.2\,\mathrm{K}$).

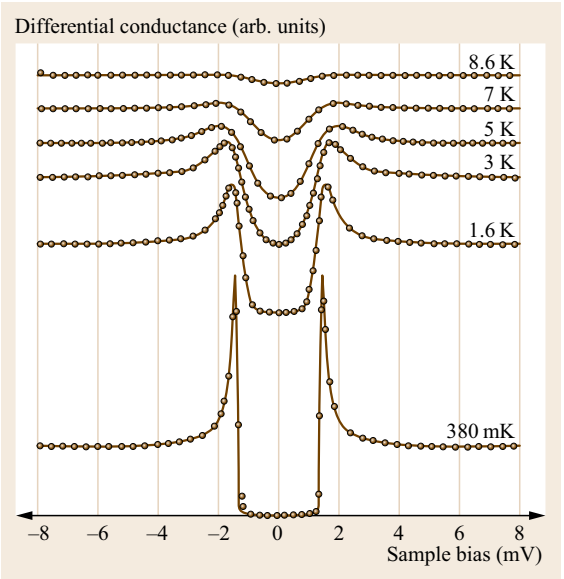


Fig. 24.3 Differential conductivity curve $\mathrm{d}I/\mathrm{d}V(V)$ measured on a Au surface by a Nb tip (*circles*). Different temperatures are indicated; the *lines* are fits according to the superconducting gap of Nb folded with the temperature-broadened Fermi distribution of the Au (© S.H. Pan, Houston)

Lifetime Broadening

Besides ΔE, intrinsic properties of the sample lead to a broadening of spectroscopic features. Basically, the finite lifetime of the electron or hole in the corresponding state broadens its energetic width. Any kind of interaction such as electron–electron interaction can be responsible. Lifetime broadening has usually been measured by photoemission spectroscopy (PES), but it turned out that lifetimes of surface states on noble-metal surfaces determined by STS (Fig. 24.4a,b) are up to a factor of three larger than those measured by PES [24.69]. The reason is probably that defects broaden the PES spectrum. Defects are unavoidable in a spatially integrating technique such as PES, thus STS has the advantage of choosing a particularly clean area for lifetime measurements. The STS results can be successfully compared with theory, highlighting the dominating influence of intraband transitions for the surface-state lifetime on Au(111) and Cu(111), at least close to the onset of the surface band [24.29].

With respect to band electrons, the analysis of the width of the band onset on $\mathrm{d}I/\mathrm{d}V(V)$ curves has the disadvantage of being restricted to the onset energy. Another method circumvents this problem by mea-

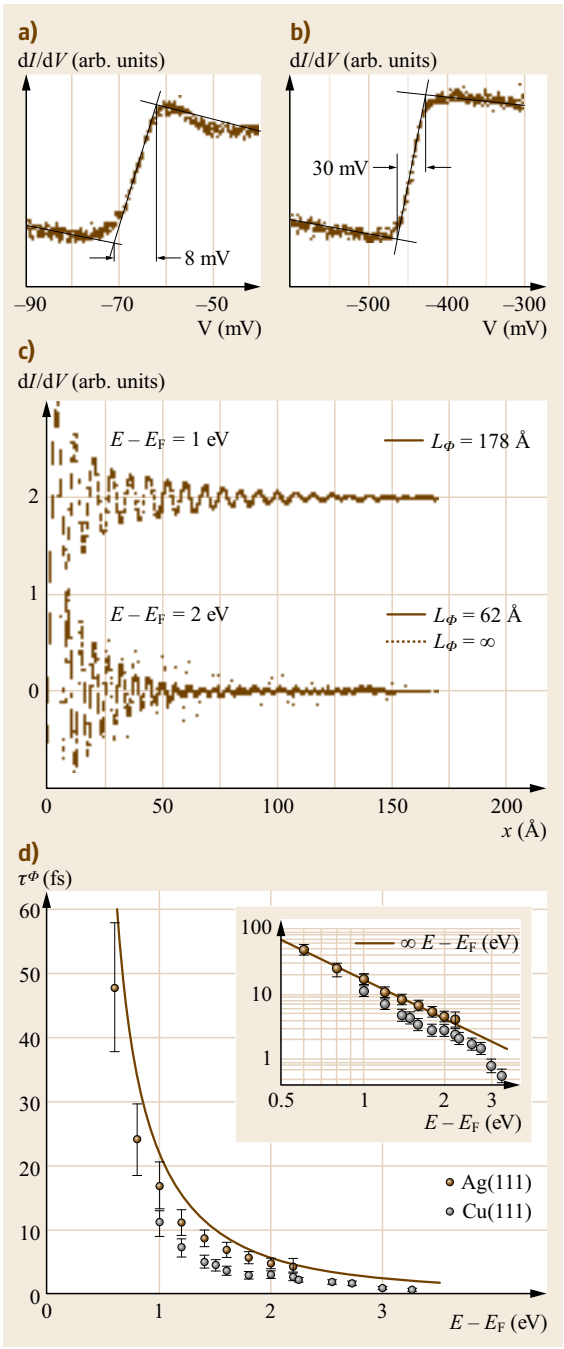


Fig. 24.4 (a,b) Spatially averaged dI/dV curves of Ag(111) and Cu(111); both surfaces exhibit a surface state with parabolic dispersion, starting at −65 and −430 meV, respectively. The *lines* are drawn to determine the energetic width of the onset of these surface bands (**(a,b)** © R. Berndt, Kiel); **(c)** dI/dV intensity as a function of position away from a step edge of Cu(111) measured at the voltages $(E - E_F)$, as indicated (*points*); the *lines* are fits assuming standing electron waves with a phase coherence length L_Φ as marked; **(d)** resulting phase coherence time as a function of energy for Ag(111) and Cu(111). *Inset* shows the same data on a double-logarithmic scale, evidencing the E^{-2} dependence (*line*) (**(c,d)** © H. Brune, Lausanne) ◀

suring the decay of standing electron waves scattered from a step edge as a function of energy [24.70]. Figure 24.4c,d shows the resulting oscillating dI/dV signal measured for two different energies. To deduce the coherence length L_Φ, which is inversely proportional to the lifetime τ_Φ, one has to consider that the finite energy resolution ΔE in the experiment also leads to a decay of the standing wave away from the step edge. The dotted fit line using $L_\Phi = \infty$ indicates this effect and, more importantly, shows a discrepancy from the measured curve. Only including a finite coherence length of 6.2 nm results in good agreement, which in turn determines L_Φ and thus τ_Φ, as displayed in Fig. 24.4c. The determined $1/E^2$ dependence of τ_Φ points to a dominating influence of electron–electron interactions at higher energies in the surface band.

Landau and Spin Levels

Moreover, the increased energy resolution at LT allows the resolution of electronic states that are not resolvable at room temperature (RT); for example, Landau and spin quantization appearing in a magnetic field B have been probed on InAs(110) [24.11, 71]. The corresponding quantization energies are given by $E_{\text{Landau}} = \hbar eB/m_{\text{eff}}$ and $E_{\text{spin}} = g\mu B$. Thus, InAs is a good choice, since it exhibits a low effective mass $m_{\text{eff}}/m_e = 0.023$ and a high g-factor of 14 in the bulk conduction band. The values in metals are $m_{\text{eff}}/m_e \approx 1$ and $g \approx 2$, resulting in energy splittings of only 1.25 and 1.20 meV at $B = 10$ T. This is obviously lower than the typical lifetime broadenings discussed in the previous section and also close to $\Delta E = 1.1$ meV achievable at $T = 4$ K.

Fortunately, the electron density in doped semiconductors is much lower, and thus the lifetime increases significantly. Figure 24.5a shows a set of spectroscopy curves obtained on InAs(110) in different magnetic fields [24.11]. Above E_F, oscillations with increasing intensity and energy distance are observed. They show the separation expected from Landau quantization. In turn, they can be used to deduce m_{eff} from the peak separation (Fig. 24.5b). An increase of m_{eff} with increasing E has been found, as expected from theory. Also, at high fields, spin quantization is observed (Fig. 24.5c). It is larger than expected from the bare g-factor due to contributions from exchange enhancement [24.72].

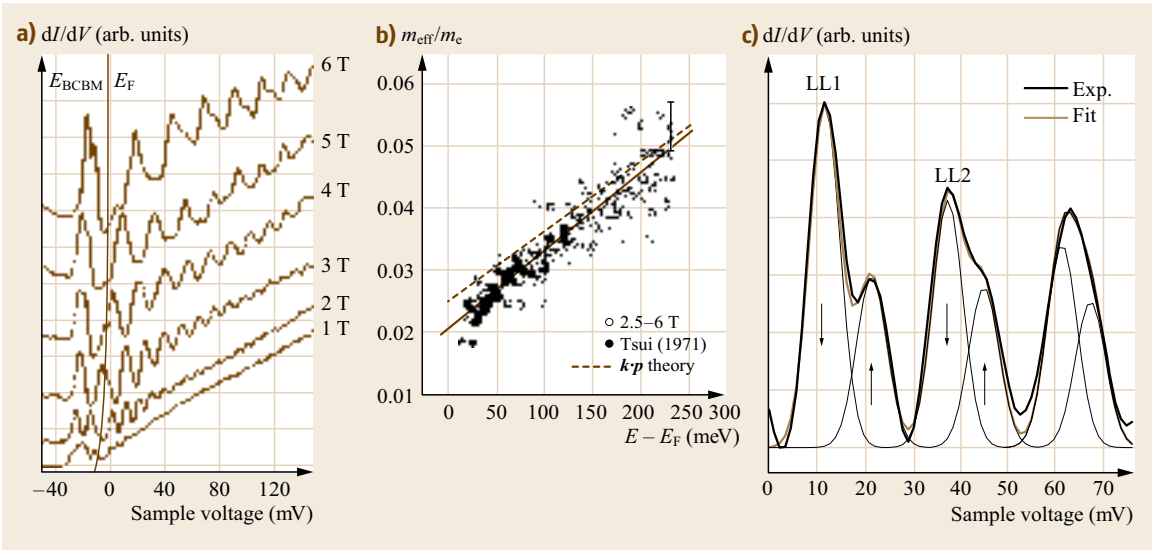


Fig. 24.5 **(a)** dI/dV curves of *n*-InAs(110) at different magnetic fields, as indicated; E_{BCBM} marks the bulk conduction band minimum; oscillations above E_{BCBM} are caused by Landau quantization; the *double peaks* at $B = 6\,T$ are caused by spin quantization. **(b)** Effective-mass data deduced from the distance of adjacent Landau peaks ΔE according to $\Delta E = heB/m_{eff}$ (*open symbols*); *filled symbols* are data from planar tunnel junctions (Tsui), the *solid line* is a mean-square fit of the data and the *dashed line* is the expected effective mass of InAs according to $\boldsymbol{k}\cdot\boldsymbol{p}$ theory. **(c)** Magnification of a dI/dV curve at $B = 6\,T$, exhibiting spin splitting; the Gaussian curves marked by *arrows* are the fitted spin levels

Atomic Energy Levels

Another opportunity at LT is to study electronic states and resonances of single adatoms. A complicated resonance is the Kondo resonance described below. A simpler resonance is a surface state bound at the adatom potential. It appears as a spatially localized peak below the onset of the extended surface state (Fig. 24.4a) [24.73, 74]. A similar resonance caused by a mixing of bulk states of the NiAl(110) substrate with atomic Au levels has been used to detect exchange splitting in Au dimers as a function of interatomic distance [24.75]. Single magnetic adatoms on the same surface also exhibit a double-peak resonance, but here due to the influence of spin-split d-levels of the adsorbate [24.76]. Atomic and molecular states decoupled from the substrate have finally been observed, when the atoms or molecules are deposited on an insulating thin film [24.22, 57].

Vibrational Levels

Inelastic tunneling processes contribute to the tunneling current. The coupling of electronic states to vibrational levels is one source of inelastic tunneling [24.33]. It provides additional channels contributing to $dI/dV(V)$ with final states at energies different from V. The final energy is simply shifted by the energy of the vibrational level. If only discrete vibrational energy levels couple to a smooth electronic DOS, one expects a peak in d^2I/dV^2 at the vibrational energy. This situation appears for molecules on noble-metal surfaces. As usual, the isotope effect can be used to verify the vibrational origin of the peak. First indications of vibrational levels have been found for H_2O and D_2O on TiO_2 [24.77], and completely convincing work has been performed for C_2H_2 and C_2D_2 on Cu(001) [24.33] (Fig. 24.6a). The technique has been used to identify individual molecules on the surface by their characteristic vibrational levels [24.60]. Moreover, the orientation of complexes with respect to the surface can be determined to a certain extent, since the vibrational excitation depends on the position of the tunneling current within the molecule. Finally, the excitation of certain molecular levels can induce such corresponding motions as hopping [24.50], rotation [24.52] (Fig. 24.6b–e), or desorption [24.49] leading to additional possibilities for manipulation on the atomic scale.

In turn, the manipulation efficiency as a function of applied voltage can be used to identify vibrational energies within the molecule, even if they are not detectable directly by d^2I/dV^2 spectroscopy [24.78]. Multiple vibronic excitations are found by positioning the molecule on an insulating film, leading to the observation of equidistant peaks in $d^2I/dV^2(V)$ [24.79].

Kondo Resonance

A rather intricate interaction effect is the Kondo effect. It results from a second-order scattering process be-

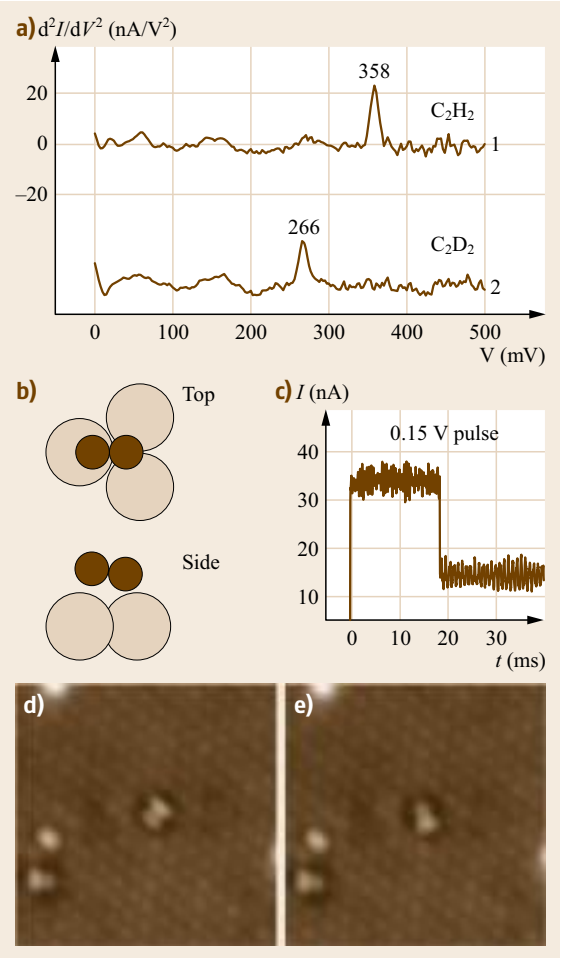


Fig. 24.6 **(a)** d^2I/dV^2 curves taken above a C_2H_2 and a C_2D_2 molecule on Cu(100); the *peaks* correspond to the C−H or C−D stretch-mode energy of the molecule, respectively. **(b)** Sketch of O_2 molecule on Pt(111). **(c)** Tunneling current above an O_2 molecule on Pt(111) during a voltage pulse of 0.15 V; the jump in current indicates rotation of the molecule. **(d,e)** STM images of an O_2 molecule on Pt(111) ($V = 0.05$ V), prior to and after rotation induced by a voltage pulse to 0.15 V (**(a–e)** © W. Ho, Irvine)

tween itinerate states and a localized state [24.80]. The two states exchange some degree of freedom back and forth, leading to a divergence of the scattering probability at the Fermi level of the itinerate state. Because of the divergence, the effect strongly modifies sample properties. For example, it leads to an unexpected increase in resistance with decreasing temperature for metals containing magnetic impurities [24.5]. Here, the exchanged degree of freedom is the spin. A spectroscopic signature of the Kondo effect is a narrow peak in the DOS at the Fermi level, continuously disappearing above a characteristic temperature (the Kondo temperature). STS provides the opportunity to study this effect on the local scale [24.81, 82].

Figure 24.7a–d shows an example of Co clusters deposited on a carbon nanotube [24.83]. While only a small dip at the Fermi level, probably caused by curvature influences on the π-orbitals, is observed without Co (Fig. 24.7b) [24.84], a strong peak is found around a Co cluster deposited on top of the tube (Co cluster is marked in Fig. 24.7a). The peak is slightly shifted with respect to $V = 0$ mV due to the so-called Fano resonance [24.85] which results from interference of the tunneling processes into the localized Co level and the itinerant nanotube levels. The resonance disappears within several nanometers of the cluster, as shown in Fig. 24.7d.

The Kondo effect has also been detected for different magnetic atoms deposited on noble-metal surfaces [24.81, 82]. There, it disappears at about 1 nm from the magnetic impurity, and the effect of the Fano resonance is more pronounced, contributing to dips in $dI/dV(V)$ curves instead of peaks. Detailed investigations show that the d-level occupation of the adsorbate [24.86] as well as the surface charge density [24.87, 88] matter for the Kondo temperature. Exchange interaction between adsorbates tunable by their mutual distance can be used to tune the Kondo temperature [24.89] or even to destroy the Kondo resonance completely [24.90]. Meanwhile, magnetic molecules have also been shown to exhibit Kondo resonances. This increases the tunability of the Kondo effect, e.g., by the selection of adequate ligands surrounding the localized spins [24.91, 92], by distant association of other molecules [24.93], or by conformational changes within the molecule [24.94].

A fascinating experiment has been performed by *Manoharan* et al. [24.95], who used manipulation to form an elliptic cage for the surface states of Cu(111) (Fig. 24.7e, *bottom part*). This cage was constructed to have a quantized level at E_F. Then, a Co atom was placed in one focus of the elliptic cage, producing a Kondo resonance. Surprisingly, the same resonance reappeared in the opposite focus, but not away from the focus (Fig. 24.7e, *top part*). This shows amazingly that complex local effects such as the Kondo resonance can be wave-guided to remote points.

24.3.3 Imaging Electronic Wave Functions

Bloch Waves

Since STS measures the sum of squared wavefunctions (24.1), it is an obvious task to measure the local appearance of the simplest wavefunctions in solids,

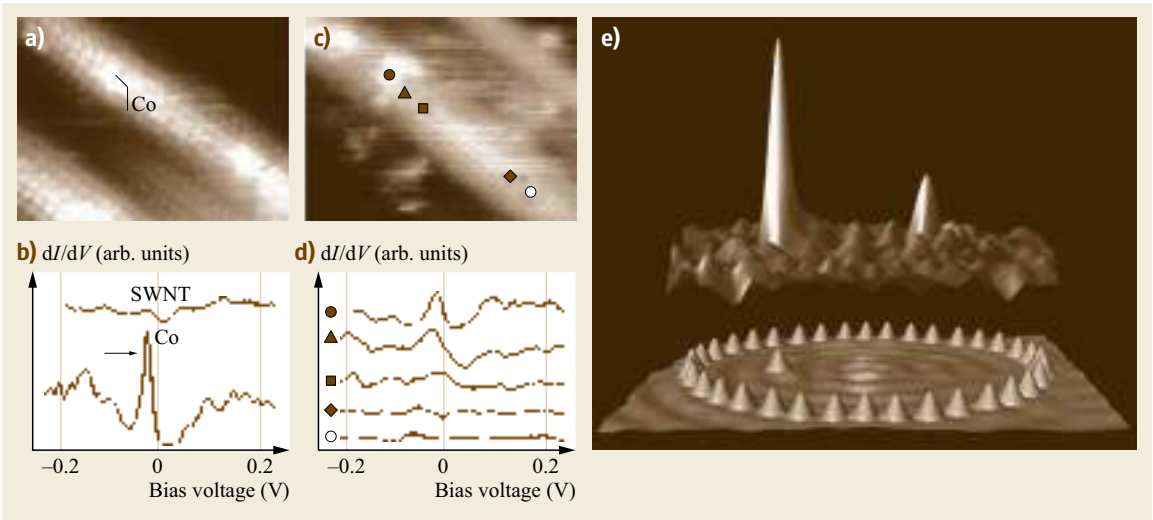


Fig. 24.7 **(a)** STM image of a Co cluster on a single-wall carbon nanotube (SWNT). **(b)** dI/dV curves taken directly above the Co cluster (Co) and far away from the Co cluster (SWNT); the *arrow* marks the Kondo peak. **(c)** STM image of another Co cluster on a SWNT with *symbols* marking the positions where the dI/dV curves displayed in **(d)** are taken. **(d)** dI/dV curves taken at the positions marked in **(c)** (**(a–d)** © C. Lieber, Cambridge). **(e)** *Lower part*: STM image of a quantum corral of elliptic shape made from Co atoms on Cu(111); one Co atom is placed at one of the foci of the ellipse. *Upper part*: map of the strength of the Kondo signal in the corral; note that there is also a Kondo signal at the focus that is not covered by a Co atom (**(e)** © D. Eigler, Almaden)

namely Bloch waves. The atomically periodic part of the Bloch wave is always measured if atomic resolution is achieved (*inset* of Fig. 24.9a). However, the long-range wavy part requires the presence of scatterers. The electron wave impinges on the scatterer and is reflected, leading to self-interference. In other words, the phase of the Bloch wave becomes fixed by the scatterer.

Such self-interference patterns were first found on graphite (0001) [24.96] and later on noble-metal surfaces, where adsorbates or step edges scatter the surface states (Fig. 24.8a) [24.30]. Fourier transforms of the real-space images reveal the k-space distribution of the corresponding states [24.97], which may include additional contributions besides the surface state [24.98]. Using particular geometries such as so-called quantum corrals, the Bloch waves can be confined (Fig. 24.8b). Depending on the geometry of the corral, the resulting state looks rather complex, but it can usually be reproduced by simple calculations involving single-particle states only [24.99].

Meanwhile, Bloch waves in semiconductors scattered at charged dopants (Fig. 24.8c,d) [24.100], Bloch states confined in semiconducting or organic quantum dots (Fig. 24.8e–g) [24.101–103] and quantum wells [24.104], as well as Bloch waves confined in short-cut carbon nanotubes (Fig. 24.8h,i) [24.105, 106] have been visualized. In special nanostructures, it was even possible to extract the phase of the wavefunction by using the mathematically known transformation matrices of so-called isospectral structures, i. e., geometrically different structures exhibiting exactly the same spatially averaged density of states. The resulting wavefunctions $\Psi(x)$ are shown in Fig. 24.8j [24.107].

More localized structures, where a Bloch wave description is not appropriate, have been imaged, too. Examples are the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of pentacene molecules deposited on NaCl/Cu(100) (Fig. 24.8k,l) [24.57], the different molecular states of C_{60} on Ag(110) (Fig. 24.8m–o) [24.108], the anisotropic states of Mn acceptors in a semiconducting host [24.109, 110], and the hybridized states developing within short monoatomic Au chains, which result in particular states at the end of the chains [24.111, 112]. Using pairs of remote Mn acceptors, even symmetric and antisymmetric pair wavefunctions have been imaged in real space [24.113].

The central requirements for a detailed imaging of wavefunctions are LT for an appropriate energetic distinction of an individual state, adequate decoupling of the state from the substrate in order to decrease lifetime-induced broadening effects, and, partly, the selection of a system with an increased Bohr radius in order to increase the spatial extension of details above the lat-

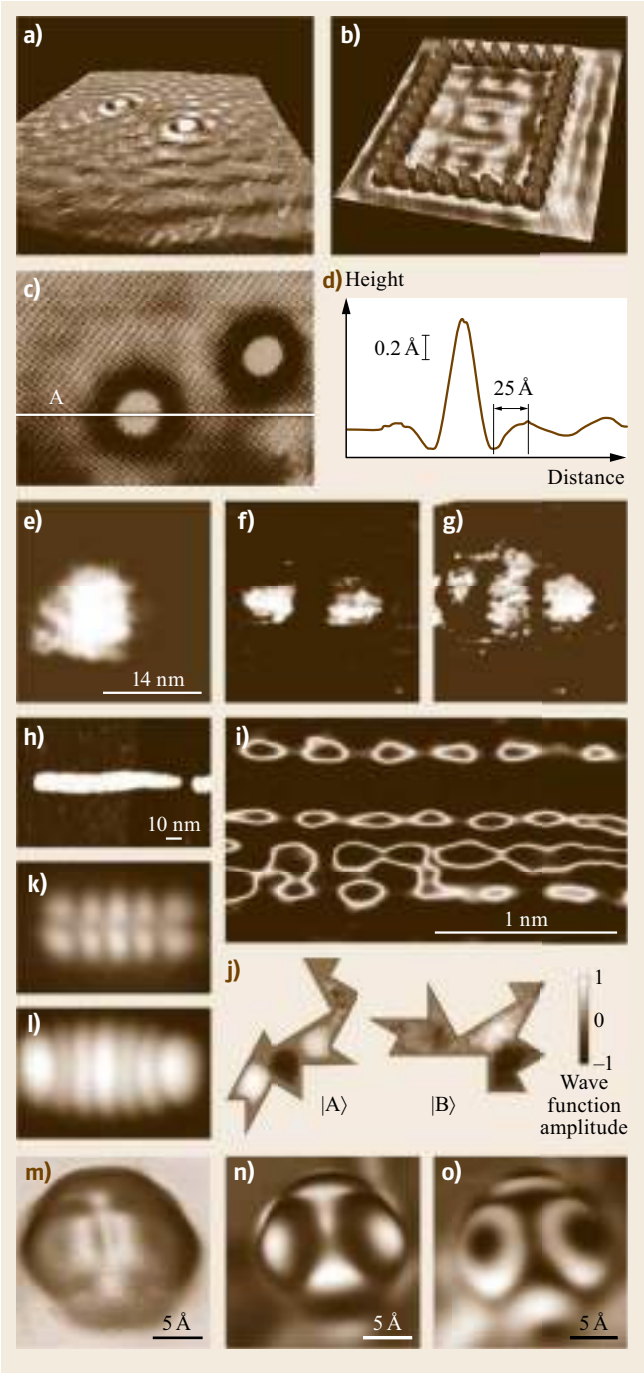


Fig. 24.8a–d **(a)** Low-voltage STM image of Cu(111) including two defect atoms; the waves are electronic Bloch waves scattered at the defects; **(b)** low-voltage STM image of a rectangular quantum corral made from single atoms on Cu(111); the pattern inside the corral is the confined state of the corral close to E_F; (© D. Eigler, Almaden **(a,b)**); **(c)** STM image of GaAs(110) around a Si donor, $V = -2.5$ V; the line scan along A, shown in **(d)**, exhibits an additional oscillation around the donor caused by a standing Bloch wave; the grid-like pattern corresponds to the atomic corrugation of the Bloch wave (© H. van Kempen, Nijmegen **(c,d)**); **(e–g)** dI/dV images of a self-assembled InAs quantum dot deposited on GaAs(100) and measured at different V (**(e)** 1.05 V, **(f)** 1.39 V, **(g)** 1.60 V). The *images* show the squared wavefunctions confined within the quantum dot, which exhibit zero, one, and two nodal lines with increasing energy. **(h)** STM image of a short-cut carbon nanotube; **(i)** greyscale plot of the dI/dV intensity inside the short-cut nanotube as a function of position (x-axis) and tunneling voltage (y-axis); four wavy patterns of different wavelength are visible in the voltage range from −0.1 to 0.15 V (© C. Dekker, Delft **(h,i)**); **(j)** two reconstructed wavefunctions confined in so-called isospectral corrals made of CO molecules on Cu(111). Note that $\Psi(x)$ instead of $|\Psi(x)|^2$ is displayed, exhibiting positive and negative values. This is possible since the transplantation matrix transforming one isospectral wavefunction into another is known (© H. Manoharan, Stanford **(j)**); **(k,l)** STM images of a pentacene molecule deposited on NaCl/Cu(100) and measured with a pentacene molecule at the apex of the tip at $V = -2.5$ V (**(k)** HOMO = highest occupied molecular orbital) and $V = 2.5$ V (**(l)** LUMO = lowest unoccupied molecular orbital) (© J. Repp, Regensburg **(k,l)**); **(m)** STM image of a C_{60} molecule deposited on Ag(100), $V = 2.0$ V; **(n,o)** dI/dV images of the same molecule at $V = 0.4$ V **(n)**, 1.6 V **(n)** (**(m–o)** © M. Crommie, Berkeley) ►

eral resolution of STM, thereby improving, e.g., the visibility of bonding and antibonding pair states within a dimer [24.113].

Wavefunctions in Disordered Systems

More complex wavefunctions result from interactions. A nice playground to study such interactions is doped semiconductors. The reduced electron density with respect to metals increases the importance of electron interactions with potential disorder and other electrons. Applying a magnetic field quenches the kinetic energy, thus enhancing the importance of interactions. A dramatic effect can be observed on InAs(110), where three-dimensional (3-D) bulk states are measured. While the usual scattering states around individual dopants are observed at $B = 0$ T (Fig. 24.9a) [24.114], stripe structures are found at high magnetic field (Fig. 24.9b) [24.115]. They run along equipotential lines of the disorder potential. This can be understood by recalling that the electron tries to move in a cyclotron circle, which becomes a cycloid path along an equipotential line within an inhomogeneous electrostatic potential [24.116].

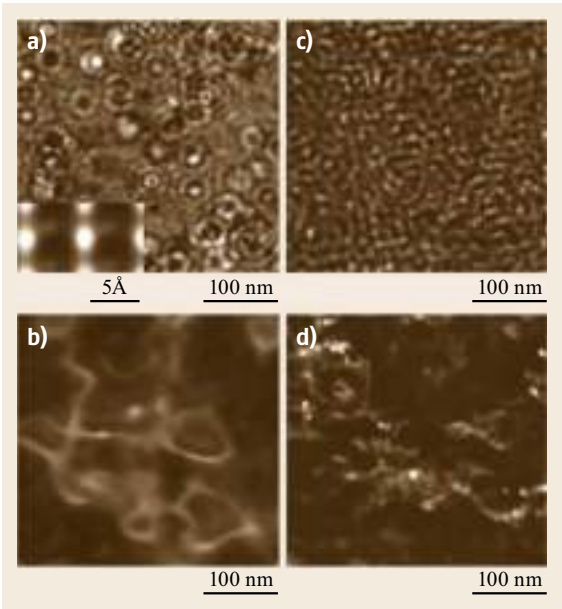


Fig. 24.9 **(a)** dI/dV image of InAs(110) at $V = 50\,\mathrm{mV}$, $B = 0\,\mathrm{T}$; circular wave patterns corresponding to standing Bloch waves around each sulphur donor are visible; *inset* shows a magnification revealing the atomically periodic part of the Bloch wave. **(b)** Same as **(a)**, but at $B = 6\,\mathrm{T}$; the stripe structures are drift states. **(c)** dI/dV image of a 2-D electron system on InAs(110) induced by the deposition of Fe, $B = 0\,\mathrm{T}$. **(d)** Same as **(c)** but at $B = 6\,\mathrm{T}$; note that the contrast in **(a)** is increased by a factor of ten with respect to **(b–d)**

The same effect has been found in two-dimensional (2-D) electron systems (2-DES) of InAs at the same large B-field (Fig. 24.9d) [24.117]. However, the scattering states at $B = 0\,\mathrm{T}$ are much more complex in 2-D (Fig. 24.9c) [24.118]. The reason is the tendency of a 2-DES to exhibit closed scattering paths [24.119]. Consequently, the self-interference does not result from scattering at individual scatterers, but from complicated self-interference paths involving many scatterers. Nevertheless, the wavefunction pattern can be reproduced by including these effects within the calculations.

Reducing the dimensionality to one dimension (1-D) leads again to complicated self-interference patterns due to the interaction of the electrons with several impurities [24.120, 121]. For InAs, they can be reproduced by single-particle calculations. However, experiments imaging self-interference patterns close to the end of a C-nanotube are interpreted as indications of spin charge separation, a genuine property of 1-D electrons not feasible within the single-particle description [24.122].

Charge Density Waves, Jahn–Teller Distortion

Another interaction modifying the LDOS is the electron–phonon interaction. Phonons scatter electrons between different Fermi points. If the wavevectors connecting Fermi points exhibit a preferential orientation, a so-called Peierls instability occurs [24.123]. The corresponding phonon energy goes to zero, the atoms are slightly displaced with the periodicity of the corresponding wavevector, and a charge density wave (CDW) with the same periodicity appears. Essentially, the CDW increases the overlap of the electronic states with the phonon by phase fixing with respect to the atomic lattice. The Peierls transition naturally occurs in one-dimensional (1-D) systems, where only two Fermi points are present and hence preferential orientation is excessive. It can also occur in 2-D systems if large parts of the Fermi line run in parallel.

STS studies of CDWs are numerous [24.124, 125]. Examples of a 1-D CDW on a quasi-1-D bulk material and of a 2-D CDW are shown in Fig. 24.10a–d and Fig. 24.10e–h, respectively [24.126, 127]. In contrast to usual scattering states, where LDOS corrugations are only found close to the scatterer, the corrugations of CDWs are continuous across the surface. Heating the substrate toward the transition temperature leads to a melting of the CDW lattice, as shown in Fig. 24.10f–h.

CDWs have also been found on monolayers of adsorbates such as a monolayer of Pb on Ge(111) [24.128]. These authors performed a nice temperature-dependent study revealing that the CDW is nucleated by scattering states around defects, as one might expect [24.129]. Some of the transitions have been interpreted as more complex Mott–Hubbard transitions caused primarily by electron–electron interactions [24.130]. One-dimensional systems have also been prepared on surfaces showing Peierls transitions [24.131, 132]. Finally, the energy gap occurring at the transition has been studied by measuring $dI/dV(V)$ curves [24.133].

A more local crystallographic distortion due to electron–lattice interactions is the Jahn–Teller effect. Here, symmetry breaking by elastic deformation can lead to the lifting of degeneracies close to the Fermi level. This results in an energy gain due to the lowering of the energy of the occupied levels. By tuning the Fermi level of an adsorbate layer to a degeneracy via doping, such a Jahn–Teller deformation has been induced on a surface and visualized by STM [24.134].

Superconductors

An intriguing effect resulting from electron–phonon interaction is superconductivity. Here, the attractive part of the electron–phonon interaction leads to the coupling of electronic states with opposite wave vector

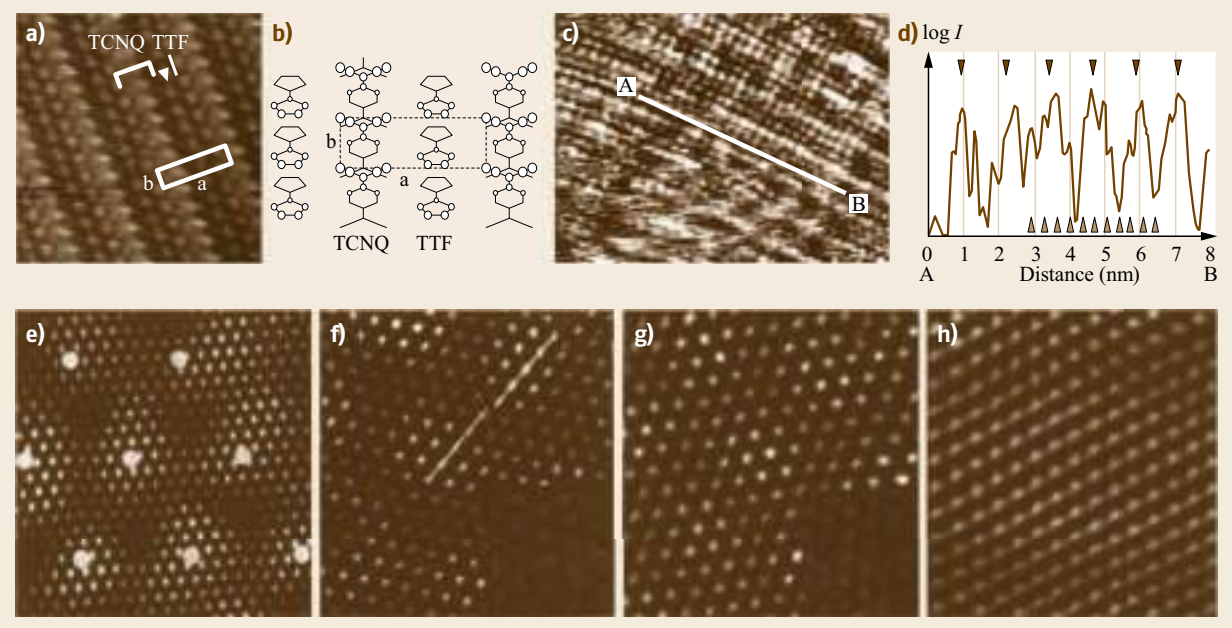


Fig. 24.10 **(a)** STM image of the *ab*-plane of the organic quasi-1-D conductor tetrathiafulvalene tetracyanoquinodimethane (TTF-TCNQ), $T = 300\,\mathrm{K}$; while the TCNQ chains are conducting, the TTF chains are insulating. **(b)** Stick-and-ball model of the *ab*-plane of TTF-TCNQ. **(c)** STM image taken at $T = 61\,\mathrm{K}$; the additional modulation due to the Peierls transition is visible in the profile along line AB shown in **(d)**; the *brown triangles* mark the atomic periodicity and the *black triangles* the expected CDW periodicity (**(a–d)** © M. Kageshima, Kanagawa). **(e–h)** Low-voltage STM images of the two-dimensional CDW system 1 T-TaS_2 at $T = 242\,\mathrm{K}$ **(e)**, 298 K **(f)**, 349 K **(g)**, and 357 K **(h)**. A long-range, hexagonal modulation is visible besides the atomic spots; its periodicity is highlighted by *large white dots* in **(e)**; the additional modulation obviously weakens with increasing T, but is still apparent in **(f)** and **(g)**, as evidenced in the lower-magnification images in the *insets* (**(e–h)** © C. Lieber, Cambridge)

and mostly opposite spin [24.135]. Since the resulting Cooper pairs are bosons, they can condense at LT, forming a coherent many-particle phase, which can carry current without resistance. Interestingly, defect scattering does not influence the condensate if the coupling along the Fermi surface is homogeneous (s-wave superconductor). The reason is that the symmetry of the scattering of the two components of a Cooper pair effectively leads to a scattering from one Cooper pair state to another without affecting the condensate. This is different if the scatterer is magnetic, since the different spin components of the pair are scattered differently, leading to an effective pair breaking, which is visible as a single-particle excitation within the superconducting gap. On a local scale, this effect was first demonstrated by putting Mn, Gd, and Ag atoms on a Nb(110) surface [24.136]. While the nonmagnetic Ag does not modify the gap shown in Fig. 24.11a, it is modified in an asymmetric fashion close to Mn or Gd adsorbates, as shown in Fig. 24.11b. The asymmetry of the additional intensity is caused by the breaking of the particle–hole symmetry due to the exchange interaction between the localized Mn state and the itinerate Nb states.

Another important local effect is caused by the relatively large coherence length of the condensate. At a material interface, the condensate wavefunction cannot stop abruptly, but overlaps into the surrounding material (proximity effect). Consequently, a superconducting gap can be measured in areas of nonsuperconducting material. Several studies have shown this effect on the local scale using metals and doped semiconductors as surrounding materials [24.137, 138].

While the classical type I superconductors are ideal diamagnets, the so-called type II superconductors can contain magnetic flux. The flux forms vortices, each containing one flux quantum. These vortices are accompanied by the disappearance of the superconducting gap and, therefore, can be probed by STS [24.139]. LDOS maps measured inside the gap lead to bright features in the area of the vortex core. Importantly, the length scale of these features is different from the length scale of the magnetic flux due to the difference between the London penetration depth and the electronic coherence length. Thus, STS probes a different property of the vortex than the usual magnetic imaging techniques (Sect. 24.4.5). Surprisingly, first measurements of the

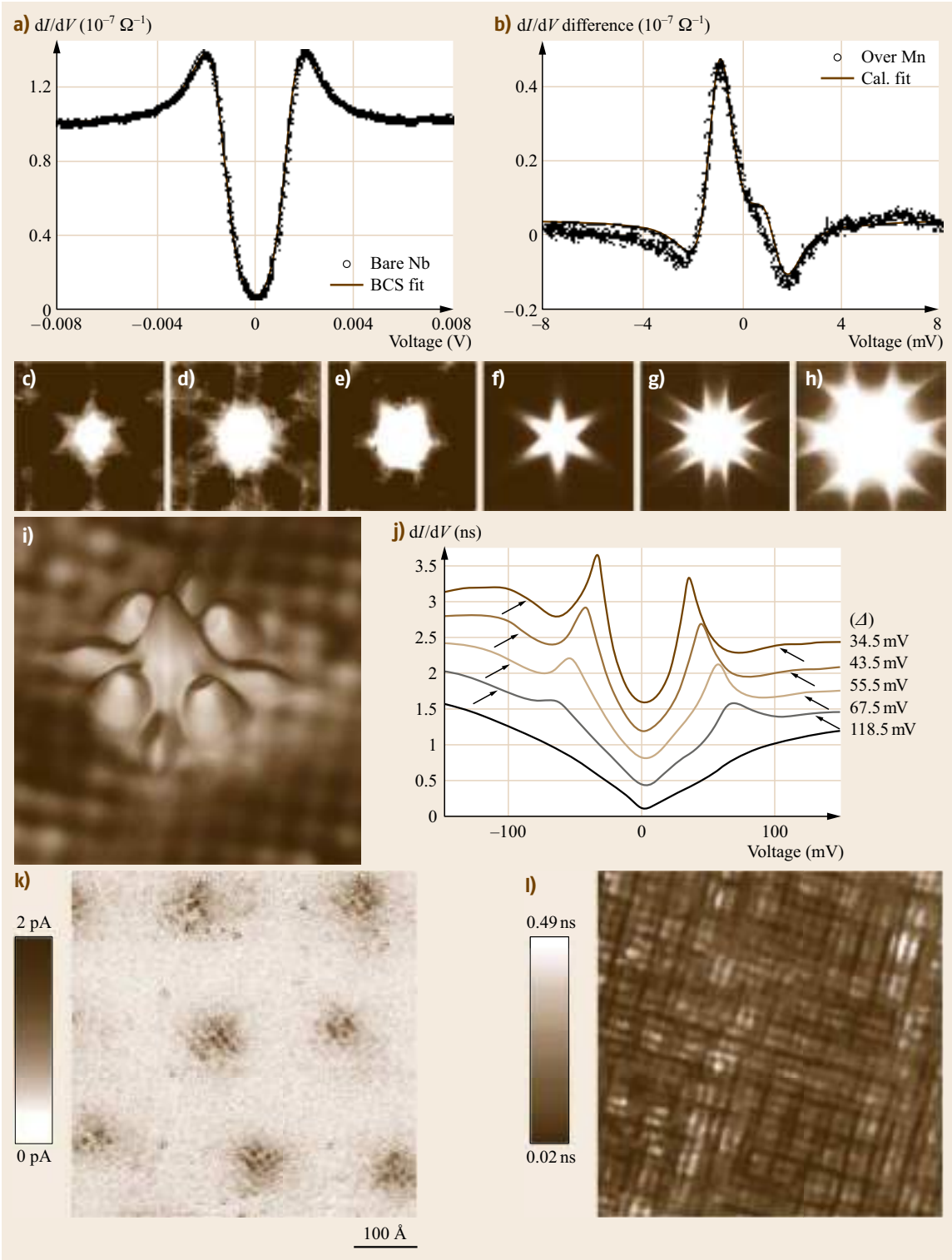

a) dI/dV (10⁻⁷ Ω⁻¹)
1.2
0.8
0.4
0
−0.008
−0.004
0
0.004
0.008
Voltage (V)
Bare Nb
BCS fit
b) dI/dV difference (10⁻⁷ Ω⁻¹)
0.4
0.2
0
−0.2
−8
−4
0
4
8
Voltage (mV)
Over Mn
Cal. fit
c)
d)
e)
f)
g)
h)
i)
j) dI/dV (ns)
3.5
3
2.5
2
1.5
1
0.5
0
−100
0
100
Voltage (mV)
(Δ)
34.5 mV
43.5 mV
55.5 mV
67.5 mV
118.5 mV
k)
2 pA
0 pA
100 Å
l)
0.49 ns
0.02 ns

Fig. 24.11 **(a)** dI/dV curve of Nb(110) at $T = 3.8\,K$ (*symbols*) in comparison with a BCS fit of the superconducting gap of Nb (*line*). **(b)** Difference between the dI/dV curve taken directly above a Mn atom on Nb(110) and the dI/dV curve taken above clean Nb(110) (*symbols*) in comparison with a fit using the Bogulubov–de Gennes equations (*line*) (© D. Eigler, Almaden **(a,b)**). **(c–e)** dI/dV images of a vortex core in the type II superconductor 2H-$NbSe_2$ at 0 mV **(c)**, 0.24 mV **(d)**, and 0.48 mV **(e)** (**(c–e)** © H. F. Hess). **(f–h)** Corresponding calculated LDOS images within the Eilenberger framework (**(f–h)** © K. Machida, Okayama). **(i)** Overlap of an STM image at $V = -100\,mV$ (background 2-D image) and a dI/dV image at $V = 0\,mV$ (overlapped 3-D image) of optimally doped $Bi_2Sr_2CaCu_2O_{8+\delta}$ containing 0.6% Zn impurities. The STM image shows the atomic structure of the cleavage plane, while the dI/dV image shows a bound state within the superconducting gap, which is located around a single Zn impurity. The fourfold symmetry of the bound state reflects the d-like symmetry of the superconducting pairing function; **(j)** dI/dV spectra of $Bi_2Sr_2CaCu_2O_{8+\delta}$ measured at different positions of the surface at $T = 4.2\,K$; the phonon peaks are marked by *arrows*, and the determined local gap size Δ is indicated; note that the strength of the phonon peak increases with the strength of the coherence peaks surrounding the gap; **(k)** LDOS in the vortex core of slightly overdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$, $B = 5\,T$; the dI/dV image taken at $B = 5\,T$ is integrated over $V = 1-12\,mV$, and the corresponding dI/dV image at $B = 0\,T$ is subtracted to highlight the LDOS induced by the magnetic field. The checkerboard pattern within the seven vortex cores exhibits a periodicity, which is fourfold with respect to the atomic lattice shown in **(i)** and is thus assumed to be a CDW; **(l)** STM image of cleaved $Ca_{1.9}Na_{0.1}CuO_2Cl_2$ at $T = 0.1\,K$, i. e., within the superconducting phase of the material; a checkerboard pattern with fourfold periodicity is visible on top of the atomic resolution (© S. Davis, Cornell and S. Uchida, Tokyo **(i–l)**) ◄

vortices on $NbSe_2$ revealed vortices shaped as a sixfold star [24.140] (Fig. 24.11c). With increasing voltage inside the gap, the orientation of the star rotates by 30° (Fig. 24.11d,e). The shape of these stars could finally be reproduced by theory, assuming an anisotropic pairing of electrons in the superconductor (Fig. 24.11f–h) [24.141]. Additionally, bound states inside the vortex core, which result from confinement by the surrounding superconducting material, are found [24.140]. Further experiments investigated the arrangement of the vortex lattice, including transitions between hexagonal and quadratic lattices [24.142], the influence of pinning centers [24.143], and the vortex motion induced by current [24.144].

A central topic is still the understanding of high-temperature superconductors (HTSC). An almost accepted property of HTSC is their d-wave pairing symmetry, which is partly combined with other contributions [24.145]. The corresponding $\boldsymbol{k}$-dependent gap (where $\boldsymbol{k}$ is the reciprocal lattice vector) can be measured indirectly by STS using a Fourier transformation of the LDOS(x, y) determined at different energies [24.146]. This shows that LDOS modulations in HTSC are dominated by simple self-interference patterns of the Bloch-like quasiparticles [24.147]. However, scattering can also lead to pair breaking (in contrast to s-wave superconductors), since the Cooper-pair density vanishes in certain directions. Indeed, scattering states (bound states in the gap) around nonmagnetic Zn impurities have been observed in $Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCCO) (Fig. 24.11i,j) [24.148]. They reveal a d-like symmetry, but not the one expected from simple Cooper-pair scattering. Other effects such as magnetic polarization in the environment probably have to be taken into account [24.149]. An interesting topic is the importance of inhomogeneities in HTSC materials. Evidence for inhomogeneities has indeed been found in underdoped materials, where puddles of the superconducting phase identified by the coherence peaks around the gap are shown to be embedded in nonsuperconducting areas [24.36].

In addition, temperature-dependent measurements of the gap size development at each spatial position exhibit a percolation-type behavior above T_c [24.35]. This stresses the importance of inhomogeneities, but the observed percolation temperature being higher than T_c shows that T_c is not caused by percolation of superconducting puddles only. Conversely, it was found that for overdoped and optimally doped samples the gap develops continuously across T_c, showing a universal relation between the local gap size $\Delta(T = 0)$ (measured at low temperature) and the local critical temperature T_p (at which the gap completely disappears): $2\Delta(T = 0)/(k_B T_p) \approx 8$. The latter result is evidence that the so-called pseudogap phase is a phase with incoherent Cooper pairs. The results are less clear in the underdoped region, where probably two gaps complicate the analysis. Below T_c, it turns out that the strength of the coherence peak is anticorrelated to the local oxygen acceptor density [24.147] and, in addition, correlated to the energy of an inelastic phonon excitation peak in dI/dV spectra [24.37]. Figure 24.11j shows corresponding spectra taken at different positions, where the coherence peaks and the nearby phonon peaks marked by arrows are clearly visible. The phonon origin of the peak has been proven by the isotope effect, similar to Fig. 24.6a. The strong intensity of the phonon

side-peak as well as the correlation of its strength with the coherence peak intensity points towards an important role of electron–phonon coupling for the pairing mechanism. However, since the gap size does not scale with the strength of the phonon peak [24.150], other contributions must be involved too.

Of course, vortices have also been investigated for HTSC [24.151]. Bound states are found, but at energies that are in disagreement with simple models, assuming a Bardeen–Cooper–Schrieffer (BCS)-like d-wave superconductor [24.152, 153]. Theory predicts, instead, that the bound states are magnetic-field-induced spin density waves, stressing the competition between antiferromagnetic order and superconductivity in HTSC materials [24.154]. Since the spin density wave is accompanied by a charge density wave of half wavelength, it can be probed by STS [24.155]. Indeed, a checkerboard pattern of the right periodicity has been found in and around vortex cores in BSCCO (Fig. 24.11k). Similar checkerboards, which do not show any $E(k)$ dispersion, have also been found in the underdoped pseudogap phase at temperatures higher than the superconducting transition temperature [24.156] or at dopant densities lower than the critical doping [24.157]. Depending on the sample, the patterns can be either homogeneous or inhomogeneous and exhibit slightly different periodicities. However, the fact that the pattern persists within the superconducting phase as shown in Fig. 24.11l, at least for Na-CCOC, indicates that the corresponding phase can coexist with superconductivity. This raises the question of whether spin density waves are the central opponent to HTSC. Interestingly, a checkerboard pattern of similar periodicity, but without long-range order, is also found, if one displays the particle–hole asymmetry of $\mathrm{d}I/\mathrm{d}V(V)$ intensity in underdoped samples at low temperature [24.158]. Since the observed asymmetry is known to be caused by the lifting of the correlation gap with doping, the checkerboard pattern might be directly linked to the corresponding localized holes in the CuO planes appearing at low doping. Although a comprehensive model for HTSC materials is still lacking, STS contributes significantly to disentangling this puzzle.

Notice that all the measurements described above have probed the superconducting phase only indirectly by measuring the quasiparticle LDOS. The superconducting condensate itself could principally also be probed directly using Cooper-pair tunneling between a superconducting tip and a superconducting sample. A proof of principle of this detection scheme has indeed been given at low tunneling resistance ($R \approx 50\,\mathrm{k}\Omega$) [24.159], but meaningful spatially resolved data are still lacking.

24.3.4 Imaging Spin Polarization: Nanomagnetism

Conventional STS couples to the LDOS, i. e., the charge distribution of the electronic states. Since electrons also have spin, it is desirable to also probe the spin distribution of the states. This can be achieved by spin-polarized STM (SP-STM) using a tunneling tip covered by a ferromagnetic material [24.160]. The coating acts as a spin filter or, more precisely, the tunneling current depends on the relative angle α_{ij} between the spins of the tip and the sample according to $\cos(\alpha_{ij})$. Consequently, a particular tip is not sensitive to spin orientations of the sample perpendicular to the spin orientation of the tip. Different tips have to be prepared to detect different spin orientations. Moreover, the stray magnetic field of the tip can perturb the spin orientation of the sample. To avoid this, a technique using antiferromagnetic Cr as a tip coating material has been developed [24.161]. This avoids stray fields, but still provides a preferential spin orientation of the few atoms at the tip apex that dominate the tunneling current. Depending on the thickness of the Cr coating, spin orientations perpendicular or parallel to the sample surface, implying corresponding sensitivities to the spin directions of the sample, are achieved.

SP-STM has been used to image the evolution of magnetic domains with increasing B field (Fig. 24.12a–d) [24.162], the antiferromagnetic order of a Mn monolayer on W(110) [24.163] as well as of a Fe monolayer on W(100) (Fig. 24.12e) [24.164], and the out-of-plane orientation of a magnetic vortex core in the center of a nanomagnet exhibiting the flux closure configuration [24.165].

In addition, more complex atomic spin structures showing chiral or noncollinear arrangements have been identified [24.166–168]. Even the spin orientation of a single adatom could be detected, if the adatom is placed either directly on a ferromagnetic island [24.169] or close to a ferromagnetic stripe [24.26]. In the latter case, hysteresis curves of the ferromagnetic adatoms could be measured, as shown in Fig. 24.12f–h. It was found that the adatoms couple either ferromagnetically (Fig. 24.12g) or antiferromagnetically (Fig. 24.12h) to the close-by magnetic stripe; i. e., the hysteresis is either in-phase or out-of-phase with the hysteresis of the stripe. This behavior, depending on adatom–stripe distance in an oscillating fashion, directly visualizes the famous Ruderman–Kittel–Kasuya–Yoshida (RKKY) interaction [24.26].

An interesting possibility of SP-STM is the observation of magnetodynamics on the nanoscale. Nanoscale ferromagnetic islands become unstable at a certain temperature, the so-called superparamagnetic tran-

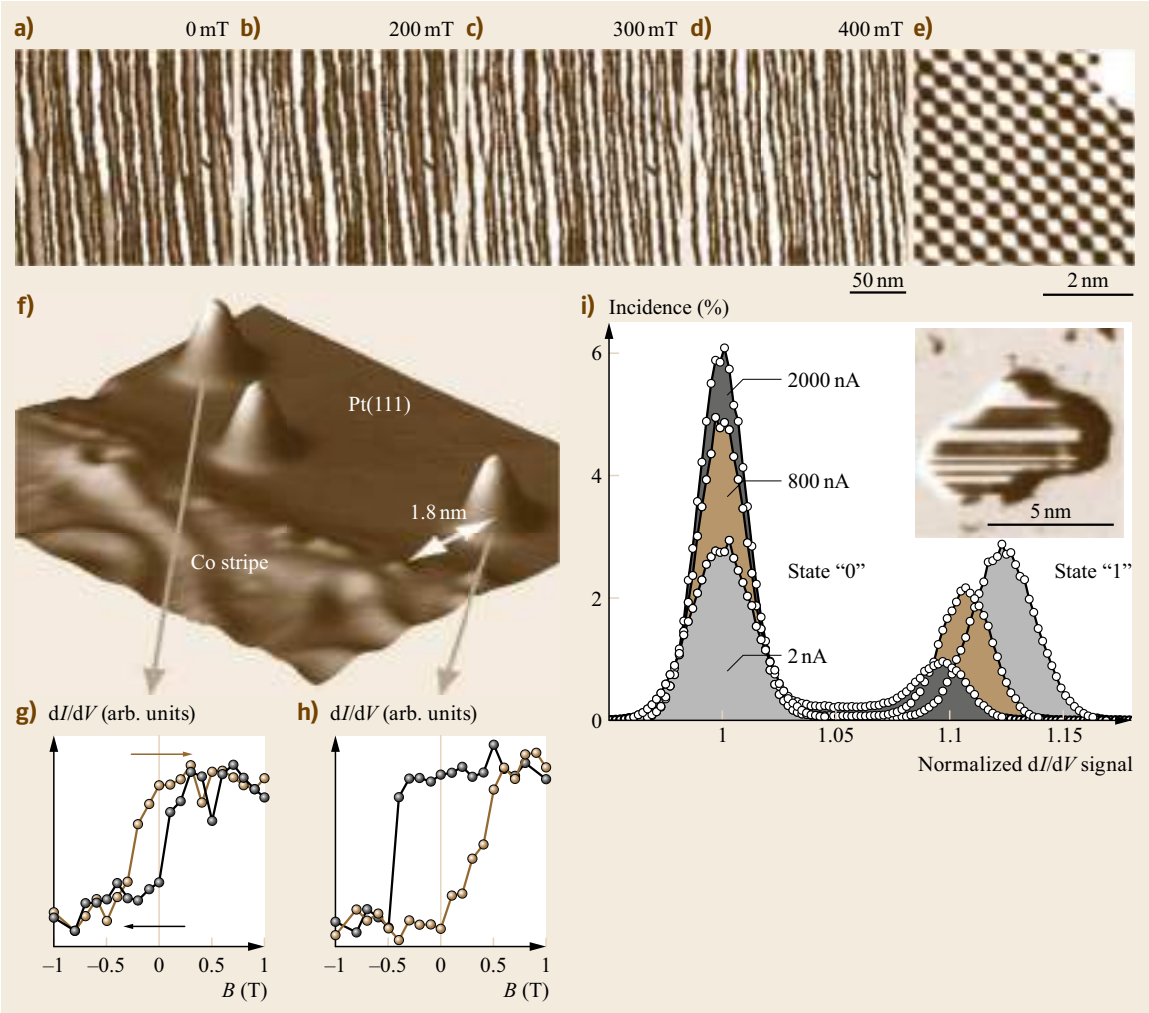


Fig. 24.12 **(a–d)** Spin-polarized STM images of 1.65 monolayers of Fe deposited on a stepped W(110) surface measured at different B fields, as indicated. Double-layer and monolayer Fe stripes are formed on the W substrate; only the double-layer stripes exhibit magnetic contrast with an out-of-plane sensitive tip, as used here. *White* and *grey areas* correspond to different domains. Note that more white areas appear with increasing field (© M. Bode, Argonne **(a–d)**). **(e)** STM image of an antiferromagnetic Fe monolayer on W(001) exhibiting a checkerboard pattern of spin-down (*dark*) and spin-up (*bright*) atoms (© A. Kubetzka, Hamburg); **(f)** STM image of a Pt(111) surface with a Co stripe deposited at the Pt edge as marked. Single Co atoms, visible as *three hills*, are deposited subsequently on the surface at $T = 25\,\mathrm{K}$; **(g,h)** $\mathrm{d}I/\mathrm{d}V(B)$ curves obtained above the Co atoms marked in **(f)** using a spin-polarized tip at $V = 0.3\,\mathrm{V}$. The *colors* mark the sweeping direction of the B field. Obviously the resulting contrast is hysteretic with B and opposite for the two Co atoms. This indicates a different sign of ferromagnetic coupling to the Co stripe. (© J. Wiebe, Hamburg **(f–h)**); **(i)** observed incidences of differential conductivities above a single monolayer Fe island on W(110) with a spin-polarized tip. The *three curves* are recorded at different tunneling currents and the increasing asymmetry shows a preferential spin direction with increasing spin-polarized current. *Inset*: $\mathrm{d}I/\mathrm{d}V$ image of the Fe island at $T = 56\,\mathrm{K}$ showing the irregular change of $\mathrm{d}I/\mathrm{d}V$ intensity (© S. Krause, Hamburg)

sition temperature. Above this temperature, the direction of magnetization switches back and forth due to thermal excitations. This switching results in a stripe-like contrast in SP-STM images, as visible in the *inset* of Fig. 24.12i. The island appears dark during the time when the orientation of the island spin is opposite to the orientation of the tip spin, and switches to bright when the island spin orientation changes. By observing

the switching as a function of time on different islands at different temperatures, the energy barriers of individual islands can be determined [24.170]. Even more importantly, the preferential orientation during switching can be tuned by the tunneling current. This is visible in Fig. 24.12i, which shows the measured orientational probability at different tunneling currents [24.171]. The observed asymmetry in the peak intensity increases with current, providing evidence that current-induced magnetization switching is possible even on the atomic scale.

24.4 Scanning Force Microscopy and Spectroscopy

The examples discussed in the previous section show the wide variety of physical questions that have been tackled with the help of LT-STM. Here, we turn to the other prominent scanning probe method that is applied at low temperatures, namely SFM, which gives complementary information on sample properties on the atomic scale.

The ability to detect *forces* sensitively with spatial resolution down to the atomic scale is of great interest, since force is one of the most fundamental quantities in physics. Mechanical force probes usually consist of a micromachined cantilever with a tip at its free end that is brought close to the sample surface. Basically, two methods exist to detect forces with cantilever-based probes: the *static* and the *dynamic* mode [24.172, 173]. They can be used to generate a laterally resolved image (*microscopy* mode) or to determine the distance dependence of force or a related quantity (*spectroscopy* mode).

In the static mode, a force that acts on the tip bends the cantilever. By measuring its deflection Δz, the tip–sample force F_{ts} can be directly calculated from Hooke's law: $F_{\mathrm{ts}} = c_{\mathrm{z}} \Delta z$, where c_{z} denotes the spring constant of the cantilever. In the various dynamic modes, the cantilever is oscillated with amplitude A, typically at or near its eigenfrequency f_0. Under ambient conditions or in liquids, amplitude modulation (AM-SFM) is typically used to detect amplitude changes or the phase shift between the driving excitation and cantilever oscillation. In vacuum, the frequency shift Δf of the cantilever due to tip–sample interaction is measured by the frequency modulation technique (FM-SFM). Converting the measured quantity (amplitude, phase or frequency shift) into a physically meaningful quantity, e.g., the tip–sample interaction force F_{ts} or potential V_{ts}, is not always straightforward and requires an analysis of the equation of motion of the oscillating tip [24.174–177]. Whatever method is used, the resolution of a cantilever-based force detection method is fundamentally limited by its intrinsic *thermomechanical* noise. If the cantilever is in thermal equilibrium at a temperature T, the equipartition theorem predicts a thermally induced *root-mean-square* (rms) motion of the cantilever in the z-direction of $z_{\mathrm{rms}} = (k_{\mathrm{B}} T / c_{\mathrm{eff}})^{1/2}$, where k_{B} is the Boltzmann constant and $c_{\mathrm{eff}} = c_{\mathrm{z}} + \partial F_{\mathrm{ts}} / \partial z$. Note that usually $\partial F_{\mathrm{ts}} / \partial z \gg c_{\mathrm{z}}$ in the contact regime and $\partial F_{\mathrm{ts}} / \partial z < c_{\mathrm{z}}$ in the noncontact regime. Evidently, this fundamentally limits the force resolution in the static mode. Of course, the same is true for the different dynamic modes, because the thermal energy $k_{\mathrm{B}} T$ excites the eigenfrequency f_0 of the cantilever. Thermal noise is *white* noise, i. e., its spectral density is flat. However, if the cantilever transfer function is taken into account, one can see that the thermal energy mainly excites f_0. This explains the term *thermo* in thermomechanical noise, but what is the *mechanical* part?

A more detailed analysis reveals that the thermally induced cantilever motion is given by

$$z_{\mathrm{rms}} = \sqrt{\frac{2 k_{\mathrm{B}} T B}{\pi c_{\mathrm{z}} f_0 Q}} \,, \tag{24.2}$$

where B is the measurement bandwidth and Q is the quality factor of the cantilever. Analogous expressions can be obtained for all quantities measured in dynamic modes, because the deflection noise translates, e.g., into frequency noise [24.8]. Note that f_0 and c_{z} are correlated with each other via $2\pi f_0 = (c_{\mathrm{z}} / m_{\mathrm{eff}})^{1/2}$, where the effective mass m_{eff} depends on the geometry, the density, and the elasticity of the material. The Q-factor of the cantilever is related to the external damping of the cantilever motion in a medium and on the intrinsic damping within the material. This is the *mechanical* part of the fundamental cantilever noise. Thus, it becomes evident from (24.2) that the smallest fundamental noise is achievable in vacuum, where the Q-factors are more than 100 times larger than in air, and at low temperatures.

Besides the reduced noise, the application of force detection at low temperatures is motivated by increased stability and the possibility to observe phenomena that appear below a certain critical temperature T_{c}, as outlined before. SFM experiments which have been performed at low temperatures until now were motivated by at least one of these reasons and have focused on atomic-scale and intramolecular imaging, force spectroscopy, atomic and molecular manipulation as well as the investigation of electrostatic and magnetic properties at increasingly small length scales. In the following, we briefly describe some exemplary results from each area.

24.4.1 Atomic-Scale and Intramolecular Imaging

In a simplified picture, the dimensions of the tip end and its distance to the surface limit the lateral resolution of force microscopy, since it is a near-field technique. Consequently, atomic resolution requires a stable single atom at the tip apex that has to be brought to a distance of some tenths of nanometers over an atomically flat surface. The latter condition can only be fulfilled in the dynamic mode, where the additional restoring force $c_z A$ at the lower turnaround point prevents the *jump-to-contact* phenomenon [24.19]. By preventing jump-to-contact, true atomic resolution is nowadays routinely obtained in vacuum (and increasingly, in liquids and air) by FM-SFM.

Si(111)-(7 × 7) was the first surface on which true atomic resolution was achieved [24.178] and several studies have been performed at low temperatures on this well-known material [24.179–181]. Because of its importance to the microelectronics industry, the Si(100) surface has also been targeted by low-temperature FM-SFM measurements. In particular, measurements performed at low temperatures have shown that the structure of the surface is defined by buckled Si dimers with tilted bonds leading to a $c(4 \times 2)$ reconstruction [24.182]. Additionally, it has been determined that tip-induced interactions at small tip–sample distances lead to dimer *flipping*, resulting in the occasional observation of a $p(2 \times 1)$ reconstruction during low temperature experiments [24.183]. Moreover, experiments by *Sweetman* et al. have demonstrated deliberate flipping of single dimers on the Si(100) surface at low temperatures via controlled tip–sample interactions, corresponding to pure mechanical toggling of an atomic-scale switch (Fig. 24.13) [24.184, 185]. Other low-temperature SFM experiments performed on semiconducting surfaces include those on indium arsenide, where noticeable displacements of tip and sample atoms due to interaction forces have been observed [24.186, 187]. Additionally, the combination of SFM and STM methods thanks to the utilization of tuning forks have allowed the precise determination of the surface structure of metal-semiconductor systems such as Ag:Si(111)-$(\sqrt{3} \times \sqrt{3})R30^\circ$ at low temperatures [24.188].

As discussed before, a particular advantage associated with low-temperature SFM operation is reduced noise and increased thermal as well as mechanical stability, which allow atomic-resolution imaging of, e.g., *weakly interacting* van der Waals surfaces. For such surfaces, much smaller atomic corrugation amplitudes are expected compared with strongly interacting surfaces of, e.g., semiconductors. A typical example is graphite, a layered material, where the carbon atoms are covalently bonded and arranged in a honeycomb structure within the (0001) plane. Individual graphene layers stick together by van der Waals forces. Because of the *ABA* stacking, three distinctive sites exist on the (0001) surface: carbon atoms with (*A*-type) and without (*B*-type) a neighbor in the next graphite layer and the *hollow site* (*H*-site) in the hexagon center. In static force microscopy operated in the contact regime as well as in STM, the contrast usually exhibits a trigonal symmetry with a periodicity of 246 pm, where *A*- and *B*-site carbon atoms could not be distinguished. However, in high-resolution SFM images acquired at low temperatures, a large maximum and two different minima can be resolved [24.10, 189]. By comparison with other experimental and simulation results in the literature [24.190–192], the large maximum can be assigned to the *H*-site, while the two different minima represent A- and B-type carbon atoms. In addition to graphite, the atomic-scale structure of a pure van der Waals crystal consisting of individual xenon atoms could also be resolved by SFM performed at low temperatures, demonstrating the outstanding resolution and stability of the approach [24.193].

Of course, low-temperature SFM imaging experiments are not limited to semiconductor and van der Waals surfaces only. In particular, metal oxide surfaces such as TiO_2(110) [24.194, 195] and ionic crystal surfaces such as KBr(001) [24.196] have also been imaged with atomic resolution at low temperatures, among others.

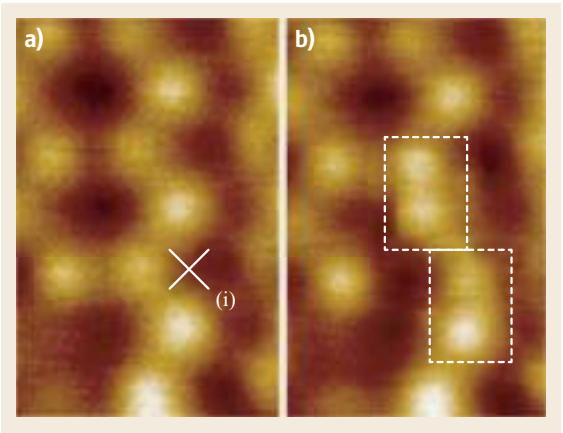


Fig. 24.13 LT-SFM images of the Si(100) surface that **(a)** demonstrate the atomic arrangement of the surface featuring the $c(4 \times 2)$ reconstruction and **(b)** *dimer flipping* (highlighted by *dashed rectangles*) induced by controlled force spectroscopy performed at location (i) in part **(a)**. Adapted from [24.185]

Performing SFM experiments at low temperatures not only allows atomic-scale imaging of clean, atomically flat surfaces but may also remarkably lead to the detailed visualization of *intramolecular* structure of individual molecules adsorbed on surfaces [24.197–202] and even bonds *between* individual molecules [24.203]. The key factor enabling the associated high resolution is the selective and deliberate *functionalization* of the SFM probe tip by, typically, an individual CO molecule. While the exact physical mechanism that gives rise to the submolecular imaging capability (Pauli repulsion versus force-interaction-induced relaxations of the CO molecule *dangling* at the tip end) is still under debate [24.204, 205], high-resolution imaging with functionalized probes certainly represents the state-of-the-art in low-temperature SFM experiments. In particular, while the visualization of the structural backbone of a single pentacene molecule [24.199] was covered worldwide by major media outlets (Fig. 24.14), the applications of the method rather quickly expanded and it recently became possible to precisely identify the structures of certain organic molecules [24.197, 200, 202], discriminate bond order [24.201], and perform KPFM with submolecular resolution [24.206, 207].

24.4.2 Force Spectroscopy

A wealth of information about the nature of the tip–sample interaction can be obtained by measuring its distance dependence. This is usually done by recording the measured quantity (deflection, frequency shift, amplitude change, phase shift) while changing the tip–sample distance by applying an appropriate voltage ramp to the z-electrode of the scan piezo. According to (24.2), low temperatures and high Q-factors (vacuum) considerably increase the force resolution. In the static mode, long-range forces and contact forces can be examined. Force measurements at small tip–sample distances are inhibited by the jump-to-contact phenomenon: If the force gradient $\partial F_{\mathrm{ts}}/\partial z$ becomes larger than the spring constant c_{z}, the cantilever cannot resist the attractive tip–sample forces and the tip snaps into the surface. Sufficiently large spring constants prevent this effect, but reduce the force resolution. In the dynamic modes, as explained before, jump-to-contact can be avoided due to the additional restoring force $c_{\mathrm{z}}A$. The highest sensitivity can be achieved in vacuum by using the FM technique, i. e., by recording $\Delta f(z)$-curves. An alternative FM spectroscopy method, the recording of $\Delta f(A)$-curves, has been suggested in [24.208]. Note that if the amplitude is much larger than the characteristic decay length of the tip–sample force, the frequency shift cannot simply be converted into force gradients by using $\partial F_{\mathrm{ts}}/\partial z = 2c_{\mathrm{z}}\Delta f/f_0$ [24.209]. Several methods have been published to convert $\Delta f(z)$ data into the tip–sample force $F_{\mathrm{ts}}(z)$ and the tip–sample potential $V_{\mathrm{ts}}(z)$ [24.174–177]. In addition to reducing noise and increasing force resolution, a major benefit of perform-

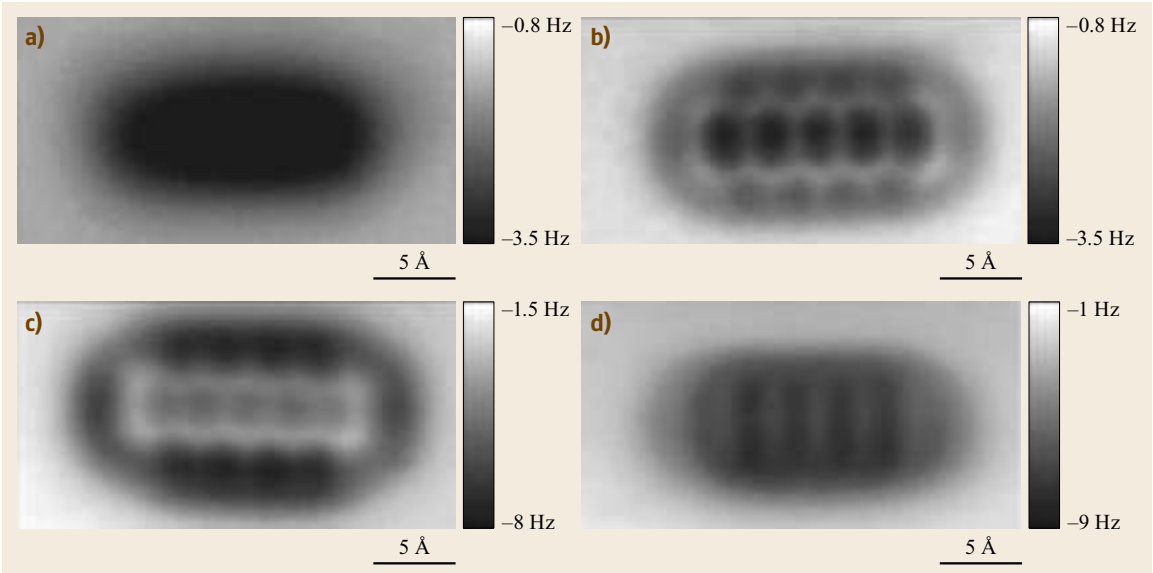


Fig. 24.14a–d Constant-height LT-SFM images of a single pentacene molecule acquired with **(a)** Ag-functionalized, **(b)** CO-functionalized, **(c)** Cl-functionalized, and **(d)** pentacene-functionalized tips. The CO-functionalized tip provides clear intramolecular resolution. The respective color bars indicate Δf values. Adapted from [24.199]. Reprinted with permission from the AAAS

ing force spectroscopy at low temperatures is drastically reduced thermal drift, which allows recording spectroscopy data on the same atomically defined surface site for extended periods of time.

Several research groups around the world have reported atomically resolved force spectroscopy results acquired at low temperatures not long after the first demonstration of atomic-resolution imaging via FM-SFM [24.191, 210–213]. In particular, force spectroscopy data obtained on graphite has been found to be in good agreement with well-known interaction laws (Fig. 24.15a) [24.211]. Additionally, small differences in interaction forces exhibited by different surface atoms on Si(111)-(7 × 7) have also been detected via low-temperature force spectroscopy [24.213]. Moreover, low-temperature force spectroscopy has been utilized on ionic crystal surfaces to, e.g., chemically identify tip and surface atoms as well as characterize elastic deformations and asymmetry artifacts [24.210, 212]. It was even shown that CO-terminated probes can be used to acquire several $\Delta f(z)$-curves above another CO molecule adsorbed on a Cu(111) surface at low temperatures, to obtain a two-dimensional, vertical map of interaction energy between the two molecules with high resolution (Fig. 24.15b) [24.214]. It should be mentioned that while spectacular results regarding chemical identification of individual atoms via force spectroscopy have been previously reported based on experiments performed at *room temperature* [24.215], the signal-to-noise-ratio had to be drastically increased by averaging about 100 curves at every atom species, which required an appropriate atom tracking scheme [24.6, 216].

While the pioneering experiments described above are all based on the careful acquisition of individual $\Delta f(z)$ curves at well-defined surface sites (or several such curves to form two-dimensional, vertical maps as in [24.214] and [24.218]) and eventual conversion to $F_{ts}(z)$ or $V_{ts}(z)$ data, it has recently been shown that it is possible to reliably combine *thousands* of such curves via the method of *three-dimensional atomic force microscopy* (3-D-AFM) to construct volumetric maps of interaction force/energy with pm spatial and pN (meV) force (energy) resolution [24.10, 194, 217, 219, 220], mainly thanks to the development of specialized data acquisition/processing procedures and operation at low temperatures which severely reduces thermal drift [24.221, 222]. Following the first *proof-of-concept* demonstration of 3-D-AFM on graphite [24.10], the method has been expanded to cover simultaneous acquisition of tunneling currents together with interaction forces. The corresponding technique of 3-D-AFM/STM (or alternatively, 3-D-SPM) has been recently shown to deliver complementary 3-D maps of interaction force

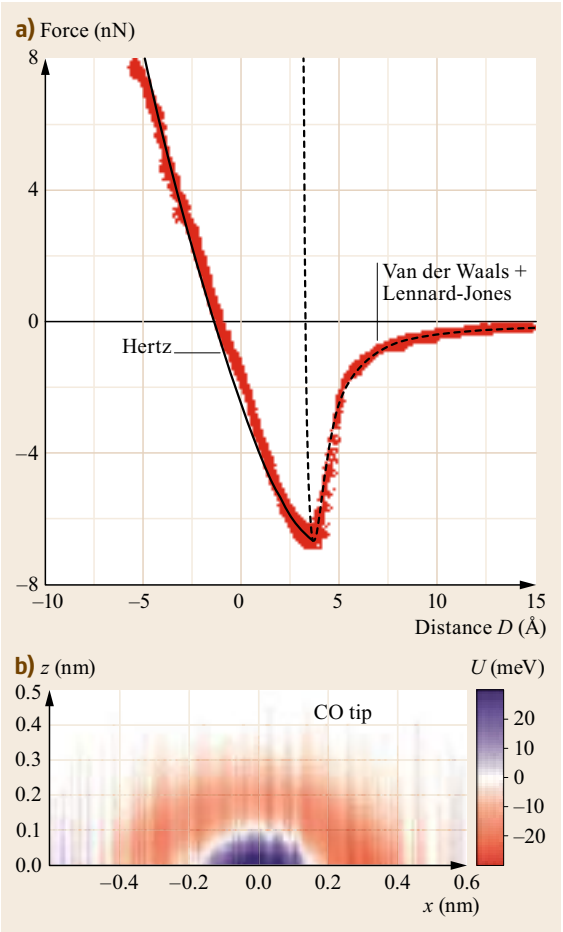


Fig. 24.15 **(a)** A force spectroscopy curve acquired on graphite with a silicon tip at low temperature. The interaction at first closely follows a combination of van der Waals and Lennard-Jones laws. At closer tip–sample distances, Hertzian contact mechanics approximates the force interaction reasonably well. **(b)** A two-dimensional map of interaction energy acquired with a CO-terminated tip on a CO molecule adsorbed on Cu(111). *Red* (*blue*) corresponds to attractive (repulsive) interactions. Adapted from [24.211] and [24.214], respectively

and tunneling current on sample surfaces of surface-oxidized copper (Cu(100)–O) [24.217] and most recently, TiO_2(110) [24.194], with atomic resolution. Specifically, Fig. 24.16 shows the atomic structure associated with the Cu(100)–O surface and the corresponding three-dimensional map of interaction forces spanning a lateral extent of 2.89 nm × 2.89 nm. The magnitude of the maximum interaction force measured in the bottom plane is ≈ 1.34 nN whereas the contrast between atomically defined lattice sites is ≈ 23 pN. Moreover, a *horizontal* cut of interaction forces ex-

tracted from the 3-D force map of Fig. 24.16b at the tip–sample distance of closest approach is shown in Fig. 24.17a, together with the simultaneously recorded map of tunneling currents at the same tip–sample distance (Fig. 24.17b). While symmetry arguments allow assigning the bright spots in the force image to individual oxygen atoms, the tunneling current map exposes a completely different structural arrangement that corresponds to the copper atoms of the surface. In addition to the simultaneous imaging of multiple atomic species on the sample surface, other exciting features of the discussed data set involve (i) the detection of minute differences in force interaction on individual oxygen atoms and (ii) the visualization and structural/chemical identification of surface defects, with support from ab initio calculations [24.217]. To further illustrate the scientific insight that 3-D-AFM/STM measurements can provide, simultaneously acquired maps of interaction force and tunneling current on the prototypical metal oxide surface of TiO_2(110) are shown in Fig. 24.17c,d. In particular, it is observed that a peculiar defect in the form of a subsurface hydrogen atom modulates the surface density of electronic states, without a detectable influence on interaction force [24.194]. Thus, it is seen that the detailed information on tip–sample interactions that is obtained by 3-D-AFM/STM definitely has the potential to provide much needed experimental input into research fields as diverse as catalysis and friction, among others.

It must be mentioned that the application of the method of 3-D-AFM is not restricted to low temperatures and it has indeed been applied multiple times at room temperature thanks to the development of software-based drift correction methods [24.223, 224].

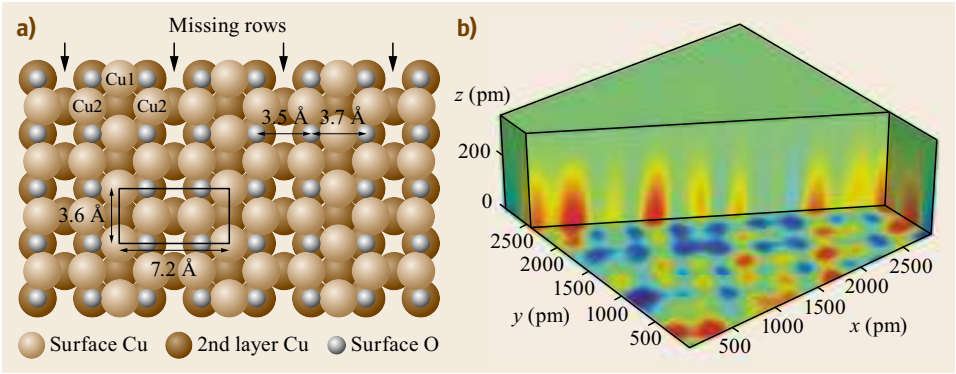


Fig. 24.16 **(a)** Structural model of the surface-oxidized Cu(100) surface (Cu(100)–O). **(b)** Atomically resolved, 3-D map of interaction forces acquired on Cu(100)–O via 3-D-AFM at low temperature. Adapted from [24.217]

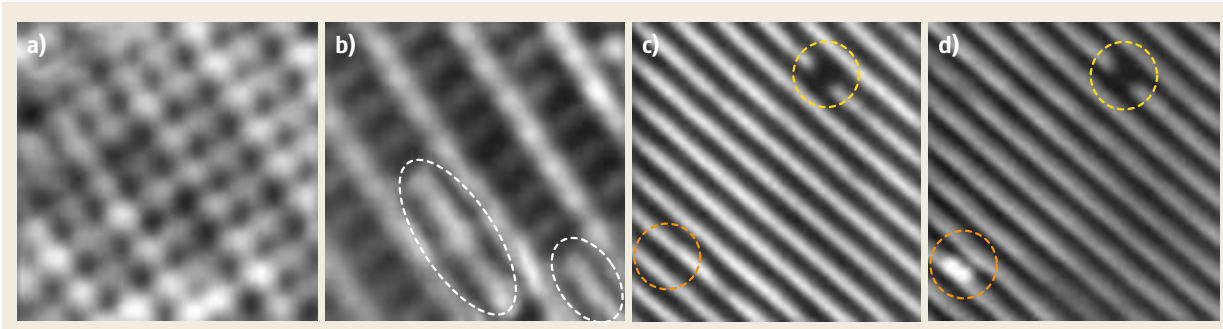


Fig. 24.17 **(a,b)** Simultaneously acquired, atomic-scale maps (2.89 nm × 2.89 nm) of interaction force **(a)** and tunneling current **(b)** on the Cu(100)–O surface at low temperature. Circular maxima in **(a)** correspond to oxygen atoms while the tunneling current channel **(b)** resolves copper atoms. Linear surface defects are highlighted by *dashed white ellipses*. **(c,d)** Simultaneously acquired, atomic-scale maps (6.55 nm × 6.55 nm) of interaction force **(c)** and tunneling current **(d)** on the TiO_2(110) surface at low temperature. *Bright rows* correspond to the arrangement of bridging oxygen ions on the surface. While a point defect in the form of an oxygen vacancy (*dashed yellow circle*) is visible in both data channels, a subsurface hydrogen interstitial (*dashed orange circle*) is found to modulate only the tunneling current data, without a detectable influence on interaction force. Adapted from [24.217] and [24.194], respectively

On the other hand, low-temperature operation results in certain significant advantages that have been discussed in detail in the literature [24.222, 225]. Finally, the application of 3-D-AFM has recently expanded from atomically flat substrate surfaces to individual molecules [24.199, 226].

24.4.3 Atomic and Molecular Manipulation

Atomic-scale manipulation has first been demonstrated via the application of LT-STM and the method is still used successfully to manipulate individual adsorbates (atoms or molecules) on conducting surfaces (Sect. 24.3.1). In most of these experiments the adsorbate is dragged with the tip by the attractive force acting between the foremost tip apex atoms and the adsorbate. By adjusting a large or a small tip–sample distance via the tunneling resistance, it is possible to switch between imaging and manipulation. In the past decade, it has been shown that controlled manipulation of individual atoms and molecules is also possible by using dynamic SFM. Vertical manipulation was performed by pressing the tip in a controlled manner into a Si(111)-(7×7) surface [24.227] at low temperatures. The strong repulsion leads to the removal of the selected silicon atom. The process could be traced by recording the frequency shift and the damping signal during the approach. For lateral manipulation, a *rubbing* technique has been utilized [24.228] – again at low temperature – where the slow scan axis is halted above a selected atom, while the tip–surface distance is gradually reduced until the selected atom hops to a new stable position. Figure 24.18 shows a Ge adatom on Ge(111)-$c(2\times8)$ that was moved during scanning in two steps from its original position (Fig. 24.18a) to its final position (Fig. 24.18c). More recently, vertical manipulation of individual Cu atoms on a reconstructed Cu(110): O $-$ $p(2\times1)$ surface has been demonstrated at low temperature and the associated mechanism elucidated via a combined density functional theory and kinetic Monte Carlo simulation approach [24.229].

While it is not unreasonable to think that operation at low temperature would be a prerequisite for atomic-scale manipulation due to the fact that the diffusion and desorption processes associated with the adsorbates would be severely suppressed, it has been shown that it is in fact possible to perform atomic-scale manipulation successfully at room temperature using dynamic SFM on semiconducting [24.230, 231] and even insulating surfaces [24.232], leading to the ability to construct complex atomic-scale patterns on surfaces via a bottom-up and deterministic approach (Fig. 24.19), with serious implications for *atomic-scale* device designs.

Atoms have not been the only species that have been controllably manipulated via SFM. In particular, it was shown a decade ago that individual, water-related defects can be manipulated on the insulating CaF_2(111) surface [24.233]. Additionally, individual molecules have also been laterally manipulated on various substrates by SFM [24.47, 234–236] and even directed rotations of individual porphyrin molecules could be induced by precisely controlled force spectroscopy [24.237].

A major advantage of SFM compared to STM within the framework of atomic-scale manipulation is the fact that SFM allows researchers to measure vertical and lateral forces involved in the manipulation process. In a seminal paper published in 2008, *Ternes* et al. have manipulated individual Co atoms and CO molecules on Pt(111) and Cu(111) surfaces, and precisely measured the lateral forces required for manip-

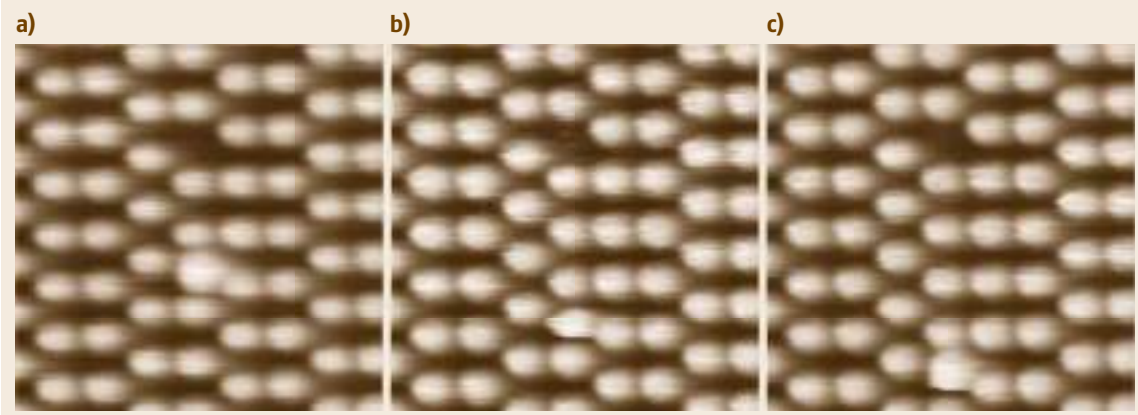


Fig. 24.18a–c Consecutively recorded FM-AFM images showing the tip-induced manipulation of a Ge adatom on Ge(111)-c(2×8) at 80 K. Scanning was performed from bottom to top (© N. Oyabu [24.228])

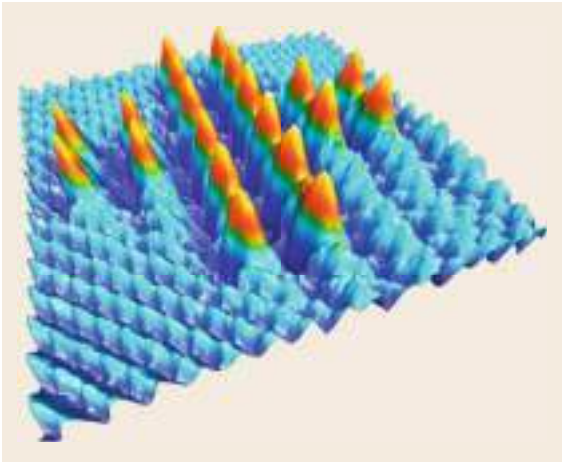


Fig. 24.19 Three-dimensional representation of a *Swiss cross* made of substitutional bromine ions that are assembled atom-by-atom on a sodium chloride surface via SFM. This result represents the first time that controllable atomic-scale manipulation on an insulator surface has been accomplished at room temperature. Adapted from [24.232]

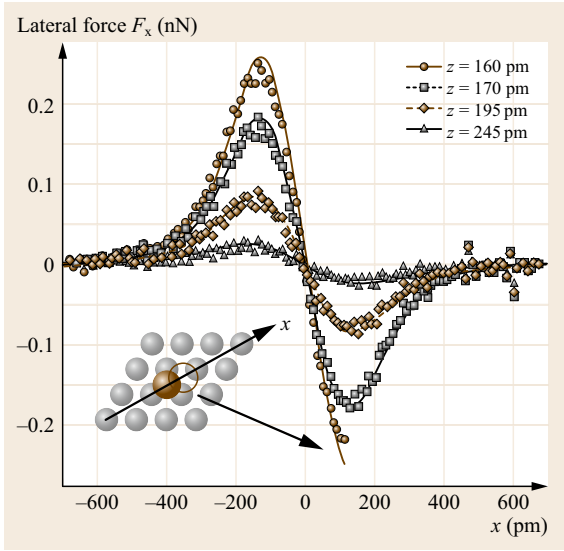


Fig. 24.20 Lateral force curves recorded at constant tip–sample separation z. At the lowest separation a discontinuity appears, which marks the jump of the Co atom form one site to the next as indicated in the *inset* (©. M. Ternes [24.47])

ulation [24.47]. This goal was achieved by recording constant-height data at different tip–sample distances to obtain first $V_{ts}(x, z)$ and then the lateral force $F_x(x, z) = \partial V_{ts}(x, z)/\partial x$. Four curves of the whole data set are displayed in Fig. 24.20. The discontinuity of the lateral force at the lowest adjusted tip–sample distance ($z = 160\,\text{pm}$) indicates the jump of a Co atom from one hollow site to the next on Pt(111) (see *inset* of Fig. 24.20). The jump takes place at about 210 pN, corresponding to the lateral force required for manipulation. More recently, *Kawai* et al. have performed a remarkable experiment where individual nanoribbons of graphene have been picked up from one side and controllably slid on a clean Au(111) surface via dynamic SFM at low temperature, whereby *superlubric* sliding was encountered [24.238].

24.4.4 Kelvin Probe Force Microscopy

The use of Kelvin probe force microscopy (KPFM) [24.239] in nanoscience and nanotechnology research has accelerated rapidly in recent years, thanks to the capability of the method to readily visualize local contact potential differences (LCPDs) over sample surfaces and even individual molecules with very high (down to atomic-scale and even intramolecular) resolution. The operational details of the method are explained elsewhere [24.240] and it should be sufficient here to state that KPFM is applied in conjunction with dynamic SFM, and the imaging typically involves a new feedback loop that is employed to determine the LCPD at each lateral position on a given sample surface by precisely measuring the bias voltage that needs to be applied for a minimization of the electrostatic tip–sample interaction. This leads to the simultaneous acquisition of a *KPFM map* together with the conventional topography map on the sample surface, whereby the KPFM map now reflects local LCPD variations on the surface.

While the mentioned variations in LCPD can be directly correlated to work function differences for conducting surfaces, in the case of insulating surfaces, KPFM contrast was shown to be determined by fixed charges and/or dipoles [24.240]. Even though the application of the KPFM method is not restricted to low temperatures, certain exciting results have been reported under such conditions. In particular, *König* et al. have determined the charge state of individual defects [24.241] and it was even possible to achieve submolecular resolution with KPFM and thus probe the charge distribution inside individual molecules at low temperatures [24.206, 207]. Specifically, a CO-functionalized tip has been utilized by *Mohn* et al. to perform combined NC-AFM/KPFM (NC: non-contact) experiments on a single naphthalocyanine molecule, revealing an asymmetric charge distribution inside the molecule (Fig. 24.21) [24.206]. Thus, the capability of KPFM at low temperatures to provide essential quantitative information about the physical and chemical properties of individual adsorbates is demonstrated.

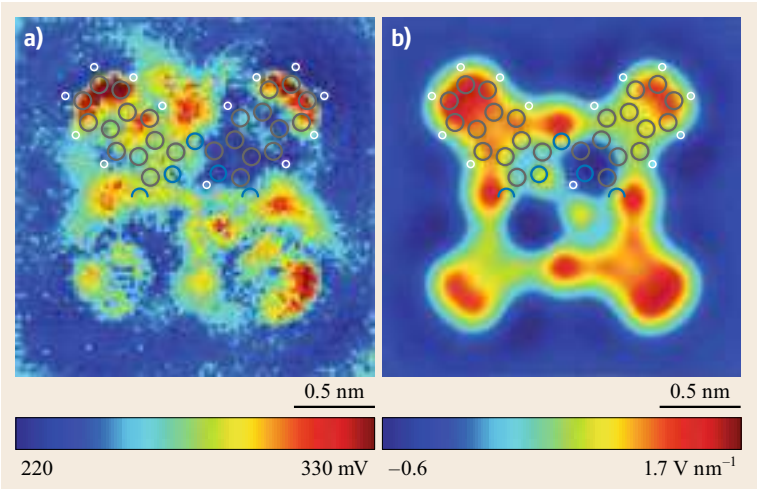


Fig. 24.21 **(a)** KPFM map of an individual naphthalocyanine molecule acquired at low temperature that highlights an asymmetric intramolecular charge distribution and **(b)** the corresponding calculation of electric field magnitude at a distance of 0.3 nm above the molecule. Adapted from [24.206]

24.4.5 Magnetic Force Microscopy

To detect magnetostatic tip–sample interactions with magnetic force microscopy (MFM), a ferromagnetic probe has to be used. Such probes are readily prepared by evaporating a thin magnetic layer, e.g., 10 nm iron, onto the tip. Because of the in-plane shape anisotropy of thin films, the magnetization of such tips lies predominantly along the tip axis, i. e., perpendicular to the surface. Since magnetostatic interactions are long-range, they can be separated from the topography by scanning at a certain constant height (typically around 20 nm) above the surface, where the z-component of the sample stray field is probed (Fig. 24.22a). Therefore, MFM is always operated in the noncontact regime. The signal from the cantilever is directly recorded while the z-feedback is switched off. MFM can be operated in the static mode or in the dynamic modes (AM-MFM at ambient pressures and FM-MFM in vacuum; AM – amplitude-modulation; FM – frequency-modulation). A lateral resolution below 50 nm can be routinely obtained.

Observation of Domain Patterns

MFM is widely used to visualize domain patterns of ferromagnetic materials. At low temperatures, *Moloni* et al. observed the domain structure of magnetite below its Verwey transition temperature (T_V = 122 K) [24.242], but most of the work concentrated on thin films of $La_{1-x}Ca_xMnO_3$ [24.243–245]. Below T_V, the conductivity decreases by two orders of magnitude and a small structural distortion is observed. The domain structure of this mixed valence manganite is of great interest, because its resistivity strongly depends on the external magnetic field, i. e., it exhibits a large colossal magnetoresistive effect. To investigate the field dependence of domain patterns under ambient conditions, electromagnets have to be used. They can cause severe thermal drift problems due to Joule heating of the coils by large currents. Flux densities on the order of 100 mT can be achieved. In contrast, much larger flux densities (more than 10 T) can be rather easily produced by implementing a superconducting magnet in low-temperature setups. With such a design, *Liebmann* et al. recorded the domain structure along the major hysteresis loop of $La_{0.7}Ca_{0.3}MnO_3$ epitaxially grown on $LaAlO_3$ (Fig. 24.22b–f) [24.243]. The film geometry (with thickness of 100 nm) favors an in-plane magnetization, but the lattice mismatch with the substrate induces an out-of-plane anisotropy. Thereby, an irregular pattern of strip domains appears at zero field. If the external magnetic field is increased, the domains with antiparallel orientation shrink and finally disappear in saturation (Fig. 24.22b,c). The residual contrast in saturation (Fig. 24.22d) reflects topographic features. If the field is decreased after saturation (Fig. 24.22e,f), cylindrical domains first nucleate and then start to grow. At zero field, the maze-type domain pattern evolves again. Such data sets can be used to analyze domain nucleation and the domain growth mode. Moreover, because of negligible drift at low temperatures, domain structure and surface morphology can be directly compared, because every MFM can be used as a regular topography-imaging force microscope.

Detection of Individual Vortices in Superconductors

Numerous low-temperature MFM experiments have been performed on superconductors [24.246–252]. Some basic features of superconductors have been mentioned already in Sect. 24.3.3, *Superconductors*. The main difference of STM/STS compared to MFM is its

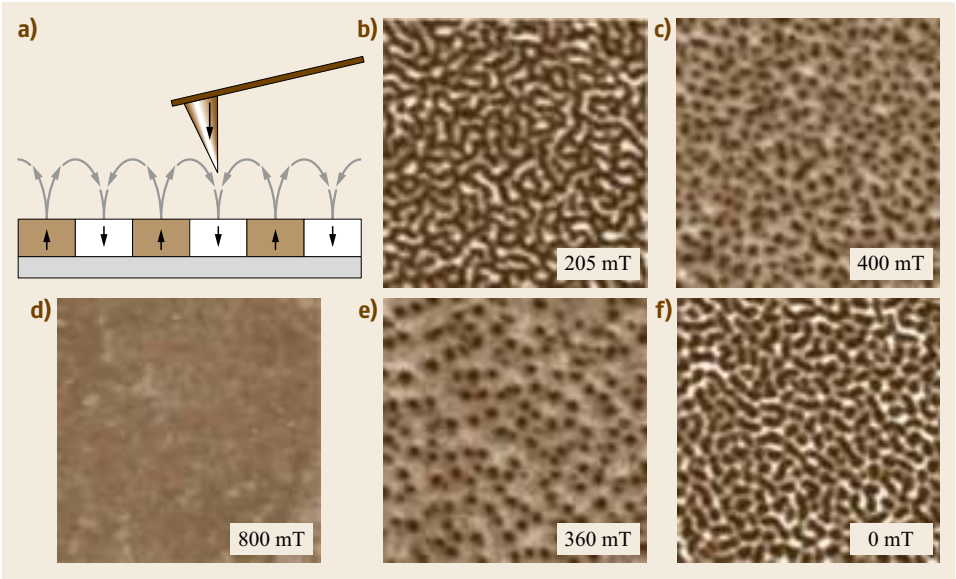


Fig. 24.22 Principle of MFM operation **(a)** and field-dependent domain structure of a ferromagnetic thin film **(b–f)** recorded at 5.2 K with FM-MFM. All images were recorded on the same $4\,\mu m \times 4\,\mu m$ scan area. The $La_{0.7}Ca_{0.3}MnO_3/LaAlO_3$ system exhibits a substrate-induced out-of-plane anisotropy. *Bright* and *dark areas* are visible and correspond to attractive and repulsive magnetostatic interactions, respectively. The series shows how the domain pattern evolves along the major hysteresis loop, i. e., from zero field to saturation at 600 mT and back to zero field. **(b–f)** adapted from [24.243]

high sensitivity to the electronic properties of the surface. Therefore, careful sample preparation is a prerequisite. This is not so important for MFM experiments, since the tip is scanned at a certain distance above the surface.

Superconductors can be divided into two classes with respect to their behavior in an external magnetic field. For type-I superconductors, all magnetic flux is entirely excluded below their critical temperature T_c (Meissner effect), while for type-II superconductors, cylindrical inclusions (*vortices*) of normal material exist in a superconducting matrix (*vortex* state). The radius of the vortex *core*, where the Cooper pair density decreases to zero, is on the order of the coherence length ξ. Since the superconducting gap vanishes in the core, they can be detected by STS. Additionally, each vortex contains one magnetic quantum flux $\Phi = h/2e$ (where h represents Planck's constant and e the electron charge). Circular supercurrents around the core screen the magnetic field associated with a vortex; their radius is given by the London penetration depth λ of the material. This magnetic field of the vortices can be detected by MFM. Investigations have been performed on the two most popular copper oxide high-T_c superconductors $YBa_2Cu_3O_7$ [24.246, 252, 253] and $Bi_2Sr_2CaCu_2O_8$ [24.247, 252], on the only elemental conventional type-II superconductor Nb [24.249, 250], and on the layered compound crystal $NbSe_2$ [24.250, 254].

Most often, vortices have been generated by cooling the sample from the normal state to below T_c in an external magnetic field. After such a *field cooling* procedure, the energetically most favorable vortex arrangement is a regular triangular Abrikosov lattice. *Volodin* et al. were able to observe such an Abrikosov lattice on $NbSe_2$ [24.254]. The intervortex distance d is related to the external field B during cool-down via $d = (4/3)^{1/4}(\Phi/B)^{1/2}$. Another way to introduce vortices into a type-II superconductor is vortex penetration from the edge by applying a magnetic field at temperatures below T_c. According to the Bean model, a vortex density gradient exists under such conditions within the superconducting material. *Pi* et al. slowly increased the external magnetic field until the vortex front approaching from the edge reached the scanning area [24.247].

If the vortex configuration is dominated by the *pinning* of vortices at randomly distributed structural defects, no Abrikosov lattice emerges. The influence of pinning centers can be studied easily by MFM, because every MFM can be used to scan the topography in its AFM mode. This has been done for natural growth defects by *Moser* et al. on $YBa_2Cu_3O_7$ [24.253] and for $YBa_2Cu_3O_7$ and niobium thin films, respectively, by *Volodin* et al. [24.251]. *Roseman* et al. investigated

the formation of vortices in the presence of an artificial structure on niobium films [24.249], while *Pi* et al. produced columnar defects by heavy-ion bombardment in a $Bi_2Sr_2CaCu_2O_8$ single crystal to study the strong pinning at these defects [24.247].

Figure 24.23 demonstrates that MFM is sensitive to the polarity of vortices. In Fig. 24.23a, six vortices have been produced in a niobium film by field cooling in $+0.5$ mT. The external magnetic field and tip magnetization are parallel, therefore, the tip–vortex interaction is attractive (bright contrast). To remove the vortices, the niobium was heated above T_c (≈ 9 K). Thereafter, vortices of opposite polarity have been produced by field cooling in -0.5 mT, which appear dark in Fig. 24.23b. The vortices are probably bound to strong pinning sites, because the vortex positions are identical in both images of Fig. 24.23. By imaging the vortices at different scanning heights, *Roseman* et al. tried to extract values for the London penetration depth from the scan-height dependence of their profiles [24.248]. While good qualitative agreement with theoretical predictions has been found, the absolute values do not agree with published literature values. The disagreement was attributed to the convolution between the tip and vortex stray fields. Better values might be obtained with calibrated tips.

24.4.6 Magnetic Exchange Force Microscopy and Spectroscopy

The resolution of MFM is limited to the nanometer range, because the long-range magnetostatic tip–sample interaction is not localized between individual surface atoms and the foremost tip apex atom [24.255].

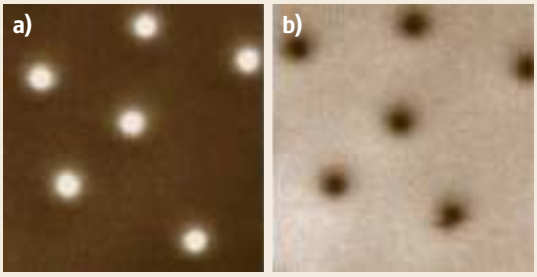


Fig. 24.23 Two $5\,\mu m \times 5\,\mu m$ FM-MFM images of vortices in a niobium thin film after field-cooling at 0.5 mT **(a)** and -0.5 mT **(b)**, respectively. Since the external magnetic field was parallel in **(a)** and antiparallel in **(b)** with respect to the tip magnetization, the vortices exhibit opposite contrast. Strong pinning dominates the position of the vortices, since they appear at identical locations in **(a)** and **(b)** and are not arranged in a regular Abrikosov lattice (© P. Grütter [24.248])

As early as 1991 *Wiesendanger* et al. [24.256] proposed that the short-range magnetic exchange interaction can be utilized to image the configuration of magnetic moments with atomic resolution. For the suggested test system – NiO(001), an antiferromagnetic insulator – *Momida* and *Oguchi* provided density functional calculations [24.257]. They found a magnetic exchange force between the magnetic moments of a single iron atom (the tip) and nickel surface atoms of more than 0.1 nN at tip–sample distances below 0.5 nm. Eventually, *Kaiser* et al. were able to prove the feasibility of magnetic exchange force microscopy (MExFM) on NiO(001) [24.258]. The super-exchange between neighboring (111) planes via bridging oxygen atoms results in a row-wise antiferromagnetic configuration of magnetic moments on the (001) surface. Hence the magnetic surface unit cell is twice as large as the chemical surface unit cell. Figure 24.24a shows the atomic-scale contrast due to a pure chemical interaction. Maxima and minima correspond to the oxygen and nickel atoms, respectively. Their arrangement represents the (1×1) surface unit cell. Figure 24.24b exhibits an additional modulation on chemically and structurally equivalent rows of nickel atoms (the minima). The structure corresponds to the (2×1) magnetic surface unit cell. Since the spin-carrying nickel 3d states are highly localized, the magnetic contrast only becomes significant at very small tip–sample distances. Subsequently, *Schmidt*

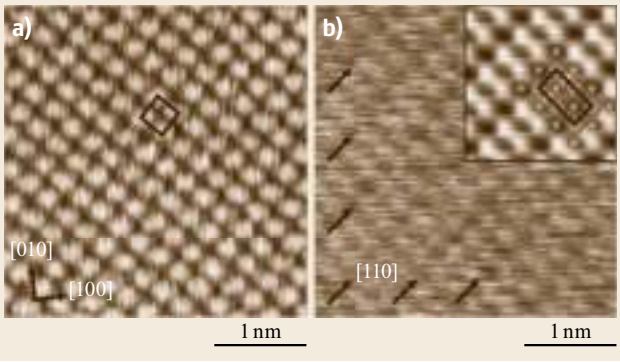


Fig. 24.24 **(a)** Pure chemical contrast on NiO(001) obtained with AFM using a nonmagnetic tip. Oxygen and nickel atoms are represented as maxima and minima, respectively, forming the (1×1) surface unit cell (*black square*). *Arrows* indicate the main crystallographic directions. **(b)** Additional modulation on neighboring nickel rows along the [110] direction (see *arrows*) due to the magnetic exchange interaction obtained with MExFM using a magnetic tip. The (2×1) structure (*black rectangle in the inset*) represents the magnetic surface unit cell. The *inset* is tiled together from the averaged magnetic unit cell calculated from the raw data, whereby the signal-to-noise ratio is significantly increased. Adapted from [24.258]

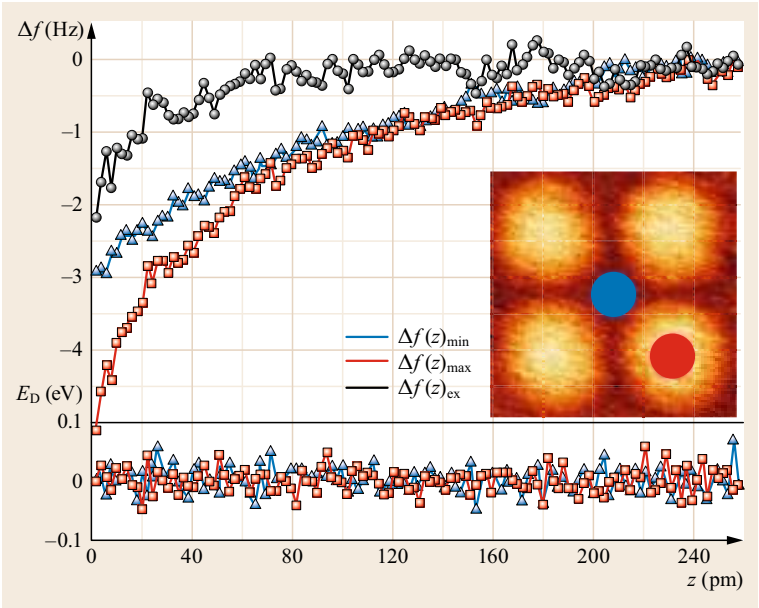


Fig. 24.25 Individual force spectroscopy curves acquired on two iron atoms with opposite spin (*blue* and *red*, respectively) and the corresponding difference (*black*) which is directly related to the magnetic exchange interaction. The curves in the bottom panel highlight the use of a *nondissipative* tip (energy dissipation $E_D \approx 0$). Adapted from [24.259]

et al. were able to perform MExFM with a much better signal-to-noise ratio on an itinerant metallic system: the antiferromagnetic iron monolayer on W(001) [24.260]. Density functional theory performed with a realistic tip model indicated significant relaxations of tip and sample atoms during imaging. Moreover, a comparison between simulation and experimental data revealed complex interplay between chemical and magnetic interaction, which results in the observed atomic-scale contrast. More recently, it became possible to quantify the distance dependence of the magnetic exchange interaction on metallic systems using magnetic exchange force spectroscopy (MExFS) [24.259] that relies on three-dimensional recording of interaction forces (Sect. 24.4.2). Specifically, force spectroscopy data obtained on atoms with opposite spin are subtracted from each other and the result corresponds to the magnetic exchange interaction (Fig. 24.25). Moreover, controllable magnetization switching via the utilization of the magnetic exchange interaction during MExFS experiments has been demonstrated [24.261].

24.4.7 Magnetic Resonance Force Microscopy

Another ambitious force-based approach utilizing magnetic tips is the proposed detection of individual spins by magnetic resonance force microscopy (MRFM) [24.262, 263]. Conventionally, nuclear spins are investigated by nuclear magnetic resonance (NMR), a spectroscopic technique to obtain microscopic chemical and physical information about molecules. An important application of NMR for medical diagnostics is magnetic resonance imaging (MRI). This tomographic imaging technique uses the NMR signal from thin slices through, e.g., the human body, to reconstruct its three-dimensional structure. Currently, at least 10^{12} nuclear spins must be present in a given volume to obtain a significant MRI signal. The ultimate goal of MRFM is to combine aspects of force microscopy with MRI to achieve true 3-D imaging with atomic resolution and elemental selectivity.

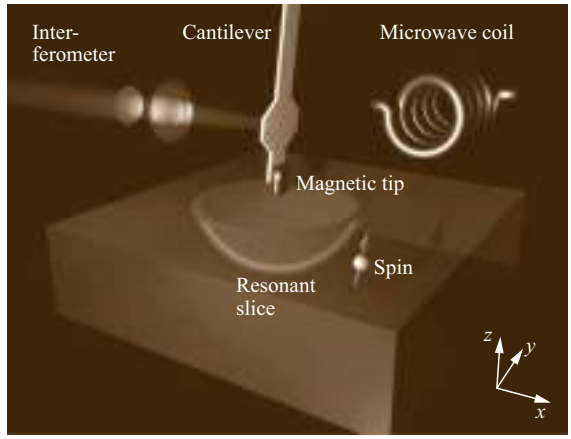


Fig. 24.26 MRFM setup. The cantilever with the magnetic tip oscillates parallel to the surface. Only electron spins within a hemispherical slice, where the stray field of the tip plus the external field matches the condition for magnetic resonance, can contribute to the MRFM signal due to cyclic spin inversion (© D. Rugar [24.264])

The experimental setup is sketched in Fig. 24.26. An oscillating cantilever with a magnetic tip at its end points toward the surface. The spherical resonant slice within the sample represents those points where the stray field from the tip and the external field match the condition for magnetic resonance. The cyclic spin flip causes a slight shift of the cantilever frequency due to the magnetic force exerted by the spin on the tip. Since the forces are extremely small, very low temperatures are required.

To date, no individual nuclear spins have been detected by MRFM. However, the design of ultrasensitive cantilevers made considerable progress, and the detection of forces below 1×10^{-18} N has been achieved [24.265]. Thereby, it became possible to perform nuclear magnetic resonance [24.266] and ferromagnetic resonance [24.267] experiments of spin ensembles with micrometer resolution. Moreover, in SiO_2 the magnetic moment of a single electron, which is three orders of magnitude larger than the nuclear magnetic moment, could be detected by *Rugar* et al. using the setup shown in Fig. 24.26 at 1.6 K [24.264]. This major breakthrough demonstrates the capability of force microscopy to detect single spins. More recently, *Degen* et al. have combined MRFM with 3-D image reconstruction to enable MRI on biological samples with a spatial resolution of $\approx$ 10 nm [24.268].

24.5 Summary

It was the intention of this chapter to give an introduction to the application of scanning probe microscopy at low temperatures. The first three sections were devoted to a general introduction, an analysis of the specialities of microscope operation at low temperatures, and an overview of related instrumentation. It was emphasized that low-temperature operation of SPMs can significantly increase both the achievable resolution and the overall stability of the instrument.

The goal of the subsequent sections was to illustrate some of the most active areas of application for LT-SPM. In particular, LT-STM allows the controlled manipulation of individual atoms and the characterization of electronic properties with atomic specificity. On essentially the same scale, LT-SFM allows the measurement of different types of force fields that provide complementary information. Nevertheless, the provided list of applications is certainly not complete, as this would go far beyond the scope of this handbook chapter.

In summary, the wide variety of associated techniques as well as the many physical questions which can be tackled by LT-SPM make it an unprecedented tool for nanotechnology. And despite the many investigations using LT-SPM that have already been performed, the authors believe that *there is still plenty of room at the bottom* [24.269].

References

24.1 G. Binnig, H. Rohrer: Scanning tunneling microscopy, Helvetica Physica Acta **55**, 726–735 (1982)

24.2 R. Wiesendanger: *Scanning Probe Microscopy and Spectroscopy* (Cambridge Univ. Press, Cambridge 1994)

24.3 D.A. Bonnell, D.N. Basov, M. Bode, U. Diebold, S.V. Kalinin, V. Madhavan, L. Novotny, M. Salmeron, U.D. Schwarz, P.S. Weiss: Imaging physical phenomena with local probes: From electrons to photons, Rev. Modern Phys. **84**, 1343–1381 (2012)

24.4 M. Tinkham: *Introduction to Superconductivity* (McGraw-Hill, New York 1996)

24.5 J. Kondo: Theory of dilute magnetic alloys, Solid State Phys. **23**, 183–281 (1969)

24.6 M. Abe, Y. Sugimoto, O. Custance, S. Morita: Room-temperature reproducible spatial force spectroscopy using atom-tracking technique, Appl. Phys. Lett. **87**, 173503 (2005)

24.7 M. Abe, Y. Sugimoto, T. Namikawa, K. Morita, N. Oyabu, S. Morita: Drift-compensated data acquisition performed at room temperature with frequency modulation atomic force microscopy, Appl. Phys. Lett. **90**, 203103 (2007)

24.8 T.R. Albrecht, P. Grutter, D. Horne, D. Rugar: Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity, J. Appl. Phys. **69**, 668–673 (1991)

24.9 F.J. Giessibl, H. Bielefeldt, S. Hembacher, J. Mannhart: Calculation of the optimal imaging parameters for frequency modulation atomic force microscopy, Appl. Surf. Sci. **140**, 352–357 (1999)

24.10 B.J. Albers, T.C. Schwendemann, M.Z. Baykara, N. Pilet, M. Liebmann, E.I. Altman, U.D. Schwarz: Three-dimensional imaging of short-range chemical forces with picometre resolution, Nat. Nanotechnol. **4**, 307–310 (2009)

24.11 M. Morgenstern, D. Haude, V. Gudmundsson, C. Wittneven, R. Dombrowski, R. Wiesendanger: Origin of Landau oscillations observed in scanning tunneling spectroscopy on n-InAs(110), Phys. Rev. B **62**, 7257–7263 (2000)

24.12 B. Uluutku, M.Z. Baykara: Effect of lateral tip stiffness on atomic-resolution force field spectroscopy, J. Vac. Sci. Technol. B **31**, 041801 (2013)

24.13 B. Uluutku, M.Z. Baykara: Artifacts related to tip asymmetry in high-resolution atomic force microscopy and scanning tunneling microscopy measurements of graphitic surfaces, J. Vac. Sci. Technol. B **33**, 031802 (2015)

24.14 D.M. Eigler, P.S. Weiss, E.K. Schweizer, N.D. Lang: Imaging Xe with a low-temperature scanning tunneling microscope, Phys. Rev. Lett. **66**, 1189–1192 (1991)

24.15 P.S. Weiss, D.M. Eigler: Site dependence of the apparent shape of a molecule in scanning tunneling micoscope images: Benzene on Pt {111}, Phys. Rev. Lett. **71**, 3139–3142 (1993)

24.16 D.M. Eigler, E.K. Schweizer: Positioning single atoms with a scanning tunnelling microscope, Nature **344**, 524–526 (1990)

24.17 H.J. Hug, B. Stiefel, P.J.A. van Schendel, A. Moser, S. Martin, H.J. Guntherodt: A low temperature ultrahigh vacuum scanning force microscope, Rev. Sci. Instrum. **70**, 3625–3640 (1999)

24.18 B.J. Albers, M. Liebmann, T.C. Schwendemann, M.Z. Baykara, M. Heyde, M. Salmeron, E.I. Altman, U.D. Schwarz: Combined low-temperature scanning tunneling/atomic force microscope for atomic resolution imaging and site-specific force spectroscopy, Rev. Sci. Instrum. **79**, 033704 (2008)

24.19 F.J. Giessibl: Advances in atomic force microscopy, Rev. Modern Phys. **75**, 949–983 (2003)

24.20 G. Dujardin, R.E. Walkup, P. Avouris: Dissociation of individual molecules with electrons from the tip of a scanning tunneling microscope, Science **255**, 1232–1235 (1992)

24.21 H.J. Lee, W. Ho: Single-bond formation and characterization with a scanning tunneling microscope, Science **286**, 1719–1722 (1999)

24.22 J. Repp, G. Meyer, F.E. Olsson, M. Persson: Controlling the charge state of individual gold adatoms, Science **305**, 493–495 (2004)

24.23 R. Berndt, R. Gaisch, J.K. Gimzewski, B. Reihl, R.R. Schlittler, W.D. Schneider, M. Tschudy: Photon emission at molecular resolution induced by a scanning tunneling microscope, Science **262**, 1425–1427 (1993)

24.24 X.H. Qiu, G.V. Nazin, W. Ho: Vibrationally resolved fluorescence excited with submolecular precision, Science **299**, 542–546 (2003)

24.25 B.G. Briner, M. Doering, H.P. Rust, A.M. Bradshaw: Microscopic molecular diffusion enhanced by adsorbate interactions, Science **278**, 257–260 (1997)

24.26 F. Meier, L. Zhou, J. Wiebe, R. Wiesendanger: Revealing magnetic interactions from single-atom magnetization curves, Science **320**, 82–86 (2008)

24.27 S. Loth, M. Etzkorn, C.P. Lutz, D.M. Eigler, A.J. Heinrich: Measurement of fast electron spin relaxation times with atomic resolution, Science **329**, 1628–1630 (2010)

24.28 F. Donati, S. Rusponi, S. Stepanow, C. Waeckerlin, A. Singha, L. Persichetti, R. Baltic, K. Diller, F. Patthey, E. Fernandes, J. Dreiser, Z. Sljivancanin, K. Kummer, C. Nistor, P. Gambardella, H. Brune: Magnetic remanence in single atoms, Science **352**, 318–321 (2016)

24.29 J. Kliewer, R. Berndt, E.V. Chulkov, V.M. Silkin, P.M. Echenique, S. Crampin: Dimensionality effects in the lifetime of surface states, Science **288**, 1399–1402 (2000)

24.30 M.F. Crommie, C.P. Lutz, D.M. Eigler: Imaging standing waves in a two-dimensional electron gas, Nature **363**, 524–527 (1993)

24.31 H. Gawronski, M. Mehlhorn, K. Morgenstern: Imaging phonon excitation with atomic resolution, Science **319**, 930–933 (2008)

24.32 A.J. Heinrich, J.A. Gupta, C.P. Lutz, D.M. Eigler: Single-atom spin-flip spectroscopy, Science **306**, 466–469 (2004)

24.33 B.C. Stipe, M.A. Rezaei, W. Ho: Single-molecule vibrational spectroscopy and microscopy, Science **280**, 1732–1735 (1998)

24.34 C.W.J. Beenakker, H. Vanhouten: Quantum transport in semiconductor nanostructures, Solid State Phys. **44**, 1–228 (1991)

24.35 K.K. Gomes, A.N. Pasupathy, A. Pushp, S. Ono, Y. Ando, A. Yazdani: Visualizing pair formation on the atomic scale in the high-T-c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, Nature **447**, 569–572 (2007)

24.36 K.M. Lang, V. Madhavan, J.E. Hoffman, E.W. Hudson, H. Eisaki, S. Uchida, J.C. Davis: Imaging the granular structure of high-T-c superconductivity in underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$, Nature **415**, 412–416 (2002)

24.37 J. Lee, K. Fujita, K. McElroy, J.A. Slezak, M. Wang, Y. Aiura, H. Bando, M. Ishikado, T. Masui, J.X. Zhu, A.V. Balatsky, H. Eisaki, S. Uchida, J.C. Davis: Interplay of electron-lattice interactions and superconductivity in $Bi_2Sr_2CaCu_2O_{8+\delta}$, Nature **442**, 546–550 (2006)

24.38 R.S. Becker, J.A. Golovchenko, B.S. Swartzentruber: Atomic-scale surface modifications using a tunnelling microscope, Nature **325**, 419–421 (1987)

24.39 P.G. Piva, G.A. DiLabio, J.L. Pitters, J. Zikovsky, M. Rezeq, S. Dogel, W.A. Hofer, R.A. Wolkow: Field regulation of single-molecule conductivity by a charged surface atom, Nature **435**, 658–661 (2005)

24.40 J.A. Stroscio, D.M. Eigler: Atomic and molecular manipulation with the scanning tunneling microscope, Science **254**, 1319–1326 (1991)

24.41 A.J. Heinrich, C.P. Lutz, J.A. Gupta, D.M. Eigler: Molecule cascades, Science **298**, 1381–1387 (2002)

24.42 L. Bartels, G. Meyer, K.H. Rieder: Basic steps of lateral manipulation of single atoms and diatomic clusters with a scanning tunneling microscope tip, Phys. Rev. Lett. **79**, 697–700 (1997)

24.43 J.J. Schulz, R. Koch, K.H. Rieder: New mechanism for single atom manipulation, Phys. Rev. Lett. **84**,

4597–4600 (2000)
24.44 J.A. Stroscio, R.J. Celotta: Controlling the dynamics of a single atom in lateral atom manipulation, Science **306**, 242–247 (2004)
24.45 M. Lastapis, M. Martin, D. Riedel, L. Hellner, G. Comtet, G. Dujardin: Picometer-scale electronic control of molecular dynamics inside a single molecule, Science **308**, 1000–1003 (2005)
24.46 J.A. Stroscio, F. Tavazza, J.N. Crain, R.J. Celotta, A.M. Chaka: Electronically induced atom motion engineered CoCun nanostructures, Science **313**, 948–951 (2006)
24.47 M. Ternes, C.P. Lutz, C.F. Hirjibehedin, F.J. Giessibl, A.J. Heinrich: The force needed to move an atom on a surface, Science **319**, 1066–1069 (2008)
24.48 J.I. Pascual, N. Lorente, Z. Song, H. Conrad, H.P. Rust: Selectivity in vibrationally mediated single-molecule chemistry, Nature **423**, 525–528 (2003)
24.49 T.C. Shen, C. Wang, G.C. Abeln, J.R. Tucker, J.W. Lyding, P. Avouris, R.E. Walkup: Atomic-scale desorption through electronic and vibrational excitation mechanisms, Science **268**, 1590–1592 (1995)
24.50 T. Komeda, Y. Kim, M. Kawai, B.N.J. Persson, H. Ueba: Lateral hopping of molecules induced by excitation of internal vibration mode, Science **295**, 2055–2058 (2002)
24.51 Y.W. Mo: Reversible rotation of antimony dimers on the silicon (001) surface with a scanning tunneling microscope, Science **261**, 886–888 (1993)
24.52 B.C. Stipe, M.A. Rezaei, W. Ho: Inducing and viewing the rotational motion of a single molecule, Science **279**, 1907–1909 (1998)
24.53 P. Liljeroth, J. Repp, G. Meyer: Current-induced hydrogen tautomerization and conductance switching of naphthalocyanine molecules, Science **317**, 1203–1206 (2007)
24.54 P. Maksymovych, D.B. Dougherty, X.Y. Zhu, J.T. Yates Jr.: Nonlocal dissociative chemistry of adsorbed molecules induced by localized electron injection into metal surfaces, Phys. Rev. Lett. **99**, 016101 (2007)
24.55 S. Katano, Y. Kim, M. Hori, M. Trenary, M. Kawai: Reversible control of hydrogenation of a single molecule, Science **316**, 1883–1886 (2007)
24.56 G.V. Nazin, X.H. Qiu, W. Ho: Visualization and spectroscopy of a metal-molecule-metal bridge, Science **302**, 77–81 (2003)
24.57 J. Repp, G. Meyer, S. Paavilainen, F.E. Olsson, M. Persson: Imaging bond formation between a gold atom and pentacene on an insulating surface, Science **312**, 1196–1199 (2006)
24.58 R. Yamachika, M. Grobis, A. Wachowiak, M.F. Crommie: Controlled atomic doping of a single C-60 molecule, Science **304**, 281–284 (2004)
24.59 S.W. Hla, L. Bartels, G. Meyer, K.H. Rieder: Inducing all steps of a chemical reaction with the scanning tunneling microscope tip: Towards single molecule engineering, Phys. Rev. Lett. **85**, 2777–2780 (2000)
24.60 Y. Kim, T. Komeda, M. Kawai: Single-molecule reaction and characterization by vibrational excitation, Phys. Rev. Lett. **89**, 126104 (2002)
24.61 M. Berthe, R. Stiufiuc, B. Grandidier, D. Deresmes, C. Delerue, D. Stievenard: Probing the carrier capture rate of a single quantum level, Science **319**, 436–438 (2008)
24.62 N. Neel, J. Kroeger, L. Limot, K. Palotas, W.A. Hofer, R. Berndt: Conductance and Kondo effect in a controlled single-atom contact, Phys. Rev. Lett. **98**, 016801 (2007)
24.63 J. Tersoff, D.R. Hamann: Theory and application for the scanning tunneling microscope, Phys. Rev. Lett. **50**, 1998–2001 (1983)
24.64 C.J. Chen: *Introduction to Scanning Tunneling Microscopy* (Oxford Univ. Press, Oxford 2007)
24.65 J. Wintterlin, J. Wiechers, H. Brune, T. Gritsch, H. Hofer, R.J. Behm: Atomic-resolution imaging of close-packed metal surfaces by scanning tunneling microscopy, Phys. Rev. Lett. **62**, 59–62 (1989)
24.66 J.A. Stroscio, R.M. Feenstra, A.P. Fein: Electronic structure of the Si (111) 2×1 surface by scanning-tunneling microscopy, Phys. Rev. Lett. **57**, 2579–2582 (1986)
24.67 A.L.V. de Parga, O.S. Hernan, R. Miranda, A.L. Yeyati, N. Mingo, A. Martin-Rodero, F. Flores: Electron resonances in sharp tips and their role in tunneling spectroscopy, Phys. Rev. Lett. **80**, 357–360 (1998)
24.68 S.H. Pan, E.W. Hudson, J.C. Davis: Vacuum tunneling of superconducting quasiparticles from atomically sharp scanning tunneling microscope tips, Appl. Phys. Lett. **73**, 2992–2994 (1998)
24.69 J.T. Li, W.D. Schneider, R. Berndt, O.R. Bryant, S. Crampin: Surface-state lifetime measured by scanning tunneling spectroscopy, Phys. Rev. Lett. **81**, 4464–4467 (1998)
24.70 L. Burgi, O. Jeandupeux, H. Brune, K. Kern: Probing hot-electron dynamics at surfaces with a cold scanning tunneling microscope, Phys. Rev. Lett. **82**, 4516–4519 (1999)
24.71 J.W.G. Wildoer, C. Harmans, H. van Kempen: Observation of Landau levels at the InAs(110) surface by scanning tunneling spectroscopy, Phys. Rev. B **55**, 16013–16016 (1997)
24.72 M. Morgenstern, V. Gudmundsson, R. Dombrowski, C. Wittneven, R. Wiesendanger: Nonlocality of the exchange interaction probed by scanning tunneling spectroscopy, Phys. Rev. B **201301**, 63 (2001)
24.73 L. Limot, E. Pehlke, J. Kroger, R. Berndt: Surface-state localization at adatoms, Phys. Rev. Lett. **94**, 036805 (2005)
24.74 F.E. Olsson, M. Persson, A.G. Borisov, J.P. Gauyacq, J. Lagoute, S. Folsch: Localization of the Cu(111) surface state by single Cu adatoms, Phys. Rev. Lett. **93**, 206803 (2004)
24.75 N. Nilius, T.M. Wallis, M. Persson, W. Ho: Distance dependence of the interaction between single atoms: Gold dimers on NiAl(110), Phys. Rev. Lett. **90**, 196103 (2003)

24.76 H.J. Lee, W. Ho, M. Persson: Spin splitting of s and p states in single atoms and magnetic coupling in dimers on a surface, Phys. Rev. Lett. **92**, 186802 (2004)

24.77 M.V. Grishin, F.I. Dalidchik, S.A. Kovalevskii, N.N. Kolchenko, B.R. Shub: Isotope effect in the vibrational spectra of water measured in experiments with a scanning tunneling microscope, Jetp Lett. **66**, 37–40 (1997)

24.78 Y. Sainoo, Y. Kim, T. Okawa, T. Komeda, H. Shigekawa, M. Kawai: Excitation of molecular vibrational modes with inelastic scanning tunneling microscopy processes: Examination through action spectra of cis-2-butene on Pd(110), Phys. Rev. Lett. **95**, 246102 (2005)

24.79 X.H. Qiu, G.V. Nazin, W. Ho: Vibronic states in single molecule electron transport, Phys. Rev. Lett. **92**, 2016102 (2004)

24.80 A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge 1993)

24.81 J.T. Li, W.D. Schneider, R. Berndt, B. Delley: Kondo scattering observed at a single magnetic impurity, Phys. Rev. Lett. **80**, 2893–2896 (1998)

24.82 V. Madhavan, W. Chen, T. Jamneala, M.F. Crommie, N.S. Wingreen: Tunneling into a single magnetic atom: Spectroscopic evidence of the Kondo resonance, Science **280**, 567–569 (1998)

24.83 T.W. Odom, J.L. Huang, C.L. Cheung, C.M. Lieber: Magnetic clusters on single-walled carbon nanotubes: The Kondo effect in a one-dimensional host, Science **290**, 1549–1552 (2000)

24.84 M. Ouyang, J.L. Huang, C.L. Cheung, C.M. Lieber: Energy gaps in "metallic" single-walled carbon nanotubes, Science **292**, 702–705 (2001)

24.85 U. Fano: Effects of configuration interaction on intensities and phase shifts, Phys. Rev. **124**, 1866–1878 (1969)

24.86 P. Wahl, L. Diekhoner, M.A. Schneider, L. Vitali, G. Wittich, K. Kern: Kondo temperature of magnetic impurities at surfaces, Phys. Rev. Lett. **93**, 176603 (2004)

24.87 Y.-S. Fu, S.-H. Ji, X. Chen, X.-C. Ma, R. Wu, C.-C. Wang, W.-H. Duan, X.-H. Qiu, B. Sun, P. Zhang, J.-F. Jia, Q.-K. Xue: Manipulating the Kondo resonance through quantum size effects, Phys. Rev. Lett. **99**, 256601 (2007)

24.88 J. Henzl, K. Morgenstern: Contribution of the surface state to the observation of the surface Kondo resonance, Phys. Rev. Lett. **98**, 266601 (2007)

24.89 P. Wahl, P. Simon, L. Diekhoener, V.S. Stepanyuk, P. Bruno, M.A. Schneider, K. Kern: Exchange interaction between single magnetic adatoms, Phys. Rev. Lett. **98**, 056601 (2007)

24.90 T. Jamneala, V. Madhavan, M.F. Crommie: Kondo response of a single antiferromagnetic chromium trimer, Phys. Rev. Lett. **87**, 256804 (2001)

24.91 P. Wahl, L. Diekhoner, G. Wittich, L. Vitali, M.A. Schneider, K. Kern: Kondo effect of molecular complexes at surfaces: Ligand control of the local spin coupling, Phys. Rev. Lett. **95**, 166601 (2005)

24.92 A.D. Zhao, Q.X. Li, L. Chen, H.J. Xiang, W.H. Wang, S. Pan, B. Wang, X.D. Xiao, J.L. Yang, J.G. Hou, Q.S. Zhu: Controlling the Kondo effect of an adsorbed magnetic ion through its chemical bonding, Science **309**, 1542–1544 (2005)

24.93 V. Iancu, A. Deshpande, S.-W. Hla: Manipulation of the Kondo effect via two-dimensional molecular assembly, Phys. Rev. Lett. **97**, 266603 (2006)

24.94 L. Gao, W. Ji, Y.B. Hu, Z.H. Cheng, Z.T. Deng, Q. Liu, N. Jiang, X. Lin, W. Guo, S.X. Du, W.A. Hofer, X.C. Xie, H.J. Gao: Site-specific Kondo effect at ambient temperatures in iron-based molecules, Phys. Rev. Lett. **99**, 106402 (2007)

24.95 H.C. Manoharan, C.P. Lutz, D.M. Eigler: Quantum mirages formed by coherent projection of electronic structure, Nature **403**, 512–515 (2000)

24.96 H.A. Mizes, J.S. Foster: Long-range electronic perturbations caused by defects using scanning tunneling microscopy, Science **244**, 559–562 (1989)

24.97 P.T. Sprunger, L. Petersen, E.W. Plummer, E. Laegsgaard, F. Besenbacher: Giant friedel oscillations on the beryllium(0001) surface, Science **275**, 1764–1767 (1997)

24.98 P. Hofmann, B.G. Briner, M. Doering, H.P. Rust, E.W. Plummer, A.M. Bradshaw: Anisotropic two-dimensional Friedel oscillations, Phys. Rev. Lett. **79**, 265–268 (1997)

24.99 E.J. Heller, M.F. Crommie, C.P. Lutz, D.M. Eigler: Scattering and absorption of surface electron waves in quantum corrals, Nature **369**, 464–466 (1994)

24.100 M. van der Wielen, A.J.A. van Roij, H. van Kempen: Direct observation of Friedel oscillations around incorporated sie, dopants in GaAs by low-temperature scanning tunneling microscopy, Phys. Rev. Lett. **76**, 1075–1078 (1996)

24.101 T. Maltezopoulos, A. Bolz, C. Meyer, C. Heyn, W. Hansen, M. Morgenstern, R. Wiesendanger: Wave-function mapping of InAs quantum dots by scanning tunneling spectroscopy, Phys. Rev. Lett. **91**, 196804 (2003)

24.102 O. Millo, D. Katz, Y.W. Cao, U. Banin: Imaging and spectroscopy of artificial-atom states in core/shell nanocrystal quantum dots, Phys. Rev. Lett. **86**, 5751–5754 (2001)

24.103 R. Temirov, S. Soubatch, A. Luican, F.S. Tautz: Free-electron-like dispersion in an organic monolayer film on a metal substrate, Nature **444**, 350–353 (2006)

24.104 K. Suzuki, K. Kanisawa, C. Janer, S. Perraud, K. Takashina, T. Fujisawa, Y. Hirayama: Spatial imaging of two-dimensional electronic states in semiconductor quantum wells, Phys. Rev. Lett. **98**, 136802 (2007)

24.105 S.G. Lemay, J.W. Janssen, M. van den Hout, M. Mooij, M.J. Bronikowski, P.A. Willis, R.E. Smalley, L.P. Kouwenhoven, C. Dekker: Two-dimensional imaging of electronic wavefunctions in carbon nanotubes, Nature **412**, 617–620 (2001)

24.106 L.C. Venema, J.W.G. Wildoer, J.W. Janssen, S.J. Tans, H. Tuinstra, L.P. Kouwenhoven, C. Dekker: Imaging electron wave functions of quantized energy levels in carbon nanotubes, Science **283**, 52–55 (1999)

24.107 C.R. Moon, L.S. Mattos, B.K. Foster, G. Zeltzer, W. Ko, H.C. Manoharan: Quantum phase extraction in isospectral electronic nanostructures, Science **319**, 782–787 (2008)

24.108 X.H. Lu, M. Grobis, K.H. Khoo, S.G. Louie, M.F. Crommie: Spatially mapping the spectral density of a single C-60 molecule, Phys. Rev. Lett. **90**, 096802 (2003)

24.109 F. Marczinowski, J. Wiebe, J.M. Tang, M.E. Flatte, F. Meier, M. Morgenstern, R. Wiesendanger: Local electronic structure near mn acceptors in InAs: Surface-induced symmetry breaking and coupling to host states, Phys. Rev. Lett. **99**, 157202 (2007)

24.110 A.M. Yakunin, A.Y. Silov, P.M. Koenraad, J.H. Wolter, W. Van Roy, J. De Boeck, J.M. Tang, M.E. Flatte: Spatial structure of an individual Mn acceptor in GaAs, Phys. Rev. Lett. **92**, 216806 (2004)

24.111 J.N. Crain, D.T. Pierce: End states in one-dimensional atom chains, Science **307**, 703–706 (2005)

24.112 N. Nilius, T.M. Wallis, W. Ho: Development of one-dimensional band structure in artificial gold chains, Science **297**, 1853–1856 (2002)

24.113 D. Kitchen, A. Richardella, J.-M. Tang, M.E. Flatte, A. Yazdani: Atom-by-atom substitution of Mn in GaAs and visualization of their hole-mediated interactions, Nature **442**, 436–439 (2006)

24.114 C. Wittneven, R. Dombrowski, M. Morgenstern, R. Wiesendanger: Scattering states of ionized dopants probed by low temperature scanning tunneling spectroscopy, Phys. Rev. Lett. **81**, 5616–5619 (1998)

24.115 D. Haude, M. Morgenstern, I. Meinel, R. Wiesendanger: Local density of states of a three-dimensional conductor in the extreme quantum limit, Phys. Rev. Lett. **86**, 1582–1585 (2001)

24.116 R. Joynt, R.E. Prange: Conditions for the quantum hall effect, Phys. Rev. B **29**, 3303–3317 (1984)

24.117 M. Morgenstern, J. Klijn, C. Meyer, R. Wiesendanger: Real-space observation of drift states in a two-dimensional electron system at high magnetic fields, Phys. Rev. Lett. **90**, 056804 (2003)

24.118 M. Morgenstern, J. Klijn, C. Meyer, M. Getzlaff, R. Adelung, R.A. Romer, K. Rossnagel, L. Kipp, M. Skibowski, R. Wiesendanger: Direct comparison between potential landscape and local density of states in a disordered two-dimensional electron system, Phys. Rev. Lett. **89**, 136806 (2002)

24.119 E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan: Scaling theory of localization: Absence of quantum diffusion in two dimensions, Phys. Rev. **42**, 673–676 (1979)

24.120 C. Meyer, J. Klijn, M. Morgenstern, R. Wiesendanger: Direct measurement of the local density of states of a disordered one-dimensional conductor, Phys. Rev. Lett. **91**, 076803 (2003)

24.121 N. Oncel, A. van Houselt, J. Huijben, A.S. Hallback, O. Gurlu, H.J.W. Zandvliet, B. Poelsema: Quantum confinement between self-organized Pt nanowires on Ge(001), Phys. Rev. Lett. **95**, 116801 (2005)

24.122 J. Lee, S. Eggert, H. Kim, S.J. Kahng, H. Shinohara, Y. Kuk: Real space imaging of one-dimensional standing waves: Direct evidence for a Luttinger liquid, Phys. Rev. Lett. **93**, 166403 (2004)

24.123 R.E. Peierls: *Quantum Theory of Solids* (Clarendon, Oxford 1955)

24.124 C.G. Slough, W.W. McNairy, R.V. Coleman, B. Drake, P.K. Hansma: Charge-density waves studied with the use of a scanning tunneling microscope, Phys. Rev. B **34**, 994–1005 (1986)

24.125 X.L. Wu, C.M. Lieber: Hexagonal domain-like charge density wave phase of TaS_2 determined by scanning tunneling microscopy, Science **243**, 1703–1705 (1989)

24.126 T. Nishiguchi, M. Kageshima, N. Ara-Kato, A. Kawazu: Behavior of charge density waves in a one-dimensional organic conductor visualized by scanning tunneling microscopy, Phys. Rev. Lett. **81**, 3187–3190 (1998)

24.127 X.L. Wu, C.M. Lieber: Direct observation of growth and melting of the hexagonal-domain charge-density-wave phase in 1T-TaS_2 by scanning tunneling microscopy, Phys. Rev. Lett. **64**, 1150–1153 (1990)

24.128 J.M. Carpinelli, H.H. Weitering, E.W. Plummer, R. Stumpf: Direct observation of a surface charge density wave, Nature **381**, 398–400 (1996)

24.129 H.H. Weitering, J.M. Carpinelli, A.P. Melechko, J.D. Zhang, M. Bartkowiak, E.W. Plummer: Defect-mediated condensation of a charge density wave, Science **285**, 2107–2110 (1999)

24.130 S. Modesti, L. Petaccia, G. Ceballos, I. Vobornik, G. Panaccione, G. Rossi, L. Ottaviano, R. Larciprete, S. Lizzit, A. Goldoni: Insulating ground state of Sn/Si(111)-(root 3 × root 3)R30 degrees, Phys. Rev. Lett. **98**, 126401 (2007)

24.131 K. Swamy, A. Menzel, R. Beer, E. Bertel: Charge-density waves in self-assembled halogen-bridged metal chains, Phys. Rev. Lett. **86**, 1299–1302 (2001)

24.132 H.W. Yeom, S. Takeda, E. Rotenberg, I. Matsuda, K. Horikoshi, J. Schaefer, C.M. Lee, S.D. Kevan, T. Ohta, T. Nagao, S. Hasegawa: Instability and charge density wave of metallic quantum chains on a silicon surface, Phys. Rev. Lett. **82**, 4898–4901 (1999)

24.133 J.J. Kim, W. Yamaguchi, T. Hasegawa, K. Kitazawa: Observation of mott localization gap using low temperature scanning tunneling spectroscopy in commensurate 1 T–T a Sa 2, Phys. Rev. Lett. **73**, 2103–2106 (1994)

24.134 A. Wachowiak, R. Yamachika, K.H. Khoo, Y. Wang, M. Grobis, D.H. Lee, S.G. Louie, M.F. Crommie: Visualization of the molecular Jahn–Teller effect in an insulating K_4C_{60} monolayer, Science **310**, 468–470 (2005)

24.135 J. Bardeen, L.N. Cooper, J.R. Schrieffer: Theory of superconductivity, Phys. Rev. **108**, 1175–1204 (1957)

24.136 A. Yazdani, B.A. Jones, C.P. Lutz, M.F. Crommie, D.M. Eigler: Probing the local effects of magnetic impurities on superconductivity, Science **275**, 1767–1770 (1997)

24.137 K. Inoue, H. Takayanagi: Local tunneling spectroscopy of a Nb/InAs/Nb superconducting proximity system with a scanning tunneling microscope, Phys. Rev. B **43**, 6214–6215 (1991)

24.138 S.H. Tessmer, M.B. Tarlie, D.J. VanHarlingen, D.L. Maslov, P.M. Goldbart: Probing the superconducting proximity effect in $NbSe_2$ by scanning tunneling microscopy, Phys. Rev. Lett. **77**, 924–927 (1996)

24.139 H.F. Hess, R.B. Robinson, R.C. Dynes, J.M. Valles, J.V. Waszczak: Scanning-tunneling-microscope observation of the Abrikosov flux lattice and the density of states near and inside a fluxoid, Phys. Rev. Lett. **62**, 214–216 (1989)

24.140 H.F. Hess, R.B. Robinson, J.V. Waszczak: Vortex-core structure observed with a scanning tunneling microscope, Phys. Rev. Lett. **64**, 2711–2714 (1990)

24.141 N. Hayashi, M. Ichioka, K. Machida: Star-shaped local density of states around vortices in a type-II superconductor, Phys. Rev. Lett. **77**, 4074–4077 (1996)

24.142 H. Sakata, M. Oosawa, K. Matsuba, N. Nishida, H. Takeya, K. Hirata: Imaging of a vortex lattice transition in YNi_2B_2C by scanning tunneling spectroscopy, Phys. Rev. Lett. **84**, 1583–1586 (2000)

24.143 S. Behler, S.H. Pan, P. Jess, A. Baratoff, H.J. Guntherodt, F. Levy, G. Wirth, J. Wiesner: Vortex pinning in ion-irradiated $NbSe_2$ studied by scanning tunneling microscopy, Phys. Rev. Lett. **72**, 1750–1753 (1994)

24.144 R. Berthe, U. Hartmann, C. Heiden: Influence of a transport current on the Abrikosov flux lattice observed with a low-temperature scanning tunneling microscope, Ultramicroscopy **42**, 696–698 (1992)

24.145 N.C. Yeh, C.T. Chen, G. Hammerl, J. Mannhart, A. Schmehl, C.W. Schneider, R.R. Schulz, S. Tajima, K. Yoshida, D. Garrigus, M. Strasik: Evidence of doping-dependent pairing symmetry in cuprate superconductors, Phys. Rev. Lett. **87**, 087003 (2001)

24.146 K. McElroy, R.W. Simmonds, J.E. Hoffman, D.H. Lee, J. Orenstein, H. Eisaki, S. Uchida, J.C. Davis: Relating atomic-scale electronic phenomena to wave-like quasiparticle states in superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$, Nature **422**, 592–596 (2003)

24.147 K. McElroy, J. Lee, J.A. Slezak, D.H. Lee, H. Eisaki, S. Uchida, J.C. Davis: Atomic-scale sources and mechanism of nanoscale electronic disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$, Science **309**, 1048–1052 (2005)

24.148 S.H. Pan, E.W. Hudson, K.M. Lang, H. Eisaki, S. Uchida, J.C. Davis: Imaging the effects of individual zinc impurity atoms on superconductivity in $Bi_2Sr_2CaCu_2O_{8+\delta}$, Nature **403**, 746–750 (2000)

24.149 A. Polkovnikov, S. Sachdev, M. Vojta: Impurity in a d-wave superconductor: Kondo effect and STM spectra, Phys. Rev. Lett. **86**, 296–299 (2001)

24.150 A.N. Pasupathy, A. Pushp, K.K. Gomes, C.V. Parker, J. Wen, Z. Xu, G. Gu, S. Ono, Y. Ando, A. Yazdani: Electronic origin of the inhomogeneous pairing interaction in the high-T-c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, Science **320**, 196–201 (2008)

24.151 I. Maggioaprile, C. Renner, A. Erb, E. Walker, O. Fischer: Direct vortex lattice imaging and tunneling spectroscopy of flux lines on $YBa_2Cu_3O_{7-\delta}$, Phys. Rev. Lett. **75**, 2754–2757 (1995)

24.152 S.H. Pan, E.W. Hudson, A.K. Gupta, K.W. Ng, H. Eisaki, S. Uchida, J.C. Davis: STM studies of the electronic structure of vortex cores in $Bi_2Sr_2CaCu_2O_{8+\delta}$, Phys. Rev. Lett. **85**, 1536–1539 (2000)

24.153 C. Renner, B. Revaz, K. Kadowaki, I. Maggio-Aprile, O. Fischer: Observation of the low temperature pseudogap in the vortex cores of $Bi_2Sr_2CaCu_2O_{8+\delta}$, Phys. Rev. Lett. **80**, 3606–3609 (1998)

24.154 D.P. Arovas, A.J. Berlinsky, C. Kallin, S.C. Zhang: Superconducting vortex with antiferromagnetic core, Phys. Rev. Lett. **79**, 2871–2874 (1997)

24.155 J.E. Hoffman, E.W. Hudson, K.M. Lang, V. Madhavan, H. Eisaki, S. Uchida, J.C. Davis: A four unit cell periodic pattern of quasi-particle states surrounding vortex cores in $Bi_2Sr_2CaCu_2O_{8+\delta}$, Science **295**, 466–469 (2002)

24.156 M. Vershinin, S. Misra, S. Ono, Y. Abe, Y. Ando, A. Yazdani: Local ordering in the pseudogap state of the high-T-c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, Science **303**, 1995–1998 (2004)

24.157 T. Hanaguri, C. Lupien, Y. Kohsaka, D.H. Lee, M. Azuma, M. Takano, H. Takagi, J.C. Davis: A 'checkerboard' electronic crystal state in lightly hole-doped $Ca_{2-x}Na_xCuO_2Cl_2$, Nature **430**, 1001–1005 (2004)

24.158 Y. Kohsaka, C. Taylor, K. Fujita, A. Schmidt, C. Lupien, T. Hanaguri, M. Azuma, M. Takano, H. Eisaki, H. Takagi, S. Uchida, J.C. Davis: An intrinsic bond-centered electronic glass with unidirectional domains in underdoped cuprates, Science **315**, 1380–1385 (2007)

24.159 O. Naaman, W. Teizer, R.C. Dynes: Fluctuation dominated Josephson tunneling with a scanning tunneling microscope, Phys. Rev. Lett. **87**, 097004 (2001)

24.160 M. Bode, M. Getzlaff, R. Wiesendanger: Spin-polarized vacuum tunneling into the exchange-split surface state of Gd(0001), Phys. Rev. Lett. **81**, 4256–4259 (1998)

24.161 A. Kubetzka, M. Bode, O. Pietzsch, R. Wiesendanger: Spin-polarized scanning tunneling microscopy with antiferromagnetic probe tips, Phys. Rev. Lett. **88**, 057201 (2002)

24.162 O. Pietzsch, A. Kubetzka, M. Bode, R. Wiesendanger: Observation of magnetic hysteresis at the nanometer scale by spin-polarized scanning tunneling spectroscopy, Science **292**, 2053–2056 (2001)

24.163 S. Heinze, M. Bode, A. Kubetzka, O. Pietzsch, X. Nie, S. Blugel, R. Wiesendanger: Real-space imaging of two-dimensional antiferromagnetism on the atomic scale, Science **288**, 1805–1808 (2000)

24.164 A. Kubetzka, P. Ferriani, M. Bode, S. Heinze, G. Bihlmayer, K. von Bergmann, O. Pietzsch, S. Blugel, R. Wiesendanger: Revealing antiferromagnetic order of the Fe monolayer on W(001): Spin-polarized scanning tunneling microscopy and first-principles calculations, Phys. Rev. Lett. **94**, 087204 (2005)

24.165 A. Wachowiak, J. Wiebe, M. Bode, O. Pietzsch, M. Morgenstern, R. Wiesendanger: Direct observation of internal spin structure of magnetic vortex cores, Science **298**, 577–580 (2002)

24.166 M. Bode, M. Heide, K. von Bergmann, P. Ferriani, S. Heinze, G. Bihlmayer, A. Kubetzka, O. Pietzsch, S. Bluegel, R. Wiesendanger: Chiral magnetic order at surfaces driven by inversion asymmetry, Nature **447**, 190–193 (2007)

24.167 C.L. Gao, U. Schlickum, W. Wulfhekel, J. Kirschner: Mapping the surface spin structure of large unit cells: Reconstructed Mn films on Fe(001), Phys. Rev. Lett. **98**, 107203 (2007)

24.168 K. Von Bergmann, S. Heinze, M. Bode, E.Y. Vedmedenko, G. Bihlmayer, S. Blugel, R. Wiesendanger: Observation of a complex nanoscale magnetic structure in a hexagonal Fe monolayer, Phys. Rev. Lett. **96**, 167203 (2006)

24.169 Y. Yayon, V.W. Brar, L. Senapati, S.C. Erwin, M.F. Crommie: Observing spin polarization of individual magnetic adatoms, Phys. Rev. Lett. **99**, 067202 (2007)

24.170 M. Bode, O. Pietzsch, A. Kubetzka, R. Wiesendanger: Shape-dependent thermal switching behavior of superparamagnetic nanoislands, Phys. Rev. Lett. **92**, 067201 (2004)

24.171 S. Krause, L. Berbil-Bautista, G. Herzog, M. Bode, R. Wiesendanger: Current-induced magnetization switching with a spin-polarized scanning tunneling microscope, Science **317**, 1537–1540 (2007)

24.172 P.J. Eaton, P. West: *Atomic Force Microscopy* (Oxford Univ. Press, Oxford 2010)

24.173 R. Garcia, R. Perez: Dynamic atomic force microscopy methods, Surf. Sci. Rep. **47**, 197–301 (2002)

24.174 U. Dürig: Extracting interaction forces and complementary observables in dynamic probe microscopy, Appl. Phys. Lett. **76**, 1203–1205 (2000)

24.175 F.J. Giessibl: A direct method to calculate tip–sample forces from frequency shifts in frequency-modulation atomic force microscopy, Appl. Phys. Lett. **78**, 123–125 (2001)

24.176 B. Gotsmann, B. Anczykowski, C. Seidel, H. Fuchs: Determination of tip–sample interaction forces from measured dynamic force spectroscopy curves, Appl. Surf. Sci. **140**, 314–319 (1999)

24.177 J.E. Sader, S.P. Jarvis: Accurate formulas for interaction force and energy in frequency modulation force spectroscopy, Appl. Phys. Lett. **84**, 1801–1803 (2004)

24.178 F.J. Giessibl: Atomic resolution of the silicon (111)-(7×7) surface by atomic force microscopy, Science **267**, 68–71 (1995)

24.179 M.A. Lantz, H.J. Hug, P.J.A. van Schendel, R. Hoffmann, S. Martin, A. Baratoff, A. Abdurixit, H.J. Guntherodt, C. Gerber: Low temperature scanning force microscopy of the Si(111)-(7×7) surface, Phys. Rev. Lett. **84**, 2642–2645 (2000)

24.180 N. Suehira, Y. Sugawara, S. Morita: Artifact and fact of Si(111)7 x 7 surface images observed with a low temperature noncontact atomic force microscope (LT-NC-AFM), Jpn. J. Appl. Phys. Part 2 Lett. **40**, L292–L294 (2001)

24.181 K. Suzuki, M. Iwatsuki, S. Kitamura, C.B. Mooney: Development of low temperature ultrahigh vacuum atomic force microscope/scanning tunneling microscope, Jpn. J. Appl. Phys. Part 1 Regul. Pap. Short Notes Rev. Pap. **39**, 3750–3752 (2000)

24.182 T. Uozumi, Y. Tomiyoshi, N. Suehira, Y. Sugawara, S. Morita: Observation of Si(100) surface with noncontact atomic force microscope at 5 K, Appl. Surf. Sci. **188**, 279–284 (2002)

24.183 Y.J. Li, H. Nomura, N. Ozaki, Y. Naitoh, M. Kageshima, Y. Sugawara, C. Hobbs, L. Kantorovich: Origin of $p(2\times1)$ phase on Si(001) by noncontact atomic force microscopy at 5 K, Phys. Rev. Lett. **96**, 106104 (2006)

24.184 A. Sweetman, S. Jarvis, R. Danza, J. Bamidele, S. Gangopadhyay, G.A. Shaw, L. Kantorovich, P. Moriarty: Toggling bistable atoms via mechanical switching of bond angle, Phys. Rev. Lett. **106**, 136101 (2011)

24.185 A. Sweetman, S. Jarvis, R. Danza, J. Bamidele, L. Kantorovich, P. Moriarty: Manipulating Si(100) at 5 K using qPlus frequency modulated atomic force microscopy: Role of defects and dynamics in the mechanical switching of atoms, Phys. Rev. B **84**, 085426 (2011)

24.186 A. Schwarz, W. Allers, U.D. Schwarz, R. Wiesendanger: Simultaneous imaging of the in and as sublattice on InAs(110)-(1×1) with dynamic scanning force microscopy, Appl. Surf. Sci. **140**, 293–297 (1999)

24.187 A. Schwarz, W. Allers, U.D. Schwarz, R. Wiesendanger: Dynamic-mode scanning force microscopy study of n-InAs(110)-(1×1) at low temperatures, Phys. Rev. B **61**, 2837–2845 (2000)

24.188 A. Sweetman, A. Stannard, Y. Sugimoto, M. Abe, S. Morita, P. Moriarty: Simultaneous noncontact AFM and STM of Ag:Si(111)-$(\sqrt{3}\times\sqrt{3})$R30°, Phys. Rev. B **87**, 075310 (2013)

24.189 W. Allers, A. Schwarz, U.D. Schwarz, R. Wiesendanger: Dynamic scanning force microscopy at low temperatures on a van der Waals surface: graphite (0001), Appl. Surf. Sci. **140**, 247–252 (1999)

24.190 M. Ashino, A. Schwarz, H. Holscher, U.D. Schwarz, R. Wiesendanger: Interpretation of the atomic scale contrast obtained on graphite and single-walled carbon nanotubes in the dynamic mode of atomic force microscopy, Nanotechnology **16**, S134–S137 (2005)

24.191 S. Hembacher, F.J. Giessibl, J. Mannhart, C.F. Quate: Local spectroscopy and atomic imaging of tunneling current, forces, and dissipation on graphite, Phys. Rev. Lett. **94**, 056101

(2005)
24.192 H. Holscher, W. Allers, U.D. Schwarz, A. Schwarz, R. Wiesendanger: Interpretation of "true atomic resolution" images of graphite (0001) in noncontact atomic force microscopy, Phys. Rev. B **62**, 6967–6970 (2000)
24.193 W. Allers, A. Schwarz, U.D. Schwarz, R. Wiesendanger: Dynamic scanning force microscopy at low temperatures on a noble-gas crystal: Atomic resolution on the xenon(111) surface, Europhys. Lett. **48**, 276–279 (1999)
24.194 M.Z. Baykara, H. Moenig, T.C. Schwendemann, O. Uenverdi, E.I. Altman, U.D. Schwarz: Three-dimensional interaction force and tunneling current spectroscopy of point defects on rutile TiO_2(110), Appl. Phys. Lett. **108**, 071601 (2016)
24.195 A. Yurtsever, Y. Sugimoto, M. Abe, S. Morita: NC-AFM imaging of the Ti0(2)(110)-(1 x 1) surface at low temperature, Nanotechnology **21**, 165702 (2010)
24.196 R. Hoffmann, M.A. Lantz, H.J. Hug, P.J.A. van Schendel, P. Kappenberger, S. Martin, A. Baratoff, H.J. Guntherodt: Atomic resolution imaging and force versus distance measurements on KBr (001) using low temperature scanning force microscopy, Appl. Surf. Sci. **188**, 238–244 (2002)
24.197 D.G. de Oteyza, P. Gorman, Y.-C. Chen, S. Wickenburg, A. Riss, D.J. Mowbray, G. Etkin, Z. Pedramrazi, H.-Z. Tsai, A. Rubio, M.F. Crommie, F.R. Fischer: Direct imaging of covalent bond structure in single-molecule chemical reactions, Science **340**, 1434–1437 (2013)
24.198 L. Gross: Recent advances in submolecular resolution with scanning probe microscopy, Nat. Chem. **3**, 273–278 (2011)
24.199 L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer: The chemical structure of a molecule resolved by atomic force microscopy, Science **325**, 1110–1114 (2009)
24.200 L. Gross, F. Mohn, N. Moll, G. Meyer, R. Ebel, W.M. Abdel-Mageed, M. Jaspars: Organic structure determination using atomic-resolution scanning probe microscopy, Nat. Chem. **2**, 821–825 (2010)
24.201 L. Gross, F. Mohn, N. Moll, B. Schuler, A. Criado, E. Guitian, D. Pena, A. Gourdon, G. Meyer: Bond-order discrimination by atomic force microscopy, Science **337**, 1326–1329 (2012)
24.202 B. Schuler, G. Meyer, D. Pena, O.C. Mullins, L. Gross: Unraveling the molecular structures of asphaltenes by atomic force microscopy, J. Am. Chem. Soc. **137**, 9870–9876 (2015)
24.203 A.M. Sweetman, S.P. Jarvis, H.Q. Sang, I. Lekkas, P. Rahe, Y. Wang, J.B. Wang, N.R. Champness, L. Kantorovich, P. Moriarty: Mapping the force field of a hydrogen-bonded assembly, Nat. Commun. **5**, 3931 (2014)
24.204 P. Hapala, G. Kichin, C. Wagner, F.S. Tautz, R. Temirov, P. Jelinek: Mechanism of high-resolution STM/AFM imaging with functionalized tips, Phys. Rev. B **90**, 085421 (2014)
24.205 N. Moll, L. Gross, F. Mohn, A. Curioni, G. Meyer: The mechanisms underlying the enhanced resolution of atomic force microscopy with functionalized tips, New J. Phys. **12**, 125020 (2010)
24.206 F. Mohn, L. Gross, N. Moll, G. Meyer: Imaging the charge distribution within a single molecule, Nat. Nanotechnol. **7**, 227–231 (2012)
24.207 B. Schuler, S.-X. Liu, Y. Geng, S. Decurtins, G. Meyer, L. Gross: Contrast formation in Kelvin probe force microscopy of single pi-conjugated molecules, Nano Lett. **14**, 3342–3346 (2014)
24.208 H. Holscher, W. Allers, U.D. Schwarz, A. Schwarz, R. Wiesendanger: Determination of tip–sample interaction potentials by dynamic force spectroscopy, Phys. Rev. Lett. **83**, 4780–4783 (1999)
24.209 H. Holscher, U.D. Schwarz, R. Wiesendanger: Calculation of the frequency shift in dynamic force microscopy, Appl. Surf. Sci. **140**, 344–351 (1999)
24.210 R. Hoffmann, C. Barth, A.S. Foster, A.L. Shluger, H.J. Hug, H.J. Guntherodt, R.M. Nieminen, M. Reichling: Measuring site-specific cluster-surface bond formation, J. Am. Chem. Soc. **127**, 17863–17866 (2005)
24.211 H. Holscher, A. Schwarz, W. Allers, U.D. Schwarz, R. Wiesendanger: Quantitative analysis of dynamic-force-spectroscopy data on graphite(0001) in the contact and noncontact regimes, Phys. Rev. B **61**, 12678–12681 (2000)
24.212 M.A. Lantz, R. Hoffmann, A.S. Foster, A. Baratoff, H.J. Hug, H.R. Hidber, H.J. Guntherodt: Site-specific force-distance characteristics on NaCl(001): Measurements versus atomistic simulations, Phys. Rev. B **74**, 245426 (2006)
24.213 M.A. Lantz, H.J. Hug, R. Hoffmann, P.J.A. van Schendel, P. Kappenberger, S. Martin, A. Baratoff, H.J. Guntherodt: Quantitative measurement of short-range chemical bonding forces, Science **291**, 2580–2583 (2001)
24.214 Z.X. Sun, M.P. Boneschanscher, I. Swart, D. Vanmaekelbergh, P. Liljeroth: Quantitative atomic force microscopy with carbon monoxide terminated tips, Phys. Rev. Lett. **106**, 046104 (2011)
24.215 Y. Sugimoto, P. Pou, M. Abe, P. Jelinek, R. Perez, S. Morita, O. Custance: Chemical identification of individual surface atoms by atomic force microscopy, Nature **446**, 64–67 (2007)
24.216 M. Abe, Y. Sugimoto, O. Custance, S. Morita: Atom tracking for reproducible force spectroscopy at room temperature with non-contact atomic force microscopy, Nanotechnology **16**, 3029–3034 (2005)
24.217 M.Z. Baykara, M. Todorovic, H. Monig, T.C. Schwendemann, O. Unverdi, L. Rodrigo, E.I. Altman, R. Perez, U.D. Schwarz: Atom-specific forces and defect identification on surface-oxidized Cu(100) with combined 3D-AFM and STM measurements, Phys. Rev. B **87**, 155414 (2013)
24.218 H. Holscher, S.M. Langkat, A. Schwarz, R. Wiesendanger: Measurement of three-dimensional force fields with atomic resolution using dynamic force spectroscopy, Appl. Phys. Lett. **81**, 4428–4430 (2002)
24.219 M.Z. Baykara, T.C. Schwendemann, B.J. Albers, N. Pilet, H. Monig, E.I. Altman, U.D. Schwarz: Exploring atomic-scale lateral forces in the at-

tractive regime: A case study on graphite (0001), Nanotechnology **23**, 405703 (2012)
24.220 M.Z. Baykara, T.C. Schwendemann, E.I. Altman, U.D. Schwarz: Three-dimensional atomic force microscopy-taking surface imaging to the next level, Adv. Mater. **22**, 2838–2853 (2010)
24.221 B.J. Albers, T.C. Schwendemann, M.Z. Baykara, N. Pilet, M. Liebmann, E.I. Altman, U.D. Schwarz: Data acquisition and analysis procedures for high-resolution atomic force microscopy in three dimensions, Nanotechnology **20**, 264002 (2009)
24.222 M.Z. Baykara, O.E. Dagdeviren, T.C. Schwendemann, H. Monig, E.I. Altman, U.D. Schwarz: Probing three-dimensional surface force fields with atomic resolution: Measurement strategies, limitations and artifact reduction, Beilstein J. Nanotechnol. **3**, 637–650 (2012)
24.223 S. Kawai, T. Glatzel, S. Koch, A. Baratoff, E. Meyer: Interaction-induced atomic displacements revealed by drift-corrected dynamic force spectroscopy, Phys. Rev. B **83**, 035421 (2011)
24.224 Y. Sugimoto, K. Ueda, M. Abe, S. Morita: Three-dimensional scanning force/tunneling spectroscopy at room temperature, J. Phys. Condens. Matter **24**, 084008 (2012)
24.225 S. Fremy, S. Kawai, R. Pawlak, T. Glatzel, A. Baratoff, E. Meyer: Three-dimensional dynamic force spectroscopy measurements on KBr(001): Atomic deformations at small tip–sample separations, Nanotechnology **23**, 055401 (2012)
24.226 R. Pawlak, S. Kawai, S. Fremy, T. Glatzel, E. Meyer: Atomic-scale mechanical properties of orientated C(60) molecules revealed by noncontact atomic force microscopy, ACS Nano **5**, 6349–6354 (2011)
24.227 N. Oyabu, O. Custance, I.S. Yi, Y. Sugawara, S. Morita: Mechanical vertical manipulation of selected single atoms by soft nanoindentation using near contact atomic force microscopy, Phys. Rev. Lett. **90**, 176102 (2003)
24.228 N. Oyabu, Y. Sugimoto, M. Abe, O. Custance, S. Morita: Lateral manipulation of single atoms at semiconductor surfaces using atomic force microscopy, Nanotechnology **16**, S112–S117 (2005)
24.229 J. Bamidele, S.H. Lee, Y. Kinoshita, R. Turansky, Y. Naitoh, Y.J. Li, Y. Sugawara, I. Stich, L. Kantorovich: Vertical atomic manipulation with dynamic atomic-force microscopy without tip change via a multi-step mechanism, Nat. Commun. **5**, 5476 (2014)
24.230 Y. Sugimoto, M. Abe, S. Hirayama, N. Oyabu, O. Custance, S. Morita: Atom inlays performed at room temperature using atomic force microscopy, Nat. Mater. **4**, 156–159 (2005)
24.231 Y. Sugimoto, P. Pou, O. Custance, P. Jelinek, M. Abe, R. Perez, S. Morita: Complex patterning by vertical interchange atom manipulation using atomic force microscopy, Science **322**, 413–417 (2008)
24.232 S. Kawai, A.S. Foster, F.F. Canova, H. Onodera, S.-I. Kitamura, E. Meyer: Atom manipulation on an insulating surface at room temperature, Nat. Commun. **5**, 5403 (2014)
24.233 S. Hirth, F. Ostendorf, M. Reichling: Lateral manipulation of atomic size defects on the CaF_2(111) surface, Nanotechnology **17**, S148–S154 (2006)
24.234 S. Kawai, M. Koch, E. Gnecco, A. Sadeghi, R. Pawlak, T. Glatzel, J. Schwarz, S. Goedecker, S. Hecht, A. Baratoff, L. Grill, E. Meyer: Quantifying the atomic-level mechanics of single long physisorbed molecular chains, Proc. Natl. Acad. Sci. U.S.A. **111**, 3968–3972 (2014)
24.235 G. Langewisch, J. Falter, H. Fuchs, A. Schirmeisen: Forces during the controlled displacement of organic molecules, Phys. Rev. Lett. **110**, 036101 (2013)
24.236 R. Pawlak, W. Ouyang, A.E. Filippov, L. Kalikhman-Razvozov, S. Kawai, T. Glatzel, E. Gnecco, A. Baratoff, Q. Zheng, O. Hod, M. Urbakh, E. Meyer: Single-molecule tribology: Force microscopy manipulation of a porphyrin derivative on a copper surface, ACS Nano **10**, 713–722 (2016)
24.237 R. Pawlak, S. Fremy, S. Kawai, T. Glatzel, H.J. Fang, L.A. Fendt, F. Diederich, E. Meyer: Directed rotations of single porphyrin molecules controlled by localized force spectroscopy, ACS Nano **6**, 6318–6324 (2012)
24.238 S. Kawai, A. Benassi, E. Gnecco, H. Soede, R. Pawlak, X. Feng, K. Muellen, D. Passerone, C.A. Pignedoli, P. Ruffieux, R. Fasel, E. Meyer: Superlubricity of graphene nanoribbons on gold surfaces, Science **351**, 957–961 (2016)
24.239 M. Nonnenmacher, M.P. Oboyle, H.K. Wickramasinghe: Kelvin probe force microscopy, Appl. Phys. Lett. **58**, 2921–2923 (1991)
24.240 S. Sadewasser, T. Glatzel: *Kelvin Probe Force Microscopy* (Springer, Berlin, Heidelberg 2012)
24.241 T. König, G.H. Simon, H.P. Rust, M. Heyde: Work function measurements of thin oxide films on metals-MgO on Ag(001), J. Phys. Chem. C **113**, 11301–11305 (2009)
24.242 K. Moloni, B.M. Moskowitz, E.D. Dahlberg: Domain structures in single crystal magnetite below the verwey transition as observed with a low-temperature magnetic force microscope, Geophys. Res. Lett. **23**, 2851–2854 (1996)
24.243 M. Liebmann, U. Kaiser, A. Schwarz, R. Wiesendanger, U.H. Pi, T.W. Noh, Z.G. Khim, D.W. Kim: Domain nucleation and growth of $La_{0.7}Ca_{0.3}MnO_{3-\delta}/LaAlO_3$ films studied by low temperature magnetic force microscopy, J. Appl. Phys. **93**, 8319–8321 (2003)
24.244 Q.Y. Lu, C.C. Chen, A. de Lozanne: Observation of magnetic domain behavior in colossal magnetoresistive materials with a magnetic force microscope, Science **276**, 2006–2008 (1997)
24.245 G.M. Xiao, J.H. Ross, A. Parasiris, K.D.D. Rathnayaka, D.G. Naugle: Low-temperature MFM studies of CMR manganites, Physica C **341**, 769–770 (2000)
24.246 A. Moser, H.J. Hug, I. Parashikov, B. Stiefel, O. Fritz, H. Thomas, K. Baratoff, H.J. Guntherodt, P. Chaudhari: Observation of single vortices condensed into a vortex-glass phase by magnetic force microscopy, Phys. Rev. Lett. **74**, 1847–1850 (1995)

24.247 U.H. Pi, D.H. Kim, Z.G. Khim, U. Kaiser, M. Liebmann, A. Schwarz, R. Wiesendanger: Vortex dynamics in $Bi_2Sr_2CaCu_2O_8$ single crystal with low density columnar defects studied by magnetic force microscope, J. Low Temp. Phys. **131**, 993–1002 (2003)

24.248 M. Roseman, P. Grutter: Estimating the magnetic penetration depth using constant-height magnetic force microscopy images of vortices, New J. Phys. **3**, 241–248 (2001)

24.249 M. Roseman, P. Grutter, A. Badia, V. Metlushko: Flux lattice imaging of a patterned niobium thin film, J. Appl. Phys. **89**, 6787–6789 (2001)

24.250 A. Volodin, K. Temst, C. Van Haesendonck, Y. Bruynseraede: Imaging of vortices in conventional superconductors by magnetic force microscopy, Physica C **332**, 156–159 (2000)

24.251 A. Volodin, K. Temst, C. Van Haesendonck, Y. Bruynseraede, M.I. Montero, I.K. Schuller: Magnetic-force microscopy of vortices in thin niobium films: Correlation between the vortex distribution and the thickness-dependent film morphology, Europhys. Lett. **58**, 582–588 (2002)

24.252 C.W. Yuan, Z. Zheng, A.L. de Lozanne, M. Tortonese, D.A. Rudman, J.N. Eckstein: Vortex images in thin films of $YBa_2Cu_3O_{7-x}$ and $Bi_2Sr_2CaCu_2O_{8+x}$ obtained by low-temperature magnetic force microscopy, J. Vac. Sci. Technol. B **14**, 1210–1213 (1996)

24.253 A. Moser, H.J. Hug, B. Stiefel, H.J. Guntherodt: Low temperature magnetic force microscopy on $YBa_2Cu_3O_{7-\delta}$ thin films, J. Magn. Magn. Mater. **190**, 114–123 (1998)

24.254 A. Volodin, K. Temst, C. Van Haesendonck, Y. Bruynseraede: Observation of the Abrikosov vortex lattice in $NbSe_2$ with magnetic force microscopy, Appl. Phys. Lett. **73**, 1134–1136 (1998)

24.255 A. Schwarz, R. Wiesendanger: Magnetic sensitive force microscopy, Nano Today **3**, 28–39 (2008)

24.256 R. Wiesendanger, D. Burgler, G. Tarrach, A. Wadas, D. Brodbeck, H.J. Guntherodt, G. Guntherodt, R.J. Gambino, R. Ruf: Vacuum tunneling of spin-polarized electrons detected by scanning tunneling microscopy, J. Vac. Sci. Technol. B **9**, 519–524 (1991)

24.257 H. Momida, T. Oguchi: First-principles study on exchange force image of NiO(001) surface using a ferromagnetic Fe probe, Surf. Sci. **590**, 42–50 (2005)

24.258 U. Kaiser, A. Schwarz, R. Wiesendanger: Magnetic exchange force microscopy with atomic resolution, Nature **446**, 522–525 (2007)

24.259 R. Schmidt, C. Lazo, U. Kaiser, A. Schwarz, S. Heinze, R. Wiesendanger: Quantitative measurement of the magnetic exchange interaction across a vacuum gap, Phys. Rev. Lett. **106**, 257202 (2011)

24.260 R. Schmidt, C. Lazo, H. Holscher, U.H. Pi, V. Caciuc, A. Schwarz, R. Wiesendanger, S. Heinze: Probing the magnetic exchange forces of iron on the atomic scale, Nano Lett. **9**, 200–204 (2008)

24.261 R. Schmidt, A. Schwarz, R. Wiesendanger: Magnetization switching utilizing the magnetic exchange interaction, Phys. Rev. B **86**, 174402 (2012)

24.262 J.A. Sidles, J.L. Garbini, K.J. Bruland, D. Rugar, O. Zuger, S. Hoen, C.S. Yannoni: Magnetic resonance force microscopy, Rev. Modern Phys. **67**, 249–265 (1995)

24.263 J.A. Sidles, J.L. Garbini, G.P. Drobny: The theory of oscillator-coupled magnetic resonance with potential applications to molecular imaging, Rev. Sci. Instrum. **63**, 3881–3899 (1992)

24.264 D. Rugar, R. Budakian, H.J. Mamin, B.W. Chui: Single spin detection by magnetic resonance force microscopy, Nature **430**, 329–332 (2004)

24.265 H.J. Mamin, D. Rugar: Sub-attonewton force detection at millikelvin temperatures, Appl. Phys. Lett. **79**, 3358–3360 (2001)

24.266 D. Rugar, O. Zuger, S. Hoen, C.S. Yannoni, H.M. Vieth, R.D. Kendrick: Force detection of nuclear magnetic resonance, Science **264**, 1560–1563 (1994)

24.267 Z. Zhang, P.C. Hammel, P.E. Wigen: Observation of ferromagnetic resonance in a microscopic sample using magnetic resonance force microscopy, Appl. Phys. Lett. **68**, 2005–2007 (1996)

24.268 C.L. Degen, M. Poggio, H.J. Mamin, C.T. Rettner, D. Rugar: Nanoscale magnetic resonance imaging, Proc. Natl. Acad. Sci. U.S.A. **106**, 1313–1317 (2009)

24.269 R.P. Feynman: There is plenty of room at the bottom, Eng. Sci. **23**, 22–25 (1960)

25. Biomedical Sensing with the Atomic Force Microscope

Constanze Lamprecht, Jürgen Strasser, Melanie Koehler, Sandra Posch, Yoojin Oh, Rong Zhu, Lilia A. Chtcheglova, Andreas Ebner, Peter Hinterdorfer

In this chapter we highlight the use and advantages of the atomic force microscope (AFM) in life science. Our aim is to present the wealth of experimental possibilities provided by this powerful toolbox with special regard to biomedical sensing applications. Originally invented in the 1980s to visualize solid surfaces on the nanometer scale, today AFM imaging is routinely used to nondestructively map the surface-ultrastructure of soft biological samples under physiological conditions with unprecedented lateral resolution. Owing to its force detection sensitivity that ranges from nano-Newtons down to a few pico-Newtons, the AFM has become an established technique for exploring kinetic and structural details of inter- and intramolecular interactions and biomolecular recognition processes. The combination of such single-molecule force measurements with topographical imaging has led to the development of recognition imaging, which allows for identification and mapping of specific components in complex biological samples with high spatial accuracy. In the following, the basic principles of biologically relevant AFM imaging modes, as well as the methods of single-molecule force spectroscopy (SMFS) and simultaneous topography and recognition imaging (TREC) will be introduced and discussed. Selected experiments will be presented in more detail to illustrate the combined application of these techniques in the elucidation of questions in molecular biology, pharmaceutical science and the medical field.

25.1 **Topographical Imaging of Biological Samples** 810
25.1.1 Basic Imaging Modes for Biological Application 811
25.1.2 Sample Preparation and Imaging Conditions 812
25.1.3 Imaging Artifacts 812
25.2 **Single-Molecule Force Spectroscopy (SMFS)** 814
25.2.1 Basic Principle of SMFS 814
25.2.2 Theory of Single-Molecule Binding and Force Spectroscopy 816
25.2.3 Data Evaluation and Determination of Interaction Rate Constants 818
25.2.4 AFM Tip Chemistry 819
25.3 **Simultaneous Topography and Recognition Imaging (TREC)** 821
25.3.1 Principle of TREC 821
25.3.2 Parameter Optimization in TREC 821
25.3.3 Application of TREC 822
25.4 **AFM Biomedical Sensing – Examples** 827
25.4.1 Isolated Membranes 828
25.4.2 Mammalian Cells 830
25.4.3 Bacteria 832
25.5 **Perspectives and Concluding Remarks** 836
References 837

Invented in 1986 as a further development of the scanning tunneling microscope [25.1], the atomic force microscope (AFM) has successfully complemented scanning and transmission electron microscopy as an instrument to visualize cells, biological membranes and molecules from the micrometer scale down to the subnanometer range. Surface textures and shapes of biological specimens are visualized as three-dimensional maps with an exceptionally good vertical resolution by scanning the surface with a sharp tip attached to a sensitive cantilever. Where electron microscopy operates under vacuum and requires extensive sample preparation, including steps such as freeze drying, staining or metal coating, AFM measurements can be carried out in aqueous and physiological environments without laborious sample preparation, staining or labeling. Structures and physiological functions of biological samples like membrane proteins, native membranes [25.2–4] and live cells [25.5–7] are preserved over an extended period of time making it possible to

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_25

follow structural changes and other dynamic processes at the nanoscale [25.8–11]. Next to topographical imaging, the tip of the AFM cantilever can be used to probe physical properties of biomolecules, cells and tissues [25.12–15] and to measure inter- and intramolecular forces [25.12, 16, 17] on the single-molecule level with outstanding signal-to-noise ratio and pico-Newton resolution [25.18–20]. The use of biomolecule functionalized AFM tips for observation and quantification of single ligand-receptor interactions, their affinity, rate constants and conformation in comparison with results from ensemble-average techniques is also of particular interest in life sciences and biomedicine [25.20, 21].

In recognition of the rapidly growing literature on AFM life science applications [25.22–25] we will highlight the use and advantages of AFM for biomedical sensing ranging from the elucidation of basic biological processes to pharmaceutical research and diagnostics [25.26]. After a brief introduction to the most relevant modes for topographical imaging of biological samples, a special emphasis will be placed on the principles and theory of single-molecule force spectroscopy (SMFS), biochemical functionalization of the AFM tip, and simultaneous topographical and recognition imaging (TREC) [25.27, 28] as the combination of topographical imaging with biosensing AFM tips. Representative studies on isolated membrane preparations, mammalian cells and bacteria will be presented to emphasize the biomedical sensing capabilities of the AFM toolbox.

25.1 Topographical Imaging of Biological Samples

AFM imaging offers the possibility to nondestructively map the topography of soft biological samples under physiological conditions and yields three-dimensional images of the surface ultrastructure with a high lateral resolution of a few nm. In general, a cantilever carrying a sharp tip is scanned laterally over a specimen. Due to the interaction with features on the surface, the cantilever is bent. In most commercial setups a laser beam is focused on the back of the cantilever and reflected into a four-segment photodiode to monitor the bending and separate the movements of the laser beam in the horizontal and vertical direction. The up and down deflection leads to a voltage difference, which can be directly attributed to changes in the z direction. Lateral movement of the cantilever leads to a shift of the laser beam between the left and the right section of the photodiode. Tip scanning can be performed in numerous different modes; the most basic and biologically relevant ones will be discussed in this section. In addition the issue of imaging conditions and artifacts,

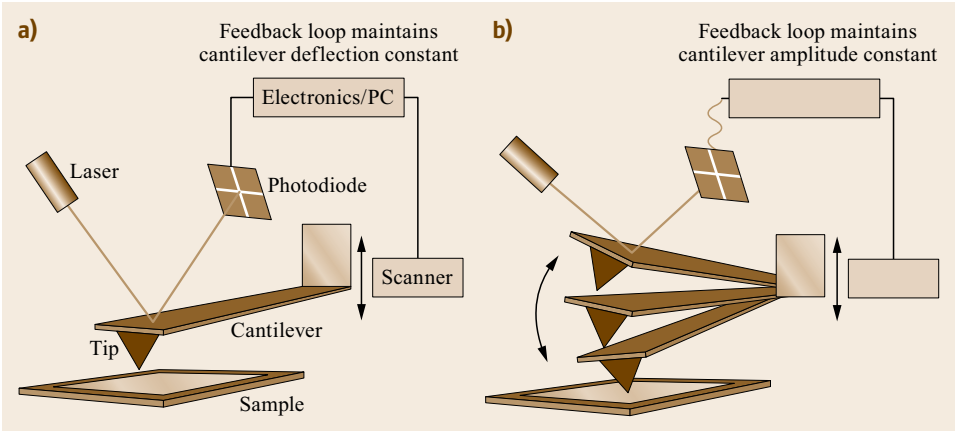


Fig. 25.1a,b Schematic illustration of the working principle of AFM. A sample is probed by an ultra-sharp tip mounted on a cantilever that scans laterally over the sample surface. Deflections of the cantilever are reported by a laser beam that is reflected from the back of the cantilever into a photodiode. The deflection signal is fed into a feedback loop that controls the movement of the piezoelectric scanner by which the tip is lifted from or lowered to the surface in order to keep the force of the tip-sample interaction constant. A topographic image of the sample is then reconstructed from the movement of the scanner in z-direction. **(a)** In contact mode the deflection is kept at a constant value. **(b)** In acoustic and magnetic AC mode the amplitude of the cantilever oscillation is maintained constant and the tip just briefly touches the sample surface

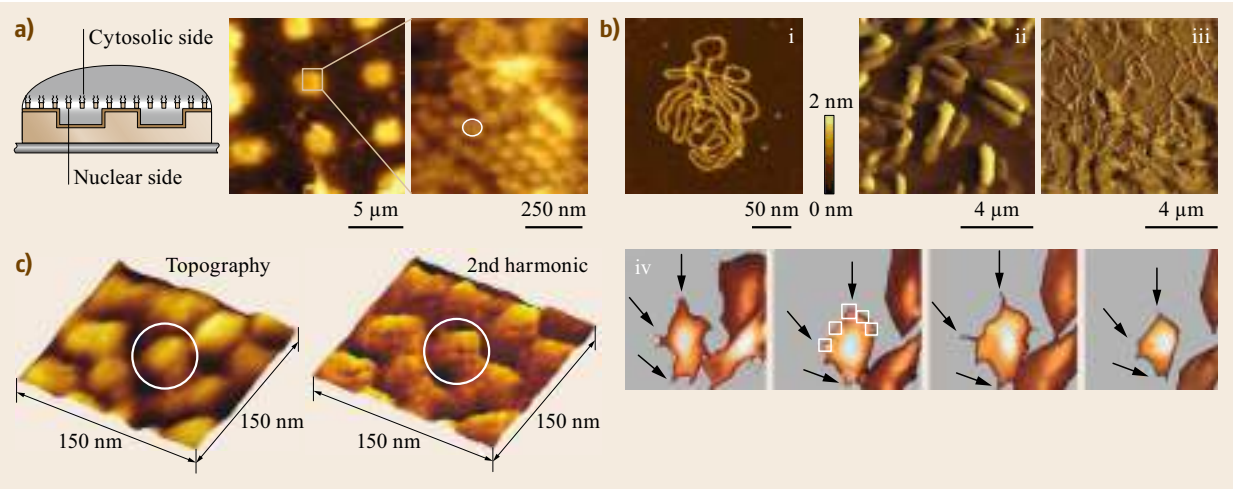


Fig. 25.2a–c Basic topographical imaging modes for biological application. **(a)** Contact mode imaging of nuclear envelope (NE) patches on microstructured PDMS (*left*, *yellow*). (Reprinted with permission from [25.29]). The thickness of the PDMS grid is about ≈ 1 mm with plateaus and wells of 3 µm in width and 0.7 µm in depth, which form microchambers into which the nuclear baskets of the nuclear pore can extend. An overview scan image of the membrane preparation performed in contact mode shows the elevated plateaus and lower regions where the membrane spans the microwells. A zoom-in on a plateau region reveals the donut-shaped nuclear pore complexes (*white circle*). **(b)** Tapping and MAC mode imaging examples of (i) a relaxed plasmid DNA cluster on 3-aminopropyl-modified mica (after [25.30]), (ii, iii) amplitude images of wild-type *E. coli* bacteria on a gelatin-coated mica surface (ii) and flagella-like nanowire structures of $CsgA^-$ mutant bacteria after removal of the cells with an AFM tip (Reprinted from [25.31] with permission from Elsevier) and (iv) visualization of shape changes of living blood platelets during activation where a series of filopodia extrusion and blebbing steps can be observed before the cells shrink and detach (Reprinted from [25.32] with permission from Elsevier). (Image size: 15 µm, color range: 0–800 nm). **(c)** Higher harmonics imaging of a dense layer of human rhinovirus serotype 2 (HRV2) on mica in buffer. The simultaneously recorded second harmonic amplitude image reveals substructures of the viral capsides (*circle*) that remain hidden in the topographical image. (Scan size: $350 \times 350\,nm^2$; color range: 0–13 nm, 0–0.35 V)

as well as some general sample preparation procedures will be addressed.

25.1.1 Basic Imaging Modes for Biological Application

AFM contact mode is the most basic imaging mode in biological applications. The tip is in permanent contact with the sample surface (Fig. 25.1a). Either the force (constant force mode) acting on the sample or the height (constant height mode) between substrate and cantilever is held constant. The latter mode is applied rather to flat samples and hard surfaces, as high shear forces might damage soft biological samples. The topography image results from the deflection of the cantilever as the tip scans over the surface. In constant force mode, the feedback loop keeps the deflection of the cantilever, i. e., the force between the tip and the sample constant. The topography image is then created from the traveling distance of the z-piezoelectric element in the scanner. The finite reaction rate of the feedback loop decreases the fastest possible scanning velocity as opposed to constant height scanning, but due to the control of forces, less sample damage can be expected. Using very soft cantilevers with a spring constant in the range of 0.1–0.301 N/m, constant force imaging can be well suited even for delicate biological applications. An example of contact mode imaging in buffer solution of a nuclear membrane preparation spanned across a polydimethylsiloxane (PDMS) grid [25.29] is presented in Fig. 25.2a. Images at higher magnification nicely resolve the embedded individual donut-shaped nuclear pore complexes (NPCs) with an outer diameter of ≈ 120 nm and inner diameter of ≈ 80 nm. Imaging of soft tissues and usually cells requires the restriction to blunt AFM tips (opening semi-angles 18° or 35°), very low loading forces (< 1 nN) and reduced scan speed (< 1 line/s) to prevent penetration of the cell membrane by the tip.

To overcome the contact mode speed issue and to eliminate frictional forces, intermittent contact or tapping mode was developed [25.33]. A piezoelectric actuator excites the cantilever to oscillate and hence the tip just briefly taps the sample surface, which minimizes the

lateral forces on the sample (Fig. 25.1b). The feedback loop keeps the oscillating amplitude constant by adjusting the distance and thus the acting force between tip and sample. The topography image is generated from the scanner's z-motion. A related dynamic imaging mode is magnetic AC mode (MAC mode) imaging, where a magnetically coated cantilever is excited to vibrate in a magnetic field. In liquid, the amplitude of the cantilevers is, in contrast to tapping mode, well defined by the driven damped harmonic oscillator model and can be easily determined. MAC mode is often the method of choice for the almost unlimited biological applications. Figure 25.2b shows some examples for MAC mode imaging DNA strands (i [25.30]), bacteria (ii, iii [25.31]) and living blood platelets (iv [25.32]). Although dynamic imaging modes induce far less shear force compared to contact mode, the viscoelastic response and dynamic nature of soft samples such as live cells essentially reduce the resolution to approximately 50–100 nm.

A more recent imaging technique is higher harmonics imaging [25.34, 35], which achieves a higher lateral resolution compared to conventional AFM imaging. In liquid, the amplitude of the second harmonic was found to be an order of magnitude higher than in air and thus more sensitive to local variations on the sample surface. Higher harmonic contributions in the movement of an oscillating cantilever are generated by nonlinear tip-sample interactions, yielding additional information on structure and physical properties such as sample stiffness. The second harmonic amplitude image – as well as the topography image – can be recorded simultaneously during scanning in tapping mode and under physiological conditions. With this technique, second harmonic amplitude images of bacterial surface layers and a layer of human rhinovirus serotype 2 (HRV2, Fig. 25.2c) showed detailed substructures with a resolution of about 0.5 nm, which were not visible in topographical images [25.36]. Later approaches used multiharmonic AFM to map local material properties of intact and disrupted virions in vitro at high resolution [25.37] and live cells [25.38].

25.1.2 Sample Preparation and Imaging Conditions

A crucial requirement for successful and reliable biological AFM experiments is proper sample preparation. Although there are many standardized nondestructive approaches for sample immobilization, they are usually only applicable for a specific system and need to be adjusted to the situation. Substrates like muscovite mica, glass and silicon oxide turned out to be the ideal support. The atomically flat mica is suitable for immobilizing e.g., DNA, single proteins or supported lipid films. Glass cover slips are rougher than mica surfaces and are thus mainly used for imaging cells and other large structures. A good alternative for mica can be silicon oxide wafers, which are more expensive but have a much smoother surface than glass [25.39].

For the immobilization of biological samples either covalent or noncovalent (adsorption) protocols can be used. Nonspecific adsorption relies on noncovalent bonds such as hydrophobic interactions, stacking, hydrogen bonds, and electrostatic attraction. Biomolecules can be adsorbed on mica or covalently bound to a chemically modified surface [25.40, 41]. Supported lipid bilayers are often prepared by the fusion of lipid vesicles on solid surfaces or by the Langmuir–Blodgett technique [25.42]. Animal cells have the ability to adhere to solid supports. Coating the surface with adhesive proteins may enhance cell immobilization. Additional chemical fixation with e.g., glutaraldehyde can prevent cell damage or detachment. Microbial cells, such as bacteria, yeast and fungi do not have the ability to spread over a surface. Thus, they have to be bound covalently to the substrate [25.43], physically adsorbed [25.44] or mechanically immobilized in e.g., a porous membrane [25.45]. In general, it is also highly recommended to investigate the effect of pH and ionic strength of the sample buffer on the image quality [25.45]. A more detailed overview of sample preparation protocols and imaging conditions may be found elsewhere [25.39].

25.1.3 Imaging Artifacts

AFM setups are commonly very well shielded against external vibrations (acoustical, mechanical) or electronic malfunctions and AFM imaging of biological specimens is quite straightforward. Nevertheless, imaging artifacts may occur rather frequently and lead to misinterpretation of data if not recognized. To achieve a high-resolution image, all scan parameters must be controlled conscientiously. Setting feedback parameters (i. e., proportional and integral gain) too high during imaging is a common mistake (Fig. 25.3a). Forces acting between the AFM probe and the sample during imaging have to be carefully monitored. Changes of pH and increasing salt concentrations due to evaporation during long-term imaging need to be considered. Temperature drifts especially in heated sample chambers influence imaging stability and the effective force might increase. Large forces can dramatically reduce the image resolution, and apart from causing molecular damage, may lead to tip contaminations, to name a few complications.

Apart from the imaging conditions, the validity of AFM images is significantly influenced by the cantilever tip as well as by the scanner (Fig. 25.3b,c). If

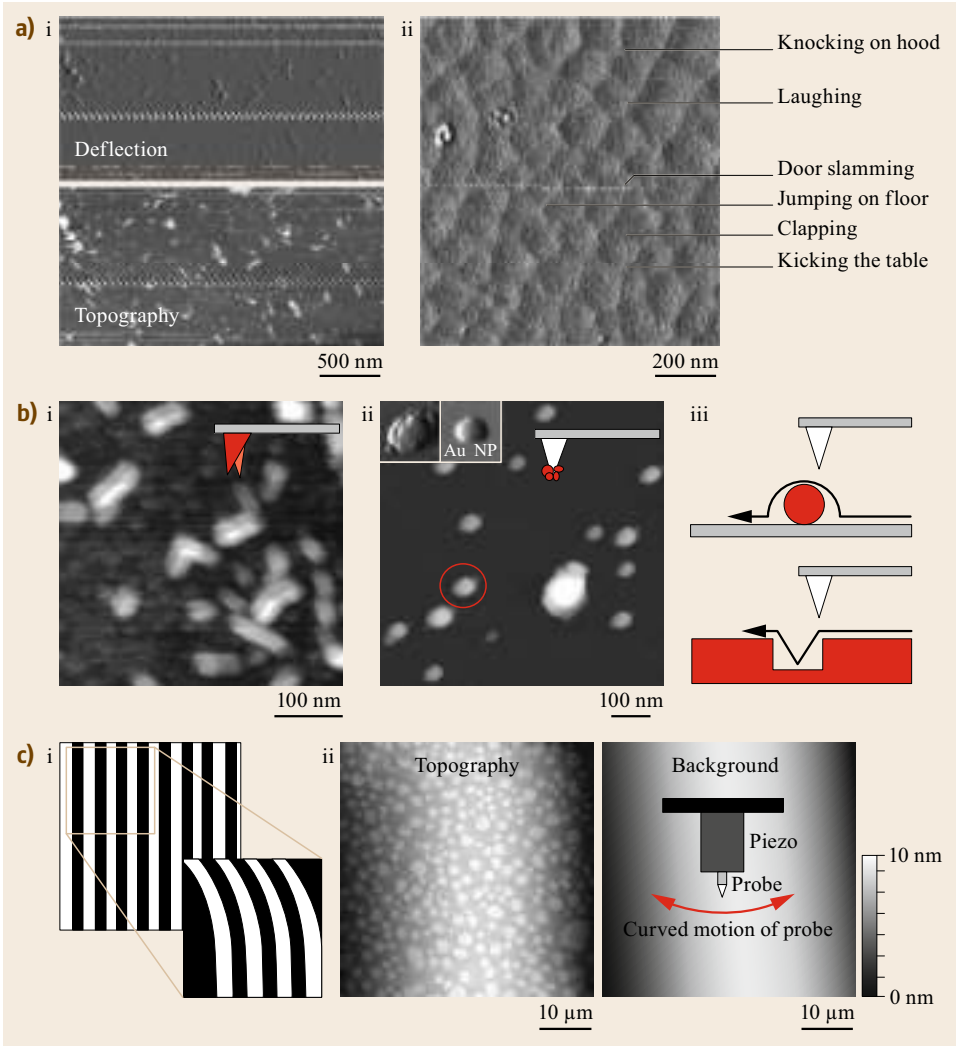


Fig. 25.3a–c Common imaging artifacts. **(a)** (i) The feedback parameters (proportional and integral gain) are set too high leading to *z*-oscillation of the piezoelectric scanner that show as *zebra-stripe*-like artifacts. Scanner oscillations appear stronger in the deflection and become visible in the topography when they are too pronounced. (ii) Perturbations in the acoustical and mechanical frequency range (e.g., vibration) are picked up by the AFM image if setups are not well enough shielded (by e.g., active vibration isolation table, acoustic enclosure). **(b)** AFM tip-associated artifacts commonly appear in the form of (i) doubled structure in the topography image, due to a double tip, or (ii) in shape of a repetitive pattern with a preferential orientation, which is caused by tip contaminations. (iii) The tip geometry radius determines the lateral resolution in AFM imaging and the tip geometry causes a broadening effect and occasionally incorrect recording of the sample features. **(c)** Artifacts caused by the piezoelectric scanner are foremost (i) a so-called scanner creep after zoom-in due to adaption of the piezo crystal to the new set voltage, resulting in stretching and compression of features at the beginning of the new image, and (ii) the nonlinearity of a tube scanner in the *z*-direction, that accounts for a nonlinear (mostly parabolic) background, which is commonly removed by postprocessing of images (flattening). The topographical image shows an overview scan of a phase-separated supported lipid bilayer (DOPC:SM:Chol 2 : 2 : 1). The liquid ordered phase (*round patches*) is about 1 nm higher than the surrounding liquid disordered phase

the tip curvature radius, which is typically between 5 and 10 nm, is comparable or bigger than the structures on the sample surface, the lateral size of the structures appears overestimated. In this case, the full width at half maximum (FWHM) value of the sample features can be determined by analyzing the cross-section of AFM to-

pography images. The following relation connects the FWHM value with the actual diameter d of the investigated structures, by implying the tip radius

$$R : \mathrm{FWHM} = 2\sqrt{d((R+d)/4)}\,.$$

In general, to achieve high-resolution AFM images of biological samples, it is important to use a sharp tip (Fig. 25.3b, iii) and a soft cantilever in order to nondestructively map the sample features on the surface [25.46].

Also the piezoelectric scanner may be directly a source of errors. Especially when imaging large areas, i. e., when the scanner has to expand far, deviations from the linear expansion behavior can occur (Fig. 25.3c). To avoid biased images of the sample topography, the scanner should be perpendicular to the surface. Hysteresis in the response of the piezo element (*piezo creep*) may cause edge overshoots and together with thermal scanner drift may cause image distortion. These effects are often observed when zooming in.

25.2 Single-Molecule Force Spectroscopy (SMFS)

Measurements of interaction forces traditionally rely on ensemble techniques such as shear flow detachment [25.47] and the surface force apparatus [25.48]. Single-molecule techniques, on the other hand, can reveal molecular processes, pathways and rare events that would be masked in ensemble measurements by analyzing individual events in large heterogeneous populations. Though they typically require more effort and time to reliably generate high-quality data, direct measurements can provide invaluable insight into complex systems. In AFM-based force spectroscopy the cantilever is used as a molecular sensor to quantify interaction forces between the tip of the cantilever and the substrate. The deflection of the AFM cantilever Δx relates to the force F acting on the cantilever through Hooke's law $F = k\Delta x$, with the spring constant k of the cantilever. The force resolution of the AFM, $\Delta F = (k_\mathrm{B} T k)^{1/2}$, is limited by the thermal noise of the cantilever, which, in turn, is determined by its spring constant. With commercially available (bio-)cantilevers, interaction forces between single molecules in the range of only a few pico-Newton (pN) [25.21] and adhesion forces between a single cell and a substrate in the order of a few nano-Newton (nN) [25.49] are routinely resolved. For single-molecule force spectroscopy (SMFS) the small size of AFM tips that ranges between 2 and 50 nm provides unprecedentedly high lateral resolution.

Over the past 20 years the technique has been refined constantly and become a versatile tool for characterizing molecular interactions [25.12, 16, 17, 50]. During this time various single-molecular pairs, such as avidin and biotin or streptavidin and biotin, antibodies and their antigens, proteins and specialized aptamers, or DNA and DNA-binding proteins, have been studied with great success [25.51, 52]. Besides the detection of intermolecular forces, AFM also shows great potential in measuring intramolecular forces acting within molecules, e.g., mechanoproteins, membrane proteins, DNA or polysaccharides [25.50, 52–55]. In these experiments, the molecule is clamped between the tip and the surface and its viscoelastic properties as well as unfolding and refolding pathways are studied by force-distance cycles [25.53].

SMFS is not only limited to the reporting of interaction forces either, and can readily be used to obtain interaction rate constants of molecular association and dissociation as well as length scales and relative heights of energy barriers traversing the free energy surface. To this end, the rate at which a pulling force is applied to the bond is varied, revealing a dependence of this *loading rate* and kinetic and thermodynamic parameters. This technique of dynamic force spectroscopy is explained in greater detail in this section.

25.2.1 Basic Principle of SMFS

Basic single-molecule force spectroscopy (SMFS) requires immobilization of one of the binding partners to the apex of the AFM tip and the other to a hard surface such as mica or glass. The functionalized tip is then repeatedly brought into contact with and retracted from the substrate, while recording the cantilever deflection, i. e., the force acting on the cantilever as illustrated in Fig. 25.4a. During most of the approach phase, when the distance of the tip to the surface is sufficiently large (1 to 3), the cantilever deflection remains zero. Upon contact (4) the cantilever bends upwards due to a repulsive force that increases linearly (4 to 5) as the tip is pushed further into the surface. After a predefined contact force (5) has been reached, the cantilever movement is reversed and the deflection relaxes (5 to 6). If binding has occurred, the cantilever will bend downwards as it is further retracted from the surface (6 to 7). In the case where the binding partners were tethered through flexible crosslinkers [25.56, 57], the shape of the attractive force-distance profile is nonlinear. It then depends on the elastic properties of the cantilever

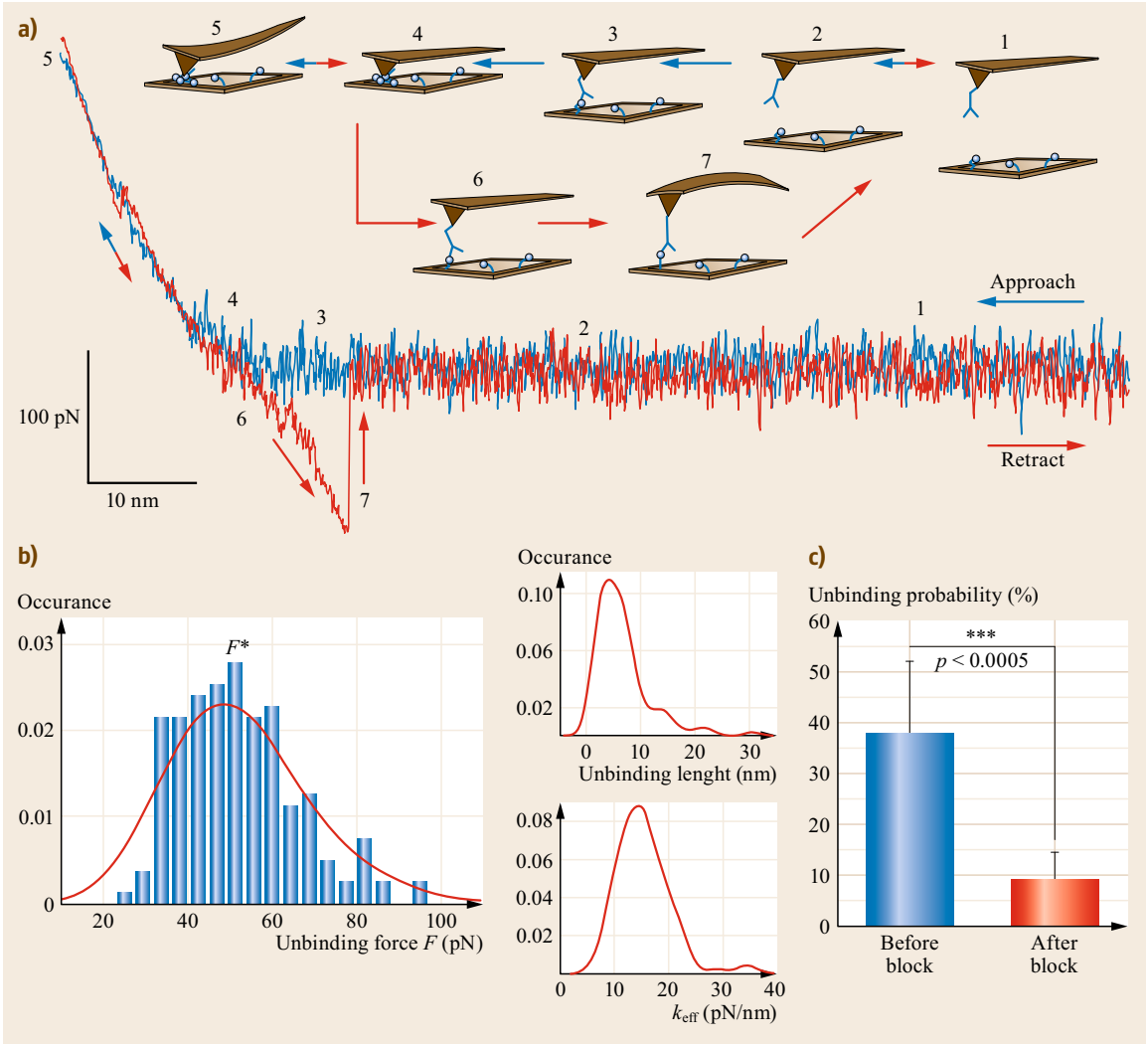


Fig. 25.4a–c Single-molecule force spectroscopy (SMFS). **(a)** A force-distance cycle (FDC) (e.g., sweep rate of 200 nm/s) with a ligand functionalized AFM tip and a dense layer of receptor molecules immobilized on mica. Binding occurs towards the end of the approach phase (1–5) when the tip comes into contact (4) and is pressed into the surface. Under further approach the tip bends upwards exerting a linearly increasing force on the sample (5). During tip retraction the bending relaxes again. When a ligand-receptor bond has formed, the cantilever progressively bends downwards (6), visible as a characteristic parabolic slope in the retract curve reflecting the extension of the distensible crosslinker mediating the ligand-receptor connection. The force increases until rupture (unbinding) of the ligand-receptor bond occurs (7). After rupture the retract curve closely follows the approach curve. **(b)** Statistical analysis of a great number of unbinding events at a given sweep rate and indentation force yields distributions of unbinding force F, unbinding length l and effective spring constant k_{eff} that can be displayed as a histogram (*blue*) or probability density functions (PDF) (*red*). In a PDF the data are weighed by their uncertainty and presented in a continuous curve that offers a measure of reliability not provided by simple binning. This eases determination of the most probable values for each parameter e.g., F^* as the most probable unbinding force. **(c)** The specificity of the ligand-receptor interaction needs to be verified by a block experiment where typically one of the binding partners is saturated by free agonist molecules. This will result in a significant decrease of the number of detected unbinding events i. e., the binding probability. The statistical significance can be rated e.g., by an unpaired two-sample t-test

(commonly modeled as a harmonic spring), and the crosslinker, which exhibits parabolic-like characteristics, reflecting an increase of the spring constant during extension and the bond itself. When the ramping force reaches a critical value the binding partners dissociate and the cantilever jumps back into the resting position (7). The force value at dissociation is termed the unbinding force and is quantified by the vertical distance from the critical bond force to the baseline (7 to 2). Similarly, the unbinding length is given by the horizontal distance from the contact point to the extension at rupture (6 to 7, horizontal component). In the absence of interaction between tip-bound ligand and surface receptors the retract curve will closely follow the approach curve. A typical SMFS experiment at a given pulling velocity comprises 1000 so-called force-distance cycles (FDCs) resulting in a distribution of unbinding forces, which can be either plotted as a histogram or as a continuous probability density function (PDF) (Fig. 25.4b). The maximum of the distribution is termed the most probable unbinding force F^*. The specificity of the observed (un)binding event is usually demonstrated by blocking the interaction either through addition of free ligands to mask receptor sites on the surface, or by saturating the binding partner on the AFM tip, drastically reducing the binding probability (i. e., the number of unbinding events observed per 1000 cycles) (Fig. 25.4c).

25.2.2 Theory of Single-Molecule Binding and Force Spectroscopy

Molecular recognition is mediated by a multitude of noncovalent interactions, the energy of which is only slightly higher than thermal energy. Due to the power law dependence of these interactions on distance, the attractive forces between noncovalently interacting molecules are extremely short ranged. A close geometrical and chemical fit within the binding interface is therefore a prerequisite for efficient association. The weak bonds that govern molecular cohesion are believed to be formed in a spatially and temporarily correlated fashion. Protein binding often involves structural rearrangements, either localized or global, that commonly bear functional significance by modulating the activity of the interaction partners. Signaling pathways, enzyme activity, and the activation and inactivation of genes all depend on conformational changes induced by ligand binding.

The strength of binding is usually given by the binding energy ΔG, which amounts to the free energy difference between the bound and the free state, and which can readily be determined by ensemble measurements. ΔG determines the ratio of bound complexes $[RL]$ to the product of free reactants $[R][L]$ at equilibrium and is related to the equilibrium dissociation constant K_D through $\Delta G = -RT \ln(K_D)$, where R is the gas constant and T the temperature in Kelvin. K_D itself is related to the empirical association (k_{on}) and dissociation (k_{off}) rate constants through $K_D = k_{off}/k_{on}$. In order to get an estimate for the interaction forces F, from binding energies ΔG, the depth of the binding pocket may be used as a characteristic length scale l. Using typical values of $\Delta G = 20 k_B T$ and $l = 0.5\,\text{nm}$, an order of magnitude estimate of $F\ (= \Delta G/l) \approx 170\,\text{pN}$ is obtained for the binding strength of a single molecular pair. Classical mechanics describes bond strength as the gradient in energy along the direction of separation. Unbinding therefore occurs when the applied force exceeds the steepest gradient in energy. This purely mechanical description of molecular bonds, however, does not provide insights into the microscopic determinants of bond formation and rupture.

Bonds will resist detachment when a pulling force is applied. Notably, the unbinding force may approach and even exceed the adiabatic limit given by the steepest energy gradient of the interaction potential if rupture occurs in less time than needed for diffusive relaxation (10^{-10}–10^{-9} s for biomolecules in viscous aqueous medium) and friction effects become dominant [25.58]. Therefore, unbinding forces do not resemble unitary values and the dynamics of the experiment critically affects the measured bond strengths. At the time scale of AFM experiments (milliseconds to seconds), thermal impulses govern the unbinding process. In the thermal activation model, the lifetime of a molecular complex in solution is described by a Boltzmann ansatz, $\tau(0) = \tau_{osc} \exp(E_b/k_B T)$ [25.59], where τ_{osc} is the inverse of the natural oscillation frequency and E_b is the height of the energy barrier for dissociation. This gives a simple Arrhenius dependency of dissociation rate on barrier height.

Force acting on a complex deforms the interaction free energy landscape and lowers barriers for dissociation (Fig. 25.5a). As a result, bond lifetime is shortened. The lifetime $\tau(F)$ of a bond loaded with a constant force F is given by $\tau(F) = \tau_{osc} \exp[(E_b - x_\beta F)/k_B T]$ [25.59], where x_β marks the thermally averaged projection of the energy barrier along the direction of the force.

The Bell–Evans Model

A detailed analysis of the relation between bond strength and lifetime was performed by *Evans* and *Ritchie* [25.60], using Kramers' theory for overdamped kinetics. For a sharp barrier, the lifetime $\tau(F)$ of a bond subjected to a constant force F relates to its characteristic lifetime, $\tau(0)$, according to $\tau(F) = \tau(0) \exp(-x_\beta F/k_B T)$ [25.12]. However, in

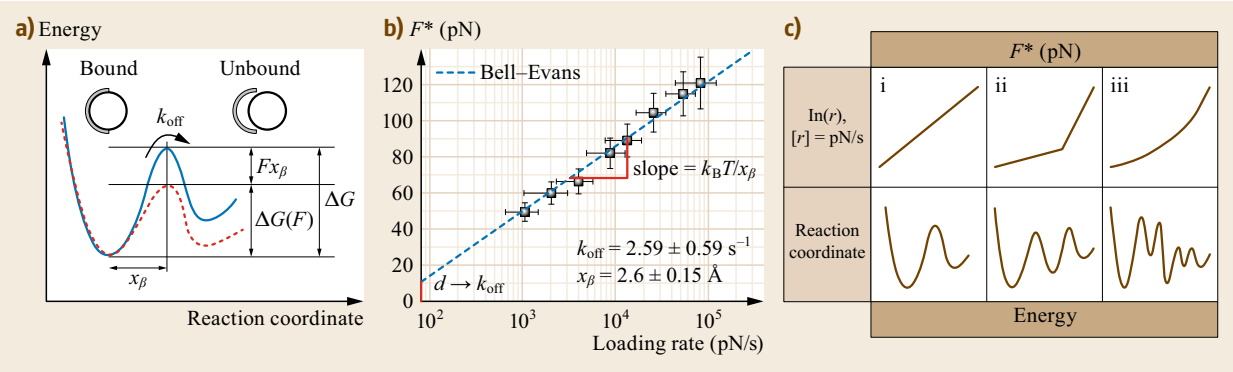


Fig. 25.5a–c Theory and interpretation of single-molecule binding and force spectroscopy. **(a)** In a simple model of molecular dissociation over a single sharp energy barrier, a constant force F lowers the barrier's height proportionally. This gives rise to a characteristic length scale x_β that is interpreted as the distance of the energy barrier from the energy minimum along the projection of the force. The barrier is crossed with the characteristic unbinding rate k_{off} (after [25.51]). **(b)** In the Bell–Evans model the most probable unbinding force F^* scales with the natural logarithm of the loading rate. The intersect with the force axis yields the characteristic unbinding rate k_{off} and the slope reveals the characteristic length scale x_β. **(c)** In the classical Bell–Evans interpretation linear segments in the logarithmic plot of the loading rate dependence reflect multiple independent barriers along the energy profile of a biological bond (ii). For nonlinear curved spectra very complex energy landscapes may result from this interpretation (after [25.51])

most pulling experiments the applied force is not constant but varies with time, and the measured bond strength depends on the loading rate [25.58, 60, 61]. In accordance with this, experimentally measured unbinding forces do not assume unitary values but rather vary with both pulling velocity [25.62, 63] and cantilever spring constant [25.16]. Nevertheless, contributions arising from thermal activation manifest mostly near the point of detachment. Therefore, the change of force with time, i. e. the loading rate r $(= \mathrm{d}F/\mathrm{d}t)$, can be derived from the product of the pulling velocity and the effective spring constant at the end of the force curve, just before unbinding occurs.

Dynamic force spectroscopy (DFS) exploits this dependence of bond strength on the loading rate to obtain detailed insights into intra- and intermolecular interactions. The dependence of the rupture force on the loading rate (force spectrum), in the thermally activated regime was first derived by *Evans* and *Ritchie* [25.60] and described further by *Strunz* et al. [25.64]. Rupture itself is a stochastic process and the likelihood of bond survival is expressed in the master equation as a time-dependent probability $N(t)$ to be in the bound state under a steady ramp of force, namely $\mathrm{d}N(t)/\mathrm{d}t = -k_{off}(rt)N(t)$ [25.64]. This results in a distribution of unbinding forces $P(F)$ parameterized by the loading rate [25.60, 64, 65]. The most probable force for unbinding F^*, given by the maximum of the distribution (Fig. 25.4b), relates to the loading rate through $F^* = F_\beta \ln(rk_{off}^{-1}/F_\beta)$, where the force scale F_β is set by the ratio of thermal energy to x_β (Fig. 25.5a) [25.60, 64]. Thus, the unbinding force scales linearly with the logarithm of the loading rate. This phenomenological approach to relating data from pulling experiments to the innate kinetic and thermodynamic parameters of the probed interaction is commonly referred to as the Bell–Evans model.

For a single barrier, the Bell–Evans model would give rise to a simple, linear force spectrum F^* versus $\ln(r)$ (Fig. 25.5c, i). In cases where the escape path traverses several barriers, the curve will follow a sequence of linear regimes, each marking a particular barrier [25.60, 65, 66] (Fig. 25.5c, ii). Transition from one regime to the other is associated with an abrupt change of slope determined by the characteristic barrier length scale and signifying that a crossover between barriers has occurred. The slopes of the force versus loading rate curves contain information about the length scale x_β of prominent energy barriers along the force-driven dissociation pathway, which may be related to the depth of the binding pocket of the interaction (Fig. 25.5a,b) [25.62].

Limitations of the Bell–Evans Model and New Approaches to DFS Data Analysis

The predicted linear dependence of the unbinding force on the logarithm of the loading rate r holds well when the barriers are stationary with force, as found by a large number of unbinding and unfolding experiments. However, if the position of the transition state is expected to vary along the reaction coordinate with the force,

as for example when the curvature at the top of the barrier is small, this simple dependence gives way to more complex forms. *Schlierf* and *Rief* [25.67] used Kramers' diffusion model to calculate the probability force distributions when the barrier cannot be assumed to be stationary. A failure to fit force distributions at high loading rates using a Bell–Evans model was also reported in by *Neuert* et al. [25.68] for the interaction of digoxigenin and antidigoxigenin. Poor matches were observed in the crossover region between the two linear regimes of the force spectrum as well.

Additionally, fitting each regime in the force spectrum with a separate function according to the Bell–Evans model can lead to very complex underlying energy landscapes even for simple experimental systems. For curved force spectra (Fig. 25.5c, iii) this interpretation may result in an arbitrary number of barriers, each characterized by at least two fit parameters, as well as values for x_β several orders of magnitude below 1 Å [25.51, 69]. Lastly, the phenomenological model described above holds for linear energy surfaces with $\Delta G \to \infty$, a simplification that has been shown to be applicable to a large number of biomolecular interactions. For some systems, however, different energy barrier and energy well geometries might correspond better to experimental data [25.70].

Alternative models that predict inherently nonlinear force spectra have emerged in recent years, extending the framework laid down by Evans and Ritchie. The newer models described below include additional fit parameters, most notably the height of the free energy barrier, which does away with the necessity of estimating ΔG using the Arrhenius equation. *Dudko* et al. [25.70] modified the Bell–Evans model by accounting for the effects of the applied force on the position of the transition state. In their treatment they find that a dependence of the mean rupture force on the loading rate r in the form $(\ln r)^\nu$, with the exponential constant ν allowing for either cusp-like ($\nu = 1/2$) or linear-cubic ($\nu = 2/3$) free-energy surfaces, corresponds more closely to experimental force spectra than a linear approach ($\nu = 1$, corresponding to the Bell–Evans model). They also find the distribution of unbinding forces to be asymmetrical, as indeed has been observed many times. In a subsequent publication *Dudko* et al. [25.71] introduced a method based on transforming rupture force distributions from DFS into force-dependent lifetimes akin to those obtained from constant-force experiments, and show how the molecular linker used in the experiment affects the loading rate.

The model put forward by *Friddle* et al. supports the notion of reversible rebinding at low loading rates when using soft springs [25.69], a process that is not taken into account by the other approaches to DFS presented here. The Friddle model introduces two distinct regimes during molecular unbinding: an equilibrium regime at low loading rates, which supports cycles of reversible unbinding and rebinding, and a kinetic regime at higher loading rates where unbinding is irreversible. In addition to k_{off} and x_β, an additional fit parameter is introduced, the equilibrium force F_{eq}. This is the force where dissociation and association rates are in balance. F_{eq} is the lowest force at which a bond breaks for a given pulling apparatus geometry and pulling speed. This also means that at forces below F_{eq}, no complete dissociation will occur, which marks an important difference between a free ligand-receptor pair, where spontaneous unbinding is possible, and a system where the interaction partners are tethered to force transducers. An interesting application of this model was presented by *Alsteens* et al. [25.72], who used it to analyze DFS data of the protease-activated receptor-1 (PAR1) in native membranes and its ligands. Whereas the majority of DFS models operate on most-probable rupture forces obtained from pooling many hundreds of measurements, Alsteens et al. present a method to apply the work of *Friddle* et al. [25.69] to data clouds of individual unbinding events. In addition to these new models for the analysis of DFS data, several modifications of the Bell–Evans model itself were proposed over the years [25.51, 73, 74]. These reflect new findings about parameters influencing experiments, such as the stiffness of the force transducer [25.74] or the range of loading rates in which this phenomenological approach is applicable [25.73].

25.2.3 Data Evaluation and Determination of Interaction Rate Constants

A comprehensive guide to modern analysis of DFS is given by *Bizzarri* and *Cannistraro* [25.51] and *Noy* and *Friddle* [25.75], whereas *Hane* et al. [25.76] have compared the three dominant models in the field in great detail. The choice of model relies first and foremost on the obtained force spectrum: nonlinear alternatives to the Bell–Evans model should only be used if the forces measured experimentally actually do not scale linearly with $\ln(r)$, since fitting a linear dependence with a nonlinear function, while possible, may result in fit parameters with little to no meaning.

While the dissociation rate constant k_{off} is a fit parameter in DFS models (Fig. 25.5), quantification of the association rate constant k_{on} has to be determined via the interaction time that is required to achieve half of the maximum possible binding probability ($t_{1/2}$). This can be obtained from experiments where the encounter time between receptor and ligand is varied over

a broad range [25.62]. Given that the concentration of ligand molecules on the tip available for interaction with the surface-bound receptors c_{eff} is known, the association rate constant can be derived from $k_{\mathrm{on}} = t_{1/2}^{-1} c_{\mathrm{eff}}^{-1}$. Determination of the effective ligand concentration requires knowledge of the effective volume V_{eff} explored by the tip-tethered ligand which, in turn, depends on the tether length. Therefore, only order-of-magnitude estimates of k_{on} can be gained from such measurements [25.62].

Note that the application of force acts to select the dissociation path. Since the kinetics of reactions are pathway dependent, such a selection implies that kinetic parameters extracted from force-probe experiments may differ from those obtained from assays conducted in the absence of external force. At extremely fast complexation-decomplexation kinetics the forces may be independent of the loading rate, indicating that the experiments were effectively carried out under thermodynamic equilibrium [25.77]. These concerns notwithstanding, SMFS has been demonstrated to be a versatile tool enabling the estimation of association and dissociation rate constants, as well as the measurement of structural parameters of the binding pocket [25.51, 55].

25.2.4 AFM Tip Chemistry

For SMFS experiments, the usually inert AFM tip has to be converted into a (mono)molecular sensor by attaching one or a few ligand molecules to the apex of the tip. Based on the tip surface material, i. e., silicon nitride or gold, different approaches are possible. While the use of gold-coated tips allows functionalization based on self-assembled monolayers (SAMs) of thiolated linker molecules, the common silicon or silicon nitride tips have to be chemically modified in the first step (e.g., by aminofunctionalization protocols). More simple strategies like adsorption or chemisorptions may provide for efficient coupling under suitable circumstances. A more detailed description can be found in [25.43]. Physisorption, meaning the nonspecific adsorption of molecules, which relies on noncovalent bonds including hydrophobic interactions, stacking, hydrogen bonds, and electrostatic attraction, offers an easy but effective coupling strategy. However, AFM tip functionalization requires very firm coupling. Though this might be realized by physisorption (e.g., with the help of crosslinking [25.63, 78]), only covalent attachment protocols will yield reliable and reproducible coupling results that render the tip nonadhesive and introduce a distensible linker between ligand and AFM tip. Figure 25.6 illustrates common strategies for functionalization of commercial silicon-silicon nitride and cold-coated AFM tips.

Functionalization of Silicon or Silicon Nitride Tips

The most common approach to make inert silicon and silicon nitride tips chemically addressable is aminofunctionalization (Fig. 25.6a). The gas-phase protocol using aminopropyl-triethoxy-silane (APTES) as well as the liquid-phase method employing ethanolamine-hydrochloride in dimethyl sulfoxide (DMSO) are frequently used. Both yield sufficient reactive sites ($\approx$ 1000–2000 amino groups/μm^2) while avoiding the formation of sponge-like multilayers, which are not suitable for force spectroscopy experiments [25.79]. A good understanding of the chemical process helps in performing successful aminofunctionalization: under ambient conditions the surface of silicon and silicon nitride tips has a silicon-oxide layer with high numbers of silanol (-Si-OH) groups. APTES, as the most commonly used aminosilanization reagent, reacts with these silanol groups. The reaction itself has to be performed in an argon-flooded desiccator to control the amount of water. The AFM cantilever chips are thoroughly cleaned with chloroform and might be treated by a mild plasma before placing them into the desiccator. In a 6 l argon-filled desiccator typically 30 µl APTES together with 10 µl triethylamine as catalyst are used and allowed to react for 120 min followed by careful purging with argon gas. A two-day *curing* at room temperature ensures sufficient and strong binding of the generated APTES layer on the cantilever. The aminofunctionalization protocol based on ethanol-hydrochloride is a simpler alternative to gas-phase APTES coating, but results in a less well-defined chemical coverage. In a nutshell, ethanolamine-hydrochloride is dissolved in DMSO, molecular sieves are added and the cleaned tips are incubated overnight followed by a careful cleaning procedure [25.80].

In the next step a heterobifunctional cross-linker is bound to the amino residues on the tip. Polyethylene glycol (PEG) consisting of 10–30 ethylene glycol units turned out to be best suited, due the properties of this polymer. PEG itself is an inert, water-soluble, and distensible polymer with the advantage of being protein repulsive. In addition, denaturation of most of the bioligands is markedly lowered compared to coupling with a short linker. Only in very special cases does this linker already contain the ligand molecule itself, e.g., biotin [25.80] or fluorescein [25.81]. These molecules can be bound as part of the linker, since their structure is stable at the typical coupling conditions, e.g., chloroform as a solvent and applying triethylamine as a catalyst.

Most other ligands – usually proteins – require aqueous coupling conditions. Proteins are often coupled to the PEG crosslinker via their lysine residues.

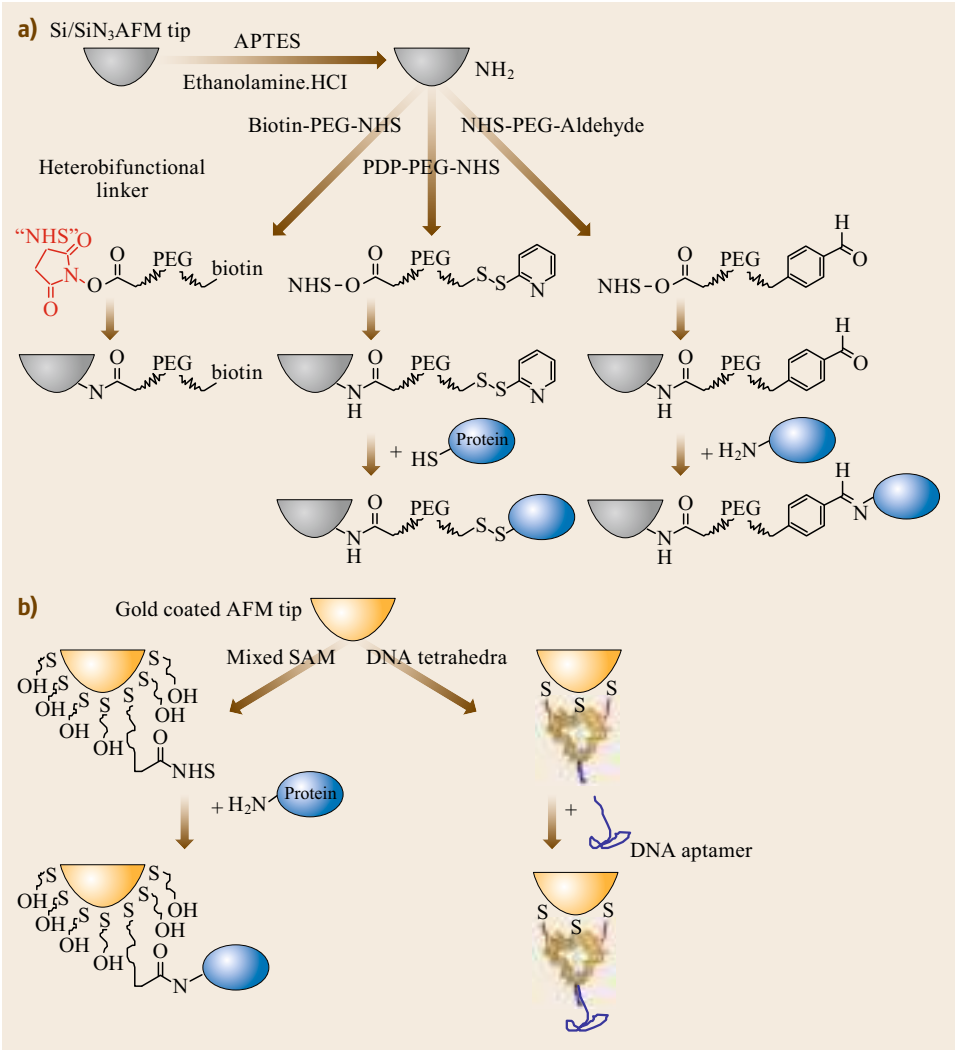


Fig. 25.6a,b Functionalization strategies for the covalent coupling of biological ligands to AFM tips. **(a)** A common approach for silicon and silicon oxide tips requires amino-functionalization of the tip surface for attachment of a heterobifunctional crosslinker that will tether the ligand. **(b)** Functionalization of gold-coated tips exploits thiol chemistry. Building blocks such as mixed self-assembled monolayers or DNA tetrahedra are used to link the ligand to the tip

N-Hydroxysuccinimide (NHS)-PEG-aldehyde [25.82] or NHS-PEG-acetal [25.83] linkers are both well suited, though the former is slightly prone to loop formation. Ligands with a free thiol end (e.g., cysteine mutants) can be tethered by NHS-PEG-PDP linkers [25.84] with the advantage that the bond can be cleaved, or by NHS-PEG-maleimide for noncleavable coupling. A very general strategy is realized by the use of the NHS-PEG-benzophenone linker [25.43]. After binding the linker via its NHS end to the aminofunctionalized tips, benzophenone can be activated by ultraviolet (UV) light (366 nm) and allows stable covalent bond formation with aliphatic carbons. Admittedly, the above-mentioned linkers do not allow for site-directed binding, as a ligand usually offers more than one coupling group, for instance when more than one lysine exists in the amino acid sequence of a protein. Synthetic proteins, however, are often tagged by hexa- or decahistidine for purification (His_6 and His_{12} respectively), which can be exploited for site-directed coupling by the use of nitrile-tri-acetic acid (NTA) functionalized linkers. Here the coupling protocol requires somewhat more effort since the synthesis of an NHS-PEG-NTA linker is not possible. A detailed description can be found in [25.85]. Table 25.1 gives an overview of the described linkers, their reactive sites and the target group in the bioligand. For in-detail AFM tip functionalization protocols the reader may refer to [25.43] and [25.86]. For completeness, it might be added that apart from these most common strategies, a number of other coupling protocols are available that may also yield stable and distensible attachment of ligands to AFM tips, which are covered for instance by [25.87].

Table 25.1 Overview of heterobifunctional PEG linkers for silicon-silicon nitride tips

Typical target molecule	Linker	Reactive site	
Biotin	NHS-PEG-biotin	–	*Ebner* et al. [25.43, 79], *Riener* et al. [25.80]
Protein, R-NH_2	NHS-PEG-aldehyde	Lysine (R-NH_2)	*Ebner* et al. [25.43, 82]
Protein, R-NH_2	NHS-PEG-acetal	Lysine (R-NH_2)	*Ebner* et al. [25.43], *Wildling* et al. [25.83]
Protein, R-SH	NHS-PEG-PDP	Cysteine, R-SH	*Ebner* et al. [25.43], *Kamruzzahan* et al. [25.84]
Protein, R-SH	NHS-PEG-maleimide	Cysteine, R-SH	*Ebner* et al. [25.43], *Lin* et al. [25.88]
Aliphatic carbons (e.g., DNA)	NHS-PEG-benzophenone	Aliphatic carbon	*Ebner* et al. [25.43]
His_6-tagged protein	Multistep procedure	His_6-tag	*Ebner* et al. [25.43], *Tang* et al. [25.85]

Functionalization of Gold-Coated Tips

Gold-coated tips are readily commercially available, but have the drawback that the gold coating, including the precoating with an adhesive layer (e.g., chromium), increases the tip radius by 10 nm or more. Still, gold is an inert noble metal that can be cleaned at harsh conditions without any troublesome effects like surface oxidation. Functionalization can be achieved with an extremely high efficiency utilizing the thiophilic properties of gold (Fig. 25.6b). To ensure sufficiently strong binding of thiolated crosslinkers, either a SAM of bifunctional alky-mercaptanes, or linkers containing a higher number of sulfur residues should be employed. In general, successful SAM formation on the AFM tip is achieved by a mixed SAM, where one component (e.g., a ω-mercaptoalkanol) is used as a nonadhesive part and the other component affords covalent coupling of the ligand of interest. In contrast, the use of thiolated building blocks such as thiolated DNA tetrahedra offers a one- or two-step functionalization procedure for the generation of reactive coatings with highly defined spacing [25.89].

25.3 Simultaneous Topography and Recognition Imaging (TREC)

In addition to measuring interaction forces, mapping the specific recognition sites on biological samples with high spatial accuracy is of great interest in many fields of life science. In contrast to the well-established but rather slow force-volume mode (FV) [25.90–96], the simultaneous topography and recognition imaging (TREC) technique has developed into a versatile tool for quick local mapping of binding sites in molecular biology and bionanotechnology [25.27, 28, 97]. Although a lot of progress has been made in terms of performance and speed [25.98–101], TREC is still considerably faster than many force-volume techniques. For comparison a complete FV image at 30 nm resolution requires typically 20 min and increasing the resolution dramatically lowers the recording speed. In contrast, TREC images with 2 nm resolution are recorded in less than 4.5 min.

25.3.1 Principle of TREC

TREC is based on a dynamic imaging mode in fluid, where the image is generated by oscillating the tip close to its resonance frequency (Sect. 25.1.1). On top of that, the topographical imaging of receptor molecules is combined with molecular recognition by tethering a ligand to the AFM tip using a distensible PEG crosslinker (Sect. 25.2.4). The tip-bound ligand is capable of binding to cognate receptors on the sample surface. With this configuration and a modified feedback loop, two independent images are recorded simultaneously – a topography image of the sample surface and a lateral map of the corresponding receptor recognition sites. The separation into topographical and recognition signals is achieved by splitting the cantilever's oscillation amplitude into downward and upward parts with respect to the cantilever's zero position using a special electronic circuit (PicoTREC, Keysight, formerly Agilent, Santa Rosa, CA, USA) (Fig. 25.7a). Provided that the cantilever's Q-factor is ≈ 1 when in contact with the sample in liquid, the maxima (or peaks) of oscillation are independently influenced by the tip-sample interaction, i. e., topographic features measurably effect only the downward amplitude, whereas the physical connection of the tip to the surface during ligand-receptor binding causes reduction of the upward amplitude of the tip oscillation. Consequently, the envelope of the lower amplitudes generates the topographic image (Fig. 25.7b) and the envelope of the upper amplitudes yields the map of the recognition sites (Fig. 25.7c). As the tip contacts the surface only intermittently, this technique provides very gentle tip-surface interactions and the high oscillation frequency provides for more than a hundred times faster mapping than conventional force-volume techniques.

25.3.2 Parameter Optimization in TREC

In 2004, the first proof of concept on biological samples was realized by visualizing the highly spe-

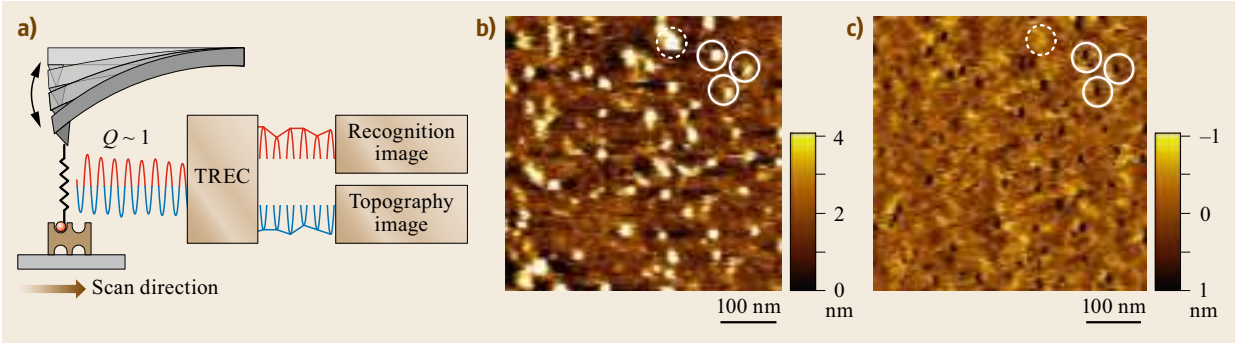


Fig. 25.7a,b Simultaneous topography and recognition imaging (TREC). **(a)** The cantilever oscillation is split into lower and upper parts allowing for simultaneous acquisition of topography and recognition image provided that the cantilever's Q-factor is ≈ 1. **(b)** Topography image of electrostatically adsorbed avidin molecules on a freshly cleaved mica imaged with a biotin-tethered tip. **(c)** Simultaneously recorded recognition image. Topography and recognition show very good correlation with avidin molecules causing corresponding recognition spots (*solid circles*). Some topographical features lack recognition (*dashed circle*) as they do not interact with the tip-bound biotin at all. (Reprinted with permission [25.28, 102])

cific lysozyme-HyHEL5 antibody binding in single-molecule interaction maps while simultaneously carrying out high-resolution topographic imaging [25.97]. Shortly afterwards, *Ebner* et al. [25.102] employed the well-characterized, high-affinity avidin-biotin interaction to determine key factors for stable, robust and reliable recognition imaging results. In 2009, *Preiner* et al. [25.103] investigated the physical properties of the tip-sample ensemble in much detail and established general handling procedures and a protocol for new TREC users to conduct meaningful TREC imaging experiments.

Unbiased values for topographical heights require operation of the feedback loop with only the lower half of the cantilever amplitude, which responds to the topographical features. Use of the full amplitude would include the influence of the recognitions signals, causing faulty height information (Fig. 25.8a). In order to reduce feedback artifacts (termed as topographical crosstalk) and unspecific adhesion in the recognition image, the driving frequency should be adjusted to a value smaller than the cantilever resonance frequency at the surface. The most important factor, however, is the magnitude of the oscillation amplitude itself. A minimum amplitude is required for the linker to exert a measurable force on the cantilever to cause an effect on the top peaks of the oscillations for the generation of the recognition signal (Fig. 25.8b). On the other hand, the receptor-ligand bond is broken at each oscillation cycle when using too large amplitudes, and recognition events pass undetected during scanning. The range of appropriate amplitudes for recognition imaging is determined by the effective linker length (including the position and the length of the molecule attached to the linker), and is therefore rather sharply localized. Variation of the amplitude while keeping the ratio of amplitude set-point and free amplitude constant can thus be used as a quick specificity proof without perturbation of the receptor-ligand system as opposed to the widely applied block of either functionalized AFM tip or surface receptors by incubation with the free analog.

25.3.3 Application of TREC

TREC imaging has become a robust, accurate and reliable technique within the past decade. It has been applied to identify and localize specific receptor binding sites on a variety of isolated systems and more complex heterogeneous biosurfaces such as membranes, cells and whole tissues. On nanoarrays of self-assembled bacterial S-layer proteins *Tang* et al. [25.104] determined the precision of TREC in the detection of recognition sites. Strep-tag II, an eight-amino acid and artificial ligand for *Strep*-Tactin had been fused to a truncated form of the S-layer protein SbpA of *Lysinibacillus sphaericus* CCM 2177. Wild-type SbpA S-layer protein mixed with rSbpA-*Strep*tag II at a ratio of 7 : 1 was brought to self-assemble into a square (p4) lattice. TREC imaging with a *Strep*-Tactin functionalized AFM tip yielded high-resolution maps of the Strep-tag II binding sites with a surprisingly high positional accuracy of 1.5 nm (Fig. 25.9a).

In one of the first medical studies involving TREC, the density of cystic fibrosis transmembrane conductor regulator (CFTR) on the membrane of red blood cells was investigated [25.105]. CFTR is involved in the transport of fluids across epithelia. Mutations in the gene encoding CFTR can lead to dysfunctions and cause cystic fibrosis (CF), which is characterized

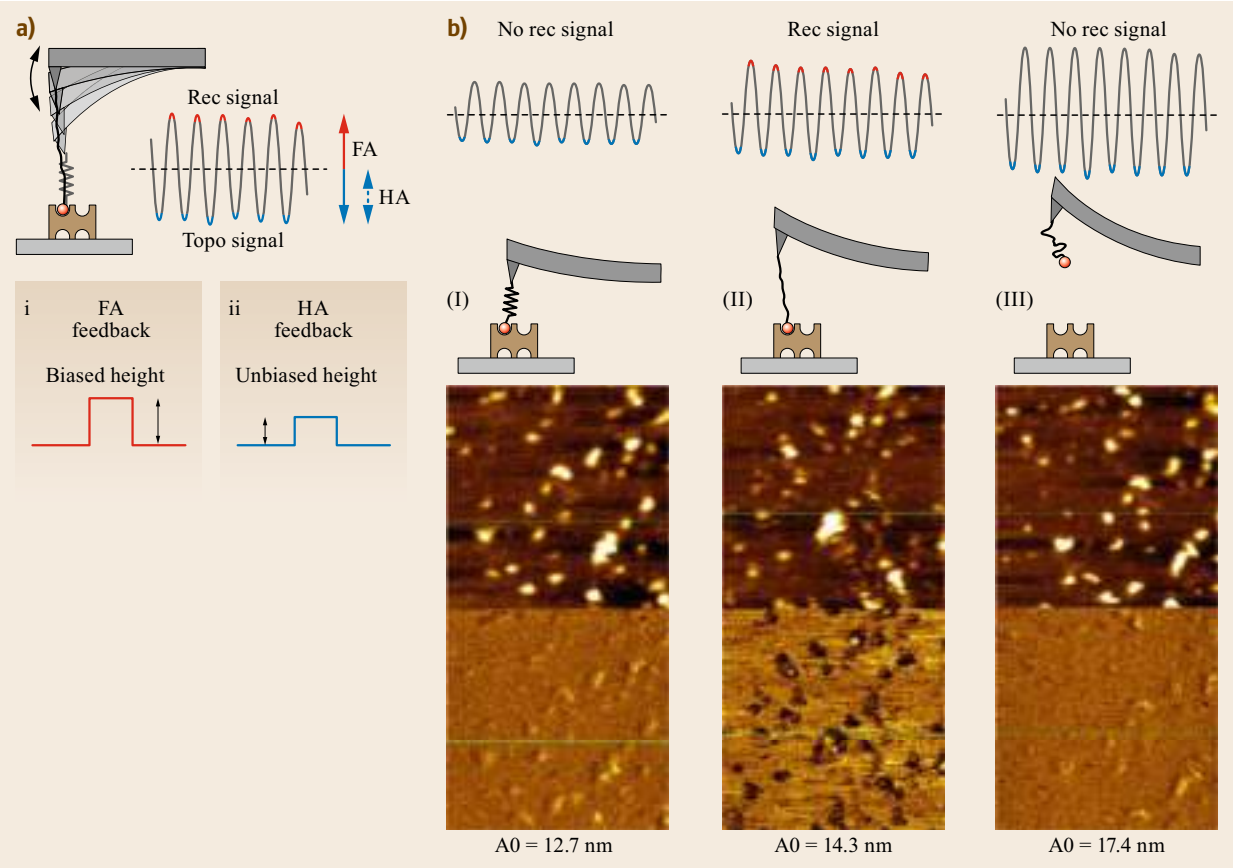


Fig. 25.8a,b Parameter optimization in TREC. **(a)** Comparison of half-amplitude feedback (HA) versus full-amplitude feedback (FA). When the tip-tethered ligand has bound to a surface-attached receptor, the topography is sensed at the bottom of the oscillation (*blue*) and stretching of the linker molecule reduces the movement in the upper parts of the oscillation (*red*). Consequently the FA parameter is biased by the additional amplitude reduction due to molecular recognition. In response the piezo-actuator will exaggerate vertical adjustment, which shows as an increased topographical height of the detected receptor molecule. Only operation of TREC using the lower HA feedback parameter will yield correct height information. **(b)** The oscillation amplitude is the most critical parameter in recognition imaging. If the oscillation amplitude is too small (I) the linker does not exert an efficient force on the cantilever upward swing and no recognition signal is generated. In the second regime (II), the amplitude has been increased, causing an efficient stretching of the linker without rupture of the ligand-receptor complex in the upward swing. This results in a pronounced recognition signal. In the third regime (III), the amplitude is too high, causing the rupture of the bond in every top peak of the oscillation and no stable recognition signal is generated

by a depletion of CFTR in the membrane of cells. A specific role of CFTR in erythrocytes is unknown at the current stage and no major hematological disorders have been reported for CF patients. Nevertheless, TREC confirmed a significant depletion of CFTR in red blood cells. About 800 CFTR proteins were localized in the erythrocyte plasma membranes of healthy donors while in CF patients the number was decreased to about 250. The advantages of TREC over methods relying on for instance quantum-dot-labeled antibodies [25.106] are the prevention of possible steric hindrance during binding of labeled antibodies to the target molecule and a direct visualization of molecules (Fig. 25.9b).

Another medical application for recognition imaging was carried out by *Rankl* et al. [25.107], where the distribution of corneodesmosin on the surface of stratum corneum (SC) was mapped at the nanoscale. Corneodesmosin, a protein known to be present in the SC, plays an important role in its physical integrity. With TREC, it was possible to characterize the morphological changes and the distribution of corneodesmosin on the surface of SC. The results are applied in the de-

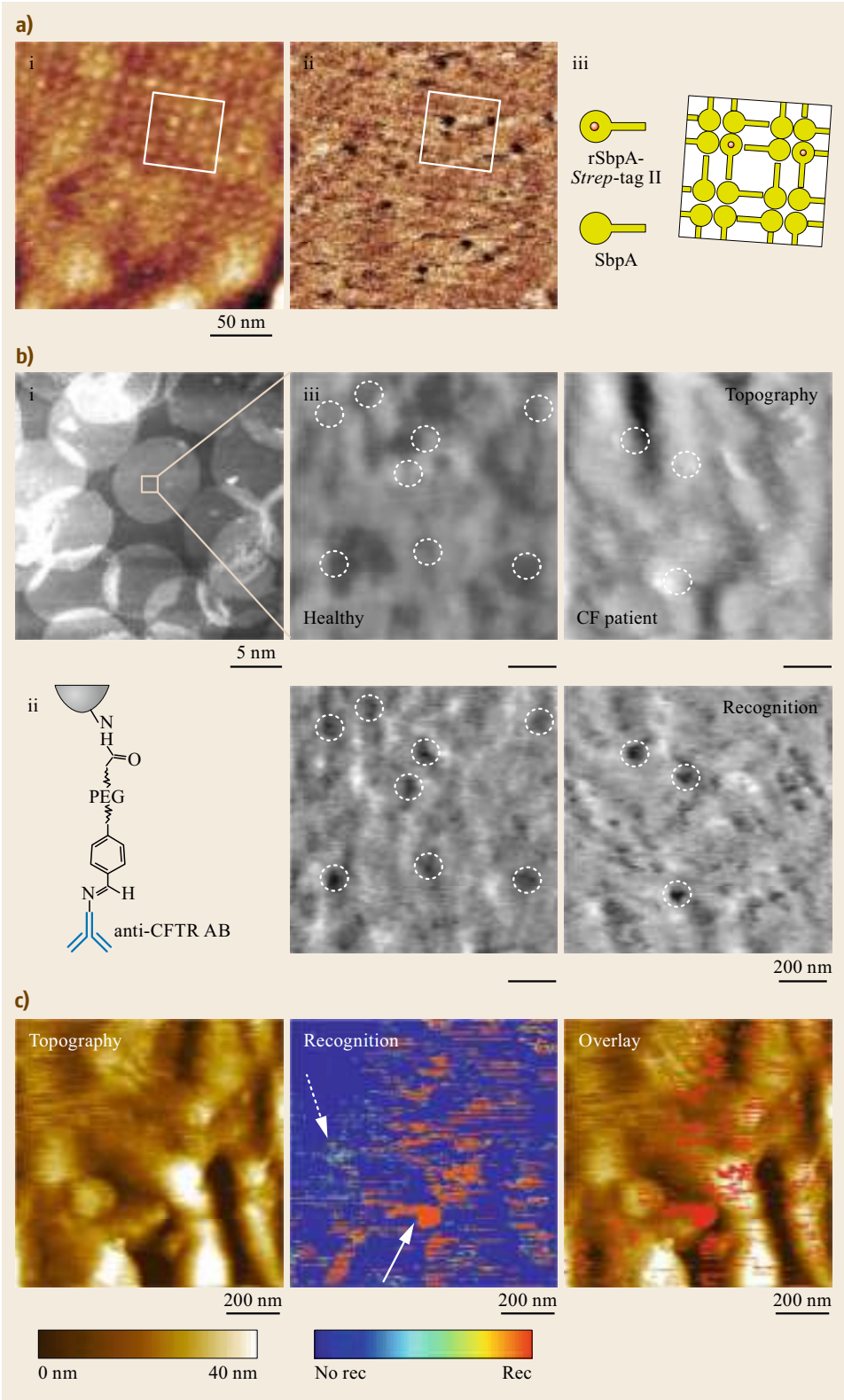

a)
i
ii
iii
rSbpA-
Strep-tag II
SbpA
50 nm
b)
i
iii
Topography
Healthy
CF patient
5 nm
ii
N
H
O
PEG
N
H
anti-CFTR AB
Recognition
200 nm
c)
Topography
Recognition
Overlay
200 nm
200 nm
200 nm
0 nm
40 nm
No rec
Rec

Fig. 25.9a–c Biosensing TREC applications. **(a)** Topography (i) and recognition (ii) images of the S-layer proteins rSbpA-*Strep*-tag II and wild-type SbpA (in a molar ratio of 1 : 7) cocrystallized on a silicon chip as imaged with a *Strep*-tag II binding *Strep*-Tactin functionalized AFM tip. The schematic drawing illustrates the regular arrangement of the cocrystallized proteins and the ratio of *Strep*-tag II, which is perfectly reflected in the recognition image at a lateral resolution of 1.5 nm. **(b)** Medical sensing application of TREC for identification of cystic fibrosis (CF) patients. (i) Inside-out oriented erythrocyte membrane patches were scanned with an anti-CFTR-antibody functionalized tip for quantification of CFTR receptors (ii). (iii) Topography (*top*) and recognition (*bottom*) images of healthy and CF donors reveal fewer recognition events between tip-bound anti-CFTR and CFTR in the erythrocyte membrane as a clear indication of the disease. **(c)** Simultaneous topography and recognition images of the stratum corneum of a skin sample using a tip coupled with an antibody against corneodesmosin, which is implicated in the process of skin peeling. Recognition spots of a few (*dashed arrow*) up to 100 nm (*white arrow*) were found. The overlay of both images shows that most of the recognition spots are located on top of pronounced topographical features ◄

sign of cosmetic compounds that will allow better interaction with the SC to the benefit and improvement of skin appearance (Fig. 25.9c).

Just recently, TREC imaging was applied to visualize uncoupling protein 1 (UCP1) reconstituted into lipid bilayers and to analyze the adenosine triphosphate (ATP)-protein interaction at a single-molecule level [25.113]. UCP1 is a mitochondrial membrane protein and is involved in nonshivering thermogenesis under cold-acclimating conditions. In the last few years, this protein has become increasingly important as a target protein for the therapy of obesity as well as inflammatory, neurodegenerative and ischemic diseases. Based on this system, a new mode for mapping and studying interaction sites at low lateral density was devised by combining recognition imaging and force spectroscopy. This method makes it possible to select a single receptor molecule reconstituted in a supported lipid membrane and then quantify receptor-ligand unbinding forces. Apart from these few selected examples a comprehensive overview of TREC studies can be found in Table 25.2.

Table 25.2 Overview of TREC studies in the years 2006–2016 (no claim to completeness)

Sample	Study	Author, reference
Endothelial cells	Identification and visualization of vascular endothelial (VE)-cadherin that is known to play a crucial role in calcium-dependent, hemophilic cell-to-cell adhesion.	*Chtcheglova* et al. [25.108, 109]
DNA aptamers	In vitro selection and evaluation of DNA aptamers selected to bind histone H4 tails.	*Lin* et al. [25.110]
Red blood cell membranes	Density of cystic fibrosis transmembrane conductor regulator (CFTR) on the membrane of red blood cells was investigated.	*Ebner* et al. [25.105]
Cardiac cells	Inhibition of hERG K^+ channel is major cause of long QT syndromes inducing fatal cardiac arrhythmias mediated by ergtoxin. TREC was used to visualize ergtoxin receptors on the voltage-sensing domain of hERG K^+ channels.	*Chtcheglova* et al. [25.111]
Bacterial surface layer (S-layer)	Visualizing binding sites on bacterial S-layer nanoarrays containing affinity tags (Strep-tag II) to within a positional accuracy of 1.5 nm. The nanoarrays were then used as a template for constructing highly ordered molecular binding blocks.	*Tang* et al. [25.104]
Chromatin	To provide insights into epigenetic processes, Wang et al. used TREC to perform single-epitope recognition imaging of native chromatin.	*Wang* et al. [25.112]
Stratum corneum	Mapping the distribution of corneodesmosin on the surface of stratum corneum (SC) at the nanoscale for designing new cosmetic compounds that will better interact with the SC surface for a specific skin benefit.	*Rankl* et al. [25.107]

Table 25.2 (continued)

Sample	Study	Author, reference
Macrophages	Precise nanomapping revealed compact cluster organization of FcγRs that would favor the rapid recognition of huge pathogens and thus their effective internalization.	*Ahmad* et al. [25.114]
Human eye lens	Recognition imaging of whole, untreated human tissue samples to map the distribution of abnormal protein aggregates on the anterior lens capsule, which is a well-known cause of irreversible blindness.	*Creasey* et al. [25.115, 116]
Protein-nucleic acid interaction	TREC can also be utilized to study protein-nucleic acid interactions, which play important roles in the physiological behavior of living cells and mapped molecular details of the ricin-aptamer interaction.	*Wang* et al. [25.117]
Plant cells	Mapping the structural changes of the cell wall of natural and pretreated plant cells provides better understanding of surface structural changes, which leads to a potential improvement of biomass pretreatment for biofuel production.	*Zhang* et al. [25.118]
Mitochondrial membrane	Visualization of uncoupling protein 1 (UCP1) reconstituted into lipid bilayers and analysis of the ATP-protein interaction at a single-molecule level. Based on this study, TREC was further developed to quantify also binding forces on interaction sites at low lateral density.	*Zhu* et al. [25.119], *Koehler* et al. [25.113]
Monocytes	CD1d molecule, a monomorphic major histocompatibility complex class I-like molecule, presents different types of glycolipids to invariant natural killer T (iNKT) cells that play an important role in immunity to infection and tumors, as well as in regulating autoimmunity. TREC analysis was conducted to nanomap single CD1d molecules on THP1 cells that were loaded with different glycolipids.	*Duman* et al. [25.120]
Bladder cancer cells	Human gonadotropin-releasing hormone receptor (GnRH-R; or type I GnRH-R) that expressed in tumor cells, has gained a great attention as a specific target for cancer therapy. Given the clinical utility, the improved nano-organization of GnRH-R is an important step in the development of more effective and possibly new therapeutic strategies.	*Zhang* et al. [25.121]
Melanoma cells	Hsp70 is known to be present at the extracellular leaflet of the plasma membrane (surface Hsp70, sHsp70) of most tumor cells but not that of the corresponding benign cells. TREC analysis was conducted to elucidate its protumor role.	*Nimmervoll* et al. [25.122]

25.4 AFM Biomedical Sensing – Examples

In the previous sections AFM techniques have been introduced that present a multifunctional toolkit for an ever-increasing diversity of biological and life sciences applications. The AFM has evolved into an enabling instrument to observe structural details of molecular assemblies, isolated membrane preparations and cell surfaces, and to monitor cellular responses to external stimuli such as antibodies or drugs. In the following studies topographical imaging, SMFS, TREC and their combination were employed to answer questions

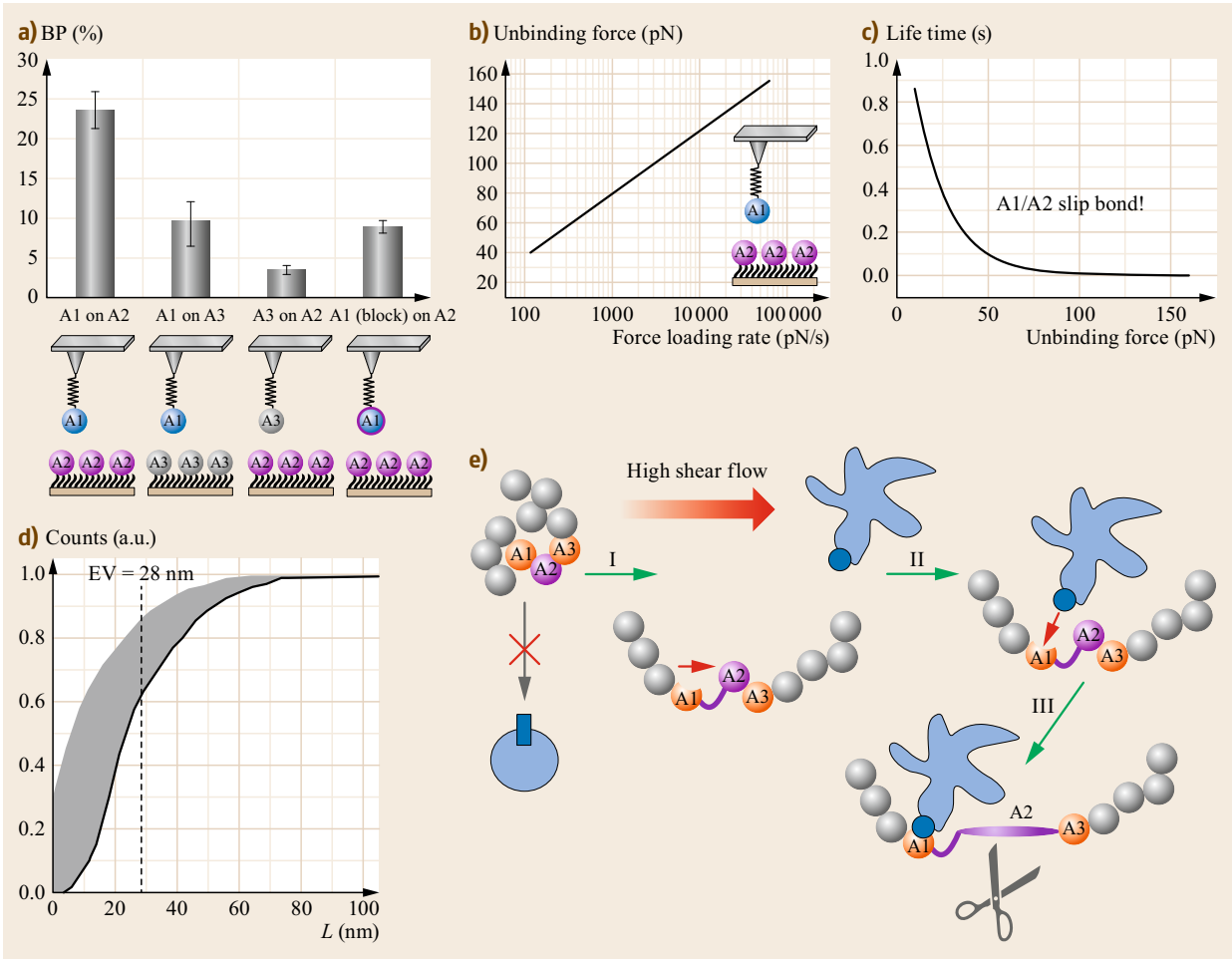


Fig. 25.10a–e **(a–d)** SMFS investigations of the functions of the von Willebrand factor (VWF) in blood coagulation. **(a)** Number of specific binding events between VWF A1 and A2 domains, where A1 was coupled to the AFM tip and A2 was immobilized on the surface. In control experiments either the A2 domains on the surface or the A1 domain at the AFM tip were replaced by the unspecific VWF A3 domain or the tip-bound A1 was saturated by soluble A2. **(b)** Loading rate dependence of the VWF A1–A2 interaction. **(c)** The lifetime versus unbinding force plot clearly indicates slip bond behavior of the A1–A2 interaction. **(d)** Cumulative distribution of the elongation L of the A1-A2 complex prior to dissociation (*black line*) and the corrected value after subtraction of the size of A1 and the linkers (*gray area*). The *dotted line* indicates the expectation value (EV) for L. The measured EV ($\approx$ 28 nm) was substantially lower than the extension of a fully stretched unfolded A2 domain ($\approx$ 80 nm). **(e)** Suggested model for the VWF A1–A2 interaction. Higher shear stress due to e.g., an injury activates VWF to interact with the platelet GPIbα receptor. The GPIbα binding site in domain A1 is made accessible by dissociation of domain A1 from A2 under high shear flow conditions (I, II). Domain A2 remains largely folded during this process protecting it from cleavage and degradation (II). SMFS data imply that dissociation predominantly occurs before exposure of A2 to cleavage. Thus, the A2 domain might prevent platelet binding in a force-dependent manner to ensure that VWF is entirely activated before cleavage

Fig. 25.11a–e Investigation of ATP binding to UCP1 using TREC. **(a)** Illustration of the TREC experiment with the uncoupling protein 1 (UCP1) reconstituted into a supported lipid bilayer on a mica surface and a UCP1-specific antibody or ATP coupled to the AFM tip. **(b)** High-resolution topographical (*top*) and UCP1 antibody-recognition (*bottom*) images. Correlation shows that some UCP1 molecules are recognized (*circles of solid line*), whereas a few remain unrecognized (*dashed*). **(c)** High-resolution ATP-recognition image of UCP1 in the bilayer reveals large (*thick arrows*) and small (*thin arrows*) spots, whereas the molecules' sizes in the topography (not shown) were homogeneous. Evaluation of the diameter of the recognition spots shows a bimodal statistical distribution. **(d)** Schematic illustration of the expected dependence of recognition spot size on the orientation of UCP1 in the bilayer and on the length of the crosslinker that tethers ATP to the AFM tip. **(e)** Proposed mechanism of ATP binding: initial binding from the c-side (I) followed by conformational change of UCP1 (II) and protein inhibition. (III) Protein binding from the m-side without conformational change and inhibition ►

in basic molecular biology, in pharmaceutical science and the medical field that so far were inaccessible by other techniques. Thus, the selected studies illustrate and highlight the biomedical sensing capabilities of the AFM.

25.4.1 Isolated Membranes

SMFS Investigation of Blood Coagulation – von Willebrand Factor

The von Willebrand factor (VWF) is a glycoprotein that plays a central role in hemostasis, the first stage of wound healing. It triggers platelet adhesion in areas of vascular damage by binding to exposed subendothelial collagen and thus causes wound closure. In an initial step the VWF binds to collagen with specific binding sites that are located in domains A1 and A3 of the glycoprotein. Collagen III is believed to interact with the A3 domain, and collagen VI with the A1 domain. The forces and the dynamics involved in these interactions were investigated by SMFS with the VWF tethered to the AFM tip and substrates coated with a dense layer of collagen [25.123]. Different mutations in the A3 domain of the VWF A1-A2-A3 construct were investigated. These mutations (H1786R, Q1734H and S1731T) correspond to the type 2 M of the current von Willebrand disease (VWD) classification system, and are characterized by a defect in the collagen binding domain but exhibit otherwise normal multimeric distribution. Surprisingly, patients carrying these mutations show no or mild symptoms, or have an alternative reason for impeded coagulation. These mutations were diagnosed purely by accident and thus their clinical significance was not clarified up to now. It could be shown that defective interaction between one VWF domain and one type of collagen can be compensated by alternative binding events. This observation might explain the inconspicuous bleeding tendency found in patients carrying these VWF A3 domain mutations.

In a follow-up study the interplay of the VWF domains A1 and A2 during the onset of blood coagulation were investigated more closely [25.124]. The VWF domain A1 is responsible for mediating platelet adhesion under flow through the platelet glycoprotein Ibα (GPIbα). The adjacent VWF domain A2 unfolds under shear, upon which it exposes a cleaving site for the protease ADAMTS13 to prevent thrombosis. A combination of molecular dynamics simulations (MDS), AFM (Fig. 25.10a), and microfluidic experiments, revealed that the A1–platelet interaction is blocked by mutual A1–A2 binding under low shear. In order to elucidate the exact shielding mechanism, SMFS was used to mimic typical elongation forces imposed by the shear flow of blood [25.125]. The interaction strength between the A1 and A2 domain (Fig. 25.10b,c) was probed together with the unfolding behavior of A2 (Fig. 25.10d) as an indicator of the shielding capacity. As a result only a small extent of A2, if at all, unfolded before dissociation from A1. These findings suggest that the platelet binding site in domain A1 is made accessible by the slip-bond dissociation of domain A1 from A2 and that domain A2 remains largely folded during this process, which keeps it protected against cleavage and degradation (Fig. 25.10e).

TREC Studies to Elucidate ATP-Binding to UCP1 in the Mitochondrial Membrane in Non-Shivering Thermogenesis

Uncoupling protein 1 (UCP1) is a transmembrane protein in the mitochondrial membrane of brown adipose tissue (BAT). The protein dissipates the energy of the proton gradient over the inner mitochondrial membrane as heat. This process is also known as non-shivering thermogenesis under cold-acclimating conditions [25.126] and is essential to newborns to counter heat loss, but also prevents energy storage as fat. Thus, UCP1 has been proposed as a pharmaceutical target for the treatment of obesity in adults [25.127, 128]. Long-chain fatty acids activate UCP1 proton transport, whereas purine nucleotides have an inhibitory effect [25.129–133]. Using TREC the mechanism of nucleotide-mediated binding and inhibition of UCP1 inside the mitochondrial membrane was elucidated on a structural level [25.119].

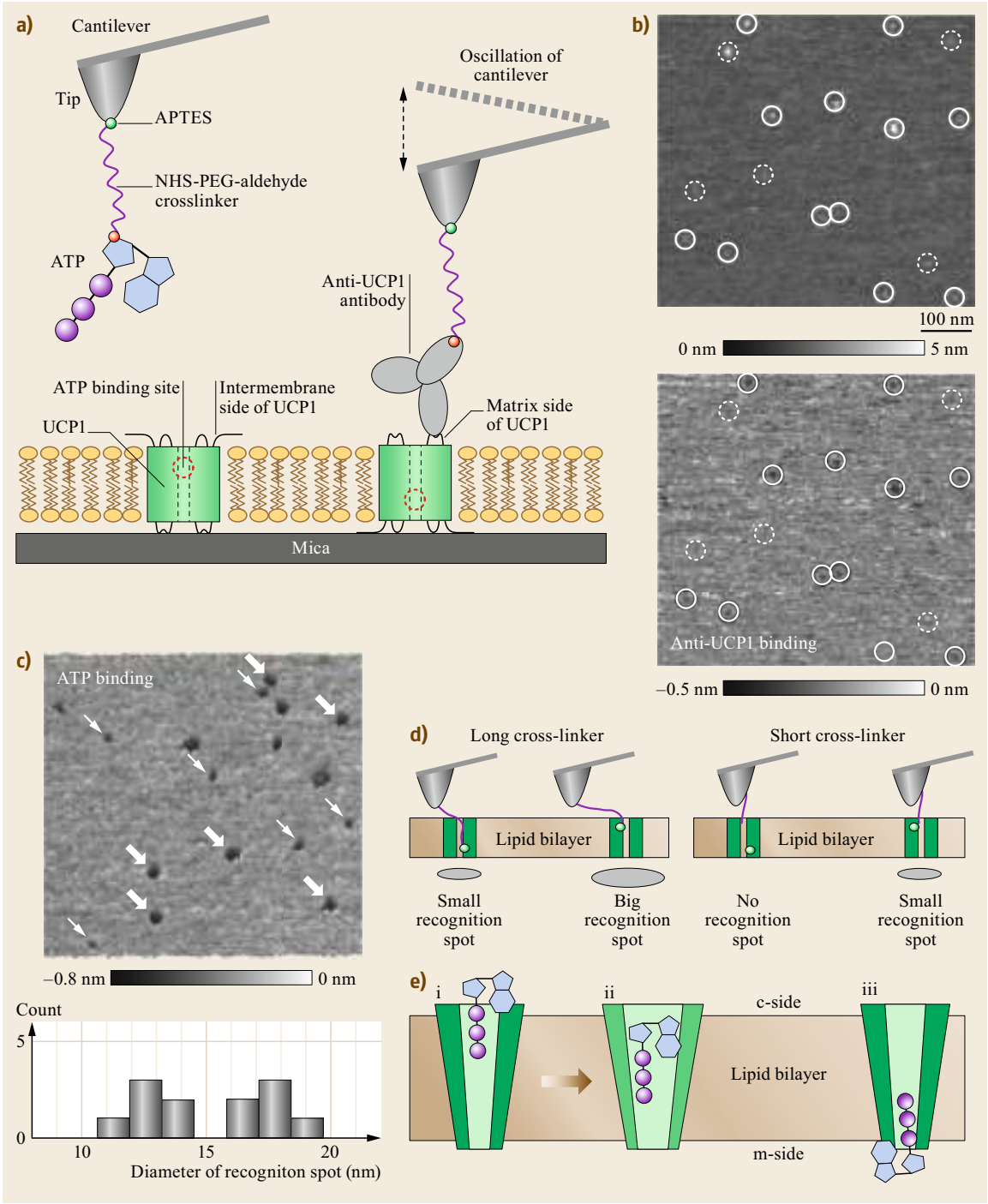


Figure 25.11a shows the experimental setup for TREC measurements of UCP1 reconstituted into lipid bilayers on a mica surface and a cantilever tip functionalized either with ATP or an anti-UCP1 antibody specific to amino acid residues 145–159 of UCP1 using acetal-linker chemistry. Only about half of the protein molecules that were visible in the topographic image were also recognized by anti-UCP1 AB (Fig. 25.11b). This result shows that the orientation of UCP1 protein in the planar lipid bilayer is random, as expected.

The specificity of this interaction was proved by injection of a peptide that competed with UCP1-anti-UCPI AB interaction. Contrarily, when the cantilever was functionalized with ATP instead of the UCP1-AB, colocalization of proteins in the topography and signals in the recognition image was close to 100%. As the orientation of UCP1 in the membrane was determined to be random, this result was unexpected and implies that the nucleotide binding sites are accessible irrespective of the orientation of the protein in the reconstituted membrane. Whereas the size of protein spots in the topographic image was uniform, the size of the recognition spots appeared different. Two groups with mean diameters of 12.65 ± 1.03 and 17.68 ± 0.86 nm were determined (Fig. 25.11c, ii). Given that the ATP binding site remains accessible no matter which side of the protein is immersed in the layer or facing upwards, the existence of two different sizes of recognition spots can be explained by differences in submersion depths of the ATP molecule on the tip and the linker. To estimate the localization of the binding site, ATP-functionalized tips with different lengths of linkers were used (Fig. 25.11d, iii). From these data the depth of the ATP binding site in UCP1 was calculated to be 1.27 ± 0.10 nm from the membrane surface.

Taking the NMR structure of UCP2 (57% homology to UCP1) [25.134, 135] and the proposed three-step purine nucleotide (PN) binding model for UCP1 [25.136] into account, we hypothesized that only binding of ATP to the cytosolic side (c-side) would lead to a conformational change in the protein, which then causes inhibition. PN binding from the matrix-facing side (m-side) would not induce a conformational change and thus not trigger UCP1 inhibition (Fig. 25.11e). This suggestion might plausibly explain one of the unresolved bio-energetical questions: how can the protein be in the active state despite the presence of mM concentrations of ATP in the cytosol? In the absence of the crystallographic structure of UCP1, TREC data yield valuable insights into the nucleotide binding mechanism in the process of nonshivering thermogenesis. High-resolution topographical AFM images provided for the first time visualization of UCPs reconstituted into lipid bilayer membranes. Simultaneously measured recognition images revealed the existence of a single ATP binding site that can be accessed by nucleotides from both the c-side and the m-side.

25.4.2 Mammalian Cells

Drug Binding to the Serotonin Transporter (SERT)

The serotonin transporter (SERT) is a member of the neurotransmitter sodium symporter (NSS) family [25.137] and retrieves extracellular serotonin back into the presynaptic neuron to terminate neurotransmission. Some neurological conditions, e.g., depression, anxiety disorder, attention-deficit hyperactivity disorder (ADHD), autism, Parkinson's disease, and addiction are related to NSS dysregulation. SERT is a target both for therapeutic medicines such as citalopram (CIT) to treat depression and for illicit drugs like cocaine, which induces severe health problems. X-ray analysis of SERT homologues, i. e., bacterial $LeuT_{Aa}$ [25.138] and dDAT (the *Drosophila melanogaster* dopamine transporter) [25.139, 140] indicate a primary binding site (S1-site) for substrates in the membrane-spanning region. Other studies provided hints for a second S2-site, located within the extracellular vestibule [25.141, 142]. To elucidate the number and the mechanism of ligand binding site(s) in SERT, we used SMFS to measure the interaction forces between SERT and S-enantiomers of CIT (S-CIT) [25.143].

An alkyne-modified 5-aminomethyl analogue [25.144] of the S-CIT was covalently conjugated onto the AFM cantilever tip (Fig. 25.12a) using click chemistry. SMFS was conducted on living CHO-K1 cells expressing human SERT [25.145] in 2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid (HEPES) buffer at room temperature. Force curves with unbinding events were observed during the repeated measurements, and the experimental probability density function (PDF) of forces was thus generated (Fig. 25.12b, solid curve). The two distinct peaks in the force PDF reflect the S-CIT rupture from binding sites with two different strengths of interaction. To identify the location of the binding sites for the two populations, force measurements were performed on SERT-G402H, which contains a point mutation in the outer vestibule [25.142]. The force PDF from SERT-G402H (Fig. 25.12b, *dotted curve*) showed only a single peak, which corresponded to the second peak in the force PDF from wild-type (wt) SERT. This indicates that the first peak in the force PDF from wt SERT originates from the vestibular S2 site. In addition, we verified that Na^+ ions were important for accessibility to the SERT binding sites by recording forces of S-CIT binding to wt SERT in Li^+ buffer. The PDF showed a single peak in the force PDF (Fig. 25.12c, *dotted curve*) that coincided with the first peak in the force PDF observed in Na^+ buffer (Fig. 25.12c, *solid curve*). The absence of the second peak is consistent with the notion that the central S1 site is Na^+ dependent. The persistence of the first peak implies that access to the vestibule S2 site does not require Na^+.

To examine if there is an allosteric effect [25.146] between the S1 and S2 sites, we performed competition experiments with 30 μM S-CIT in solution so that

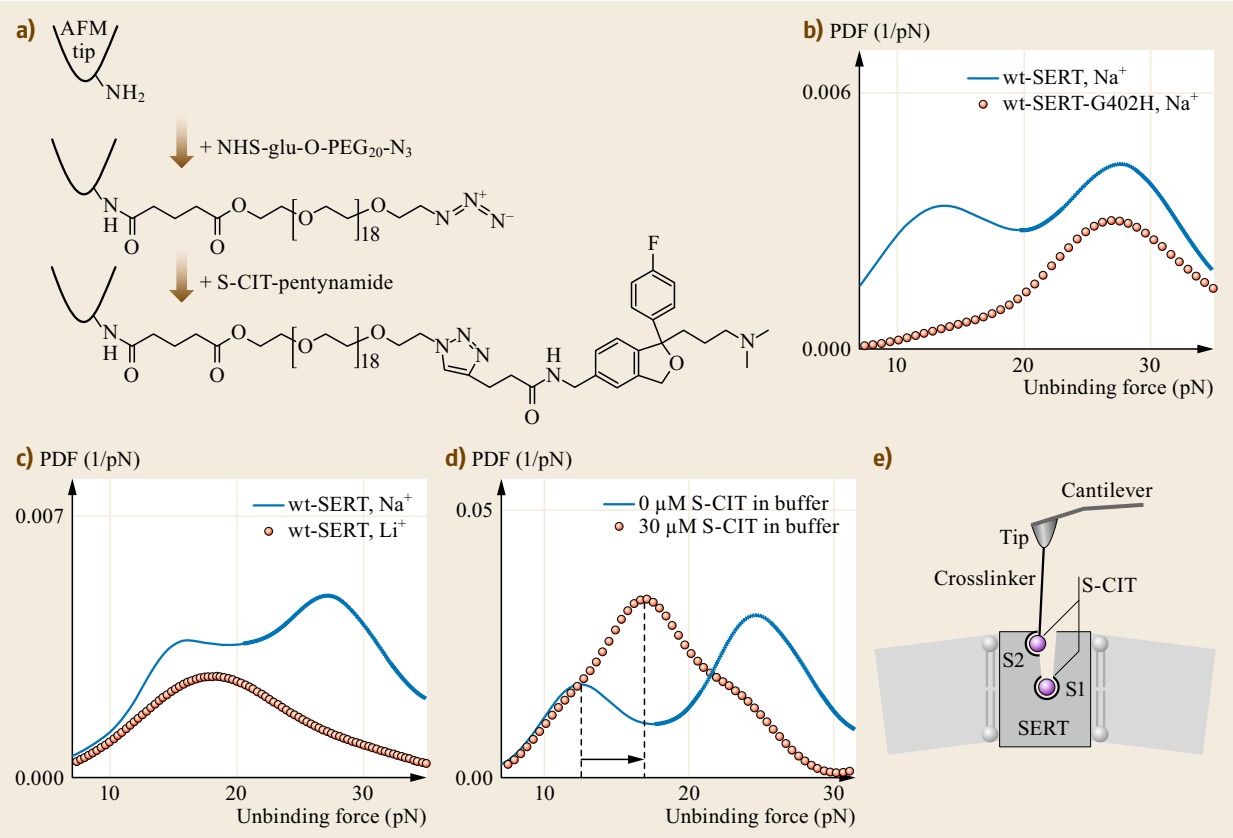


Fig. 25.12a–e Drug binding to the serotonin transporter (SERT). **(a)** Conjugation of S-CIT to the AFM cantilever tip. After the cantilever was aminofunctionalized with gas-phase APTES, it was pegylated with NHS-glu-O-PEG20-N3. The alkyne-modified S-CIT analogue was coupled to the azide-terminated PEG via cocatalyst-accelerated copper(I)-catalyzed azide-alkyne cycloaddition. **(b)** Comparison of force distribution measured on wild type (wt) SERT (*solid line*) and mutant SERT-G402 (*dotted line*). **(c)** Comparison of force distribution measured in Na^+ buffer (*solid line*) and Li^+ buffer (*dotted line*). **(d)** Comparison of force distribution measured on wt-SERT in Na^+ buffer without (*solid line*) or with 30 µM S-CIT (*dotted line*). After the S-CIT in solution bound at S1 site **(e)**, the binding strength between S-CIT and the S2 site was enhanced, shown as the shift of the peak position in **(d)**, indicating a positive allostery

the higher-affinity S1 site was occupied whereas the lower-affinity S2 site was still accessible to S-CIT on the tip. Consequently, only a single peak remained in the force PDF (Fig. 25.12d, *dotted curve*), which was attributed to the S2 site. This is due to the long lifespan (≈ 20 min) of S-CIT at the S1 site [25.141]. Meanwhile, the AFM tip-tethered S-CIT, which has a local concentration of ≈ 3.7 mM (corresponding to one S-CIT molecule within a hemisphere of radius ≈ 6 nm, defined by the length of PEG linker), is likely to effectively compete with the S-CIT in solution for binding to the S2 site. The remaining peak shifted to the right (*arrows* in Fig. 25.12d), indicating that the binding of S-CIT at the central S1 site exerts a positive allosteric effect on the S2 site (Fig. 25.12e). Such an allosterically enhanced binding strength of S-CIT at the S2 site should provide the possibility for x-ray crystallography to capture the S2 site. Three months after the demonstration of the two S-CIT binding sites in SERT by our SMFS study [25.143], the x-ray structure of SERT bound with two S-CIT molecules was published [25.147], which verified our results.

Given the importance of allosteric regulation in biology and pharmacology, the SMFS-based nanopharmacological approach paved a new avenue to exploring transient binding sites in clinically relevant membrane transporters and opened the door to quantitatively addressing the modulation of interaction forces between ligands and allosterically coupled binding sites.

Fluorescence-Guided SMFS to Confirm Increased GLUT4 Translocation

The combination of AFM instrumentation with optical microscopy [25.148] has greatly eased biomedical sensing applications of the AFM on cells with the possibility to correlate independently or simultaneously acquired topographical and optical-fluorescence images [25.149, 150] and to guide the AFM tip in SMFS experiments [25.151]. In a recent application, optically guided AFM was employed to reveal the translocation of glucose transporter 4 (GLUT4) in cells [25.152]. GLUT4 is the key molecule for uptake of glucose into adipocytes and muscle cells and thus reduces blood glucose levels to prevent hyperglycemia. In a basal state, GLUT4 is trapped in intracellular compartments. Insulin induces the translocation of GLUT4 into the plasma membrane and thus triggers glucose uptake. Resistance to insulin, an important parameter of type-2 diabetes, strongly reduces this translocation process [25.153].

There is a great demand for analytical techniques capable of quantifying the GLUT4 translocation efficiency in order to search for compounds that are able to enhance the GLUT4 translocation process in the absence of insulin (insulin mimetics substances). Most approaches such as western blot analysis of membrane fractions, enzyme-linked immunosorbent assay (ELISA)-guided binding assays, flow cytometric analysis [25.154–156] and immunoelectron techniques [25.157] suffer from being indirect, time consuming, nonquantitative and/or more expensive. An important improvement was the insertion of different epitope tags (mainly myc and HA) in the first extracellular loop of the GLUT4 protein in combination with genetically encoded fluorophore molecules (mainly green fluorescent protein(GFP)) at the c-terminus of the transporter [25.158, 159] as depicted in Fig. 25.13a. Here, dual-color total internal reflection fluorescence (TIRF) microscopy has been used to track GLUT4 translocation in the CHO-K1-hIR-GLUT4-myc-GFP cell line. Nevertheless, a critical point of such TIRF experiments is to prove that the increase of fluorescence in the evanescent field in response to insulin or insulin mimetic compounds is solely caused by the increase of GLUT4 molecules in the plasma membrane, and not by the increase of recruited vesicles in the TIRF zone. The use of fluorescently labeled antibodies that target the myc-tag in the first extracellular loop of GLUT4-myc-GFP provides the possibility to further reduce uncertainty. However, before incubation with the reporter molecule, cells are commonly fixed with paraformaldehyde, which may slightly permeabilize the cells, potentially leading to staining of intracellular GLUT4 molecules as well. In order to clarify this uncertainty, fluorescence-guided SMFS experiments were performed with the anti-myc antibody covalently coupled to the AFM tip. In this configuration the antibody can only interact with GLUT4-myc-GFP fraction on the cell membrane.

For SMFS experiments the anti-myc antibody was covalently bound to the AFM tip (nominal spring constant 0.02 N/m, MSCT, Bruker, CA, USA) via a bifunctional NHS-PEG-acetal linker (Fig. 25.13a). Highly contrasted optical (phase contrast) and epifluorescence images were acquired with a 20× objective to enable positioning of the AFM cantilever above CHO-K1 cells that stably express the GLUT4-myc-GFP construct (Fig. 25.13b) and force-distance curves (FDCs) were collected. For each condition, i. e., either deprived of serum (starved) or insulin stimulated, at least three different cells expressing GLUT4 receptors were tested using in sum three different AFM tips. Each AFM/epifluorescence experiment was performed in duplicate using independent cell preparations. For each dataset of FDCs (> 500 FDCs per set) the binding probability (BP) was determined as the number of force curves that show specific unbinding events of anti-myc antibody and GLUT4-myc-GFP molecules on the cell membrane divided by total number of collected force curves. A significant difference between starved (BP ≈ 2.4%) and insulin-stimulated cells (BP ≈ 24.6%) was clearly recognized (Fig. 25.13c).

This fluorescence-guided SMFS study elegantly and decisively confirmed the improved accessibility of the myc-tag in the first extracellular loop of GLUT4 only after insulin stimulation. In a follow-up study additional AFM-based techniques such as TREC will help to uncover the nano-organization of GLUT4 proteins on cell membranes, their oligomeric states, and as well the cortical cytoskeleton rearrangement. A detailed comparison of the effect of insulin and insulin mimetic phytochemicals will help in the identification of new compounds that increase GLUT4 translocation, oligomerization, domain formation, and cytoskeletal coupling for treatment of type 2 diabetes.

25.4.3 Bacteria

For decades, the methods of choice for studying physiology and behavior of prokaryotic organisms were growth in batch cultures and on agar plates [25.160]. In the last few years it has been recognized that apart from chemical cues, mechanical interactions through cell-cell and cell-surface contact also play a significant role in microbial biology on surfaces. Thus, characterization of surface structures, binding mechanisms and identification of the main molecular players has moved into the spotlight for understanding fundamental pro-

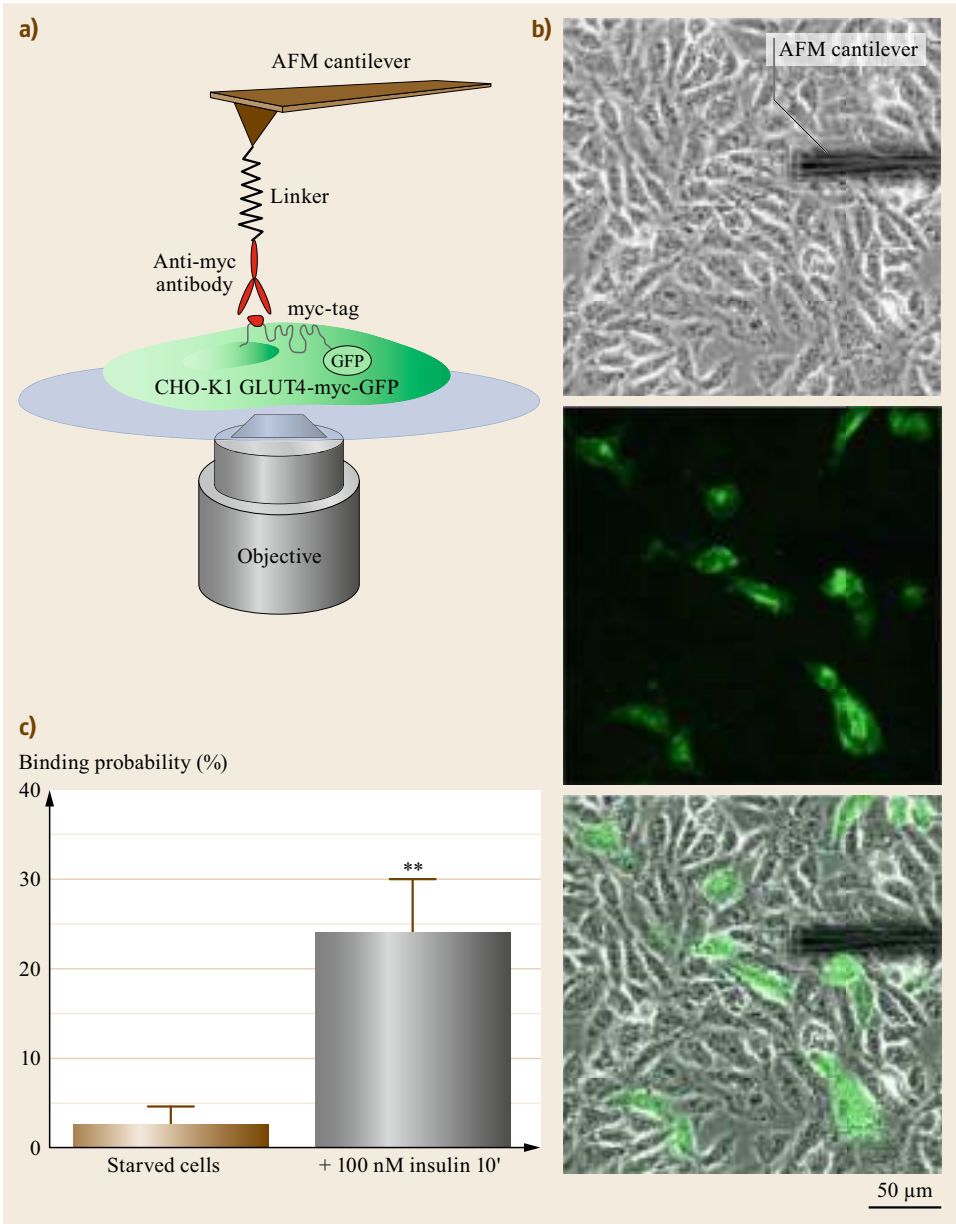


Fig. 25.13a–c Fluorescence-guided AFM force detection of GLUT4 molecules in the membrane of CHO-K1 hInsR/GLUT4-myc-GFP cells. **(a)** Schematic representation of an integrated platform combining an AFM with an inverted light microscope to visualize and recognize the insulin-dependent GLUT4 glucose transporter by using a genetically encoded fluorophore (GFP) at the c-terminus and a myc-tag in the first extracellular loop. An anti-myc antibody functionalized AFM tip was used to exclusively identify membrane-inserted GLUT4 molecules. **(b)** By overlapping highly contrasted optical (phase contrast) and fluorescence images it was possible to visualize the AFM cantilever and thus to guide it on a cell of interest. **(c)** Force measurements results (binding probability) of starved and insulin-treated (cells were incubated with 100 nM of insulin for 10 min) GLUT4-myc-GFP hInsR cells. $^{**}P < 0.01$, significant increase with respect to the starved cells

cesses such as bacterial adhesion, surface recognition, biofilm development, and the incipient stage of infection.

Advances in AFM techniques provide us with powerful tools to explore the physical and mechanical properties of bacteria in unprecedented detail. Apart from analyzing nanostructures on the cell surface by topographical imaging, indentation measurements yield physical properties of single bacteria such as cell wall stiffness, Young's modulus, internal turgor pressure, and the stretching modulus as a function of environmental conditions like medium ionic strength. Such measurements have revealed that the nanomechanical properties of bacteria can impose mechanics to drive the adaptive behavior [25.161]. Additionally, force indentation and force spectroscopy approaches using AFM to study these effects have become available only recently by measuring processes occurring at the interface between microorganisms and cells [25.162, 163].

Bacterial Surface Layer (S-Layer) of *Tannerella Forsythia*

Crystalline bacterial cell surface layers (S-layer) are the most common surface structure in bacteria. S-layers may hold great merit in nanotechnology due to their inherent ability to self-assemble into an ordered protein lattice on solid surfaces [25.164]. However, from a clinical point of view the outermost cell surface is a virulence factor, as it mediates adhesion to and invasion of host cells. Thus, it is an important task to characterize the biochemical nature and structure of S-layers in pathogenic bacteria under live cell conditions. *Tannerella forsythia* (*T. forsythia*) is a gram negative oral pathogen belonging to the so-called *red-complex* bacteria together with *Porphyromonas gingivalis* and *Treponema denticola*, and has been implicated in onset and development of periodontal diseases in humans [25.165]. Clinical evidence accumulates that *T. forsythia* also impacts systemic health, but the underlying mechanisms of pathogenicity are only slowly beginning to unravel [25.166].

MAC-mode imaging was employed to determine the S-layer ultrastructure on intact *T. forsythia* cells at the nanometer scale (Fig. 25.14) [25.167, 168]. For noninvasive imaging, 1 ml of *T. forsythia* culture was collected, washed and immobilized by mechanical trapping on 0.8 µm polycarbonate membranes. The wild-type (wt) cells are $3.5 \pm 1\,\mu m$ in length and $0.6 \pm 0.1\,\mu m$ in width, resulting in a large curvature that makes proper scaling of the height image difficult. Hence, the amplitude image provides significantly better local contrast and the fine structure in the surface topology is more clearly resolved (Fig. 25.14b). The *T. forsythia* wt bacteria S-layer possesses a distinct, periodic square (p4) lattice, as clearly seen in the high-resolution reconverted three-dimensional (3-D) fast Fourier transform (FFT) image obtained from an enlarged topography (Fig. 25.14c). Peaks are due to surface periodicities, which are represented in the power spectra as the bright peaks near to the origin (*inset* in Fig. 25.14c). The S-layer lattice periodicity is 8.74 ± 0.35 nm with an angle of $90 \pm 4°$, which is comparable to earlier observations made by electron microscopy on freeze-fractured, freeze-dried and ultrathin sectioned preparations [25.167].

The AFM data clearly indicate that the outer membrane of intact live *T. forsythia* wt bacteria has a continuous, symmetrical two-dimensional lattice structure built by two distinct S-layer protein species (*TfsA* and *TfsB*). In order to further explore the arrangement of *TfsA* and *TfsB*, SMFS was used to identify the specific interactions between an anti-*TfsA* antibody and the *TfsA* glycoprotein (*TfsA*-GP) within the *TfsA/TfsB* S-layer lattice. The anti-*TfsA* antibody was tethered to the AFM tip via a standard NHS-PEG-aldehyde linker utilizing a lysine residue as a reactive site in the antibody. Force-distance curves were measured on intact *T. forsythia* wt cells indicating specific binding of the antibody to *TfsA*-GP on the bacterial S-layer (Fig. 25.14d). Each rupture event was used to calculate the distribution of specific unbinding forces, fitted with a Gaussian function, and added up to yield the empirical density function (PDF) as shown in Fig. 25.14e. The most probable unbinding forces between the anti-*TfsA* and S-layer *TfsA*-GP at a retraction velocity of 1000 nm/s were 41 pN with wt cells and 56 pN with *TfsB*-GP-deficient cells, where only the *TfsA*-GP is exposed on the cell surface. Differences of unbinding forces between wt and *TfsB*-deficient cells may occur due to the orientation and flexibility of the binding site of the *TfsA*-GP on the bacterial cell surface. The *TfsA*-GP on wt bacteria is embedded in the S-layer lattice structure, whereas the *TfsA*-GP does not appear to be that strictly fixed on the surface of the *TfsB*-deficient mutant where the space-filling counterpart (*TfsB*-GP) is missing. Binding was also highly specific, as verified from the binding probability (Fig. 25.14f).

AFM Indentation Experiments to Measure Physical Properties of *Escherichia coli*

Bacteria produce various types of filamentous structures such as flagella, pili, and curli that promote bacterial adhesion onto the surface. In AFM indentation experiments the effects of curli on the physical properties of the bacterial membrane in *Escherichia coli* (*E.coli*) and the two mutant strains curli fibre protein CsgA deficient mutant ($CsgA^-$) and CsgA overexpressed mutant ($CsgA^+$) were studied [25.161]. Figure 25.15a shows

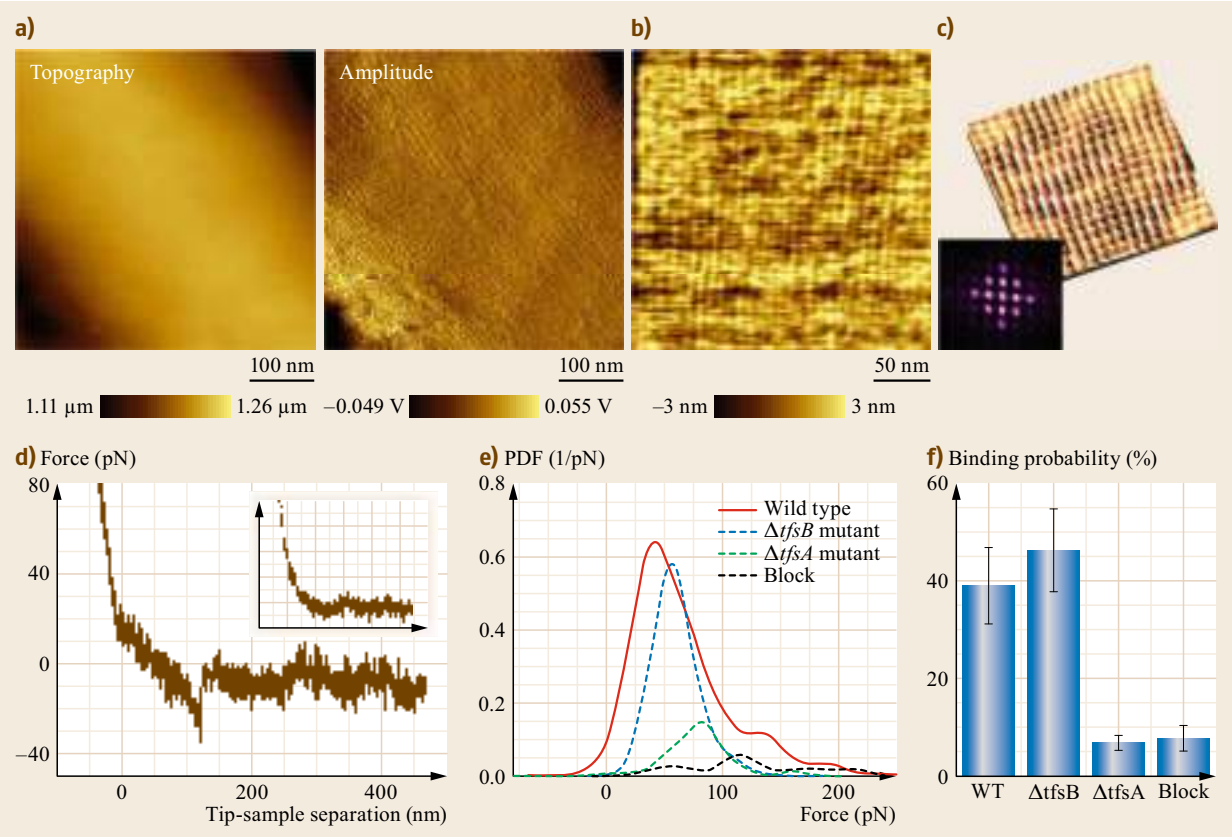


Fig. 25.14a–f Bacterial surface layer (S-layer) of *Tannerella forsythia*. **(a–d)** MAC-mode imaging of *T. forsythia* wild-type bacteria: **(a)** topography and amplitude image, **(b)** magnified topographic image and **(c)** reconverted FFT 3-D image of topography; inset, power spectrum of lattice. **(d–f)** SMFS to identify the specific interactions between anti-*TfsA* antibody and *TfsA* glycoprotein (*TfsA*-GP) in the S-layer lattice. **(d)** Typical unbinding event in a force-distance cycle (inset, blocking of the specific interaction by injecting of recombinant *TfsA* protein into the bath solution), **(e)** distribution of unbinding forces between anti-*TfsA* antibody and the bacterial cell surface (wild type, *TfsA*-deficient mutant (ΔtfsA) and *TfsB*-deficient mutant (ΔtfsB)) at a retraction velocity of 1000 nm/s and **(f)** comparison of binding probabilities of the anti-*TfsA* antibody-conjugated AFM tip to wild type, ΔtfsB and ΔtfsA cells, as well as the blocking experiment

typical approach curves measured on the bacterial cell surfaces that reflect well the different stiffness values of the three strains in relation to the production of curli structures. The measured curves were fitted using the Hertz model for a conical tip to extract the Young's modulus values [25.169–171] (Fig. 25.15b).

The effective spring constant of the cell wall and the decay length of the steric force sensed upon approaching the bacterial surface are a measure for the cell's turgor pressure [25.170, 172] and the thickness of the protruding proteins on the bacterial surface respectively. The late linear parts of the approaching force-indentation curves were used to calculate the bacterial spring constant (effective spring constants), which were fitted using the models of *Alexander* and *de Gennes*, as modified for AFM analysis to quantify the steric interactions (e.g., decay lengths) [25.173–176] (Fig. 25.15c). The data show that the effective spring constant, i. e., the turgor pressure of the bacterial cells, was not affected by bacterial surface structures such as flagella and curli fibers. Since there was no change in bacterial turgor pressure of the mutants compared to that of wt *E.coli*, this parameter most likely reflects the mechanical properties of the cell wall. Consequently, the observed decrease in Young's modulus has been induced by changes in bacterial surface properties. The decay length of the bacterial surface was clearly affected by CsgA curli production. These differences evidently indicate the presence of the CsgA protein on $CsgA^{+}$ overexpressed mutant after curli expression and the presence of flagella-like structure on $CsgA^{-}$ curli deficient mutant respectively, which

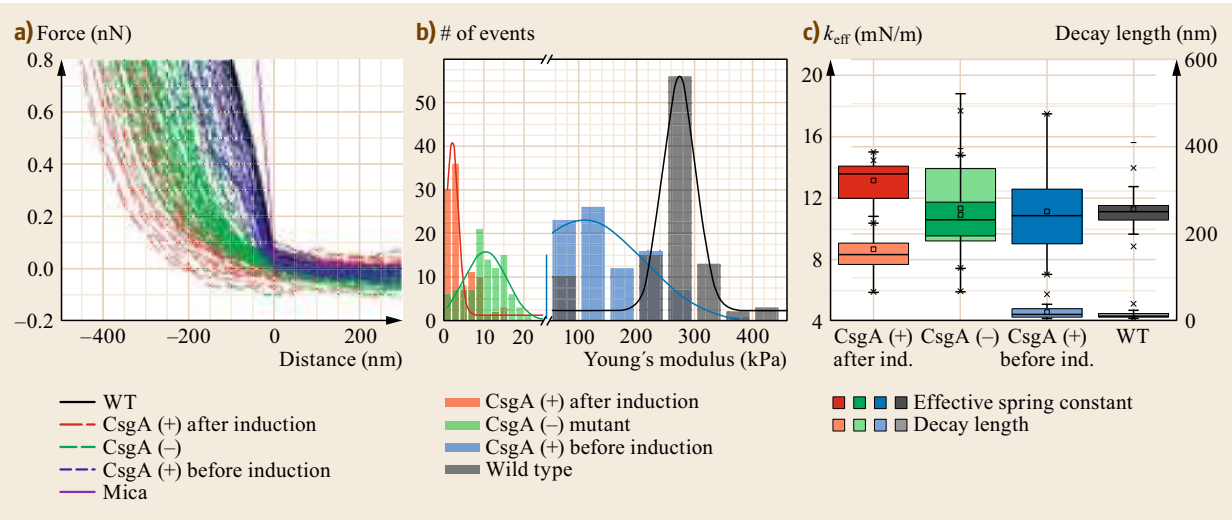


Fig. 25.15a–c AFM indentation measurements to determine the physical properties of *Escherichia coli* bacteria depending on their capacity to produce curli. **(a)** Approach indentation curves measured on a mica surface, wild type, CsgA knock-out mutant ($CsgA^-$) and CsgA overexpressed mutant ($CsgA^+$) before curli expression and after curli expression (by isopropyl-β-D-thiogalactoside (IPTG) induction) on the bacterial surface. **(b)** Distribution of Young's modulus for the different mutant strains. **(c)** Calculated effective spring constant and decay length for the different mutants

both act as a polymer brush on the bacterial cell surface.

Overall the curli overexpressed mutant after IPTG induction produced protruding CsgA protein structures that affected cell wall mechanical properties. This overproduction of curli fiber decreased Young's modulus but increased the decay length on the bacterial surface, while maintaining the inner turgor pressure. This quantitative information obtained by AFM suggests that many critical issues of how mechanical parameters influence cellular behavior can be properly addressed using this method.

25.5 Perspectives and Concluding Remarks

A few years ago live cell AFM experiments still required a high level of expertise and technologically advanced AFM equipment, and consequently were restricted to a few laboratories. Today, several manufacturers offer highly developed bio-AFM setups that are mounted on inverted optical microscopes or provided with upright optical configurations. They are equipped with environmental chambers and sample stages that offer great flexibility and control of experimental conditions. Data processing and analysis also for SMFS is being simplified more and more, as increasingly adequate software is integrated directly into the AFM instrumentation and specialized software has become available from different companies. All these advancements greatly facilitate operation of an AFM by scientist from various life science disciplines. Nevertheless, a basic understanding of the techniques and critical assessment of recorded data remain crucial to AFM experiments. Proper sample preparation and control of environmental conditions, surface chemistry, and the quality of the AFM probe, are equally important and may change over the course of an experiment (e.g., due to contamination) [25.177]. Live-cell experiments in buffer or medium, for instance, have become a routine task and temperature control is available to most setups. However, such experiments continue to require a great deal of patience and newcomers to AFM will generally need to practice before obtaining good data.

SMFS experiments are straightforward, but careful sample and AFM tip preparation together with critical data evaluation are of high importance [25.178, 179]. While standardized tip-functionalization protocols have become more and more available [25.180], validation of successful tip functionalization, i. e., that the ligand is attached and oriented correctly for binding usually happens during the actual experiment. As the interaction force is calculated from the bending of the cantilever, the spring constant has to be calibrated properly

for each cantilever and may vary tremendously from the reported manufacturer value [25.181]. Further considerations in SMFS include the separation of specific (wanted) interactions from nonspecific (unwanted) interactions. Their distinction can be facilitated through the use of defined PEG-crosslinkers as described further above, but may remain challenging even under well-controlled conditions. Unequivocal demonstration of binding specificity by blocking experiments remains absolutely necessary. Finally, data interpretation and automation of force-spectroscopy analyses continue to be time-consuming tasks that (partially) require in-house programmed algorithms. However, many AFM manufacturers have started to implement analysis algorithms into their data-processing software and continue to develop automated routines that will contribute to the simplification of SMFS data evaluation.

Apart from its capabilities, AFM remains primarily a surface technique. Ligand receptor interactions on rigid surfaces and on cell membranes were acquired *blindly* in the past, meaning that force curves were collected at individual, randomly selected sites, with no corresponding image of receptor distribution. The combination of AFM instrumentation with optical microscopy has eliminated this shortcoming and improved imaging and force spectroscopy applications on cells, with the possibility of correlating independently or simultaneously acquired topographic and fluorescence images. The continued development of combined AFM and optical techniques like confocal laser scanning microscopy (CLSM) [25.182–184] and high-resolution correlative microscopy methods such as total internal reflection fluorescence (TIRF) [25.185, 186] or stimulated emission depletion (STED) and stochastic optical reconstruction microscopy (STORM) [25.187] will further improve live-cell applications and allow sophisticated experimental settings. In addition, remarkable advances have been made in developing AFM instruments with increased data acquisition rates, giving access to unprecedented timescales (millisecond resolution) [25.188]. Time resolution is a crucial factor that currently limits investigation of rapid and highly dynamic biological processes with conventional AFM instrumentation. Acquisition of *high*-resolution images takes several minutes, whereas dynamic processes in biology usually occur at much shorter timescales. The ongoing improvements in the field of high-speed AFM open up fascinating new perspectives on the exploration of cellular dynamics and monitoring of molecule interactions with receptors [25.8–10, 189–191]. In the future, high-speed AFM force-distance-based imaging modes will become more and more important in this field. These techniques will allow direct observations of individual molecules in action at high spatiotemporal resolution and provide simultaneous nanoscale mapping of their properties.

To summarize this chapter, we have provided a snapshot of AFM techniques with a special focus on force spectroscopy and recognition imaging for biomedical sensing applications. Our declared aim was to introduce researchers of life sciences in particular, who may be new to these techniques, to the wealth of experimental possibilities provided by the AFM toolbox. The highlighted examples in the last section illustrate the considerable potential of the force-measuring tool SMFS and the recognition-imaging technique TREC to decipher receptor and cell responses to external stimuli and to explore the interaction of cell receptors with individual (drug) molecules. Studies on artificial and native membranes, tissue samples and live cells hold great promise for increasing our understanding of medically relevant molecular processes in disease patterns or pathogen invasion and may help in the development of new therapeutic approaches and pharmaceuticals.

Acknowledgments. Authors and projects presented in this chapter received in part funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 under Marie Skłodowska-Curie grant No. 656842 (to CL), the Austrian Research Promotion Agency (FFG) under MNT-ERA-net project No. 823980 (to AE), project No. 850681 (to LAC), the Austrian Science Foundation (FWF) under project(s) SFB-F35 (to PH), P25357 (to PH), I 767-B11 (to PH) and I 2083-B21 (to PH), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under FOR1543 (to PH), a fellowship from the Austrian Programme for Advanced Research and Technology (APART) of the Austrian Academy of Science and a Marie Curie International Incoming Fellowship of the 7th European Community Framework Program (both to YJO).

References

25.1 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56**(9), 930–933 (1986)

25.2 D.J. Muller, H. Janovjak, T. Lehto, L. Kuerschner, K. Anderson: Observing structure, function and assembly of single proteins by AFM, Prog. Biophys. Mol. Biol. **79**(1/3), 1–43 (2002)

25.3 J.K.H. Horber, M.J. Miles: Scanning probe evolution in biology, Science **302**(5647), 1002–1005

(2003)
25.4 S. Scheuring, D. Levy, J.L. Rigaud: Watching the components of photosynthetic bacterial membranes and their in situ organisation by atomic force microscopy, Biochim. Et Biophys. Acta-Biomembr. **1712**(2), 109–127 (2005)
25.5 A. Jeremic, M. Kelly, S.J. Cho, M.H. Stromer, B.P. Jena: Reconstituted fusion pore, Biophys. J. **85**(3), 2035–2043 (2003)
25.6 M.S. Zaman, A. Goyal, G.P. Dubey, P.K. Gupta, H. Chandra, T.K. Das, M. Ganguli, Y. Singh: Imaging and analysis of Bacillus anthracis spore germination, Microsc. Res. Tech. **66**(6), 307–311 (2005)
25.7 C.M. Franz, D.J. Muller: Analyzing focal adhesion structure by atomic force microscopy, J. Cell Sci. **118**(22), 5315–5323 (2005)
25.8 F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, S. Scheuring: High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations, Science **342**(6159), 741–743 (2013)
25.9 T. Ando, T. Uchihashi, S. Scheuring: Filming biomolecular processes by high-speed atomic force microscopy, Chem. Rev. **114**(6), 3120–3188 (2014)
25.10 J. Preiner, N. Kodera, J. Tang, A. Ebner, M. Brameshuber, D. Blaas, N. Gelbmann, H.J. Gruber, T. Ando, P. Hinterdorfer: IgGs are made for walking on bacterial and viral surfaces, Nat. Commun. **5**, 4394 (2014)
25.11 J. Preiner, A. Horner, A. Karner, N. Ollinger, C. Siligan, P. Pohl, P. Hinterdorfer: High-speed AFM images of thermal motion provide stiffness map of interfacial membrane protein moieties, Nano Lett. **15**(1), 759–763 (2015)
25.12 P. Hinterdorfer, W. Baumgartner, H.J. Gruber, K. Schilcher, H. Schindler: Detection and localization of individual antibody-antigen recognition events by atomic force microscopy, Proc. Natl. Acad. Sci. **93**(8), 3477–3481 (1996)
25.13 J.R. Withers, D.E. Aston: Nanomechanical measurements with AFM in the elastic limit, Adv. Colloid Interface Sci. **120**(1–3), 57–67 (2006)
25.14 S. Kasas, G. Dietler: Probing nanomechanical properties from biomolecules to living cells, Pflügers Archiv-Eur. J. Phys. **456**(1), 13–27 (2008)
25.15 N.E. Kurland, Z. Drira, V.K. Yadavalli: Measurement of nanomechanical properties of biomolecules using atomic force microscopy, Micron **43**(2/3), 116–128 (2012)
25.16 G.U. Lee, D.A. Kidwell, R.J. Colton: Sensing discrete streptavidin-biotin interactions with atomic force microscopy, Langmuir **10**(2), 354–357 (1994)
25.17 E.L. Florin, V.T. Moy, H.E. Gaub: Adhesion forces between individual ligand-receptor pairs, Science **264**(5157), 415–417 (1994)
25.18 T. Hugel, M. Seitz: The study of molecular interactions by AFM force spectroscopy, Macromol. Rapid Commun. **22**(13), 989–1016 (2001)
25.19 A. Alessandrini, P. Facci: AFM: A versatile tool in biophysics, Meas. Sci. Technol. **16**(6), R65–R92 (2005)
25.20 F.A. Carvalho, N.C. Santos: Atomic force microscopy-based force spectroscopy u biological and biomedical applications, Iubmb Life **64**(6), 465–472 (2012)
25.21 F. Kienberger, A. Ebner, H.J. Gruber, P. Hinterdorfer: Molecular recognition imaging and force spectroscopy of single biomolecules, Acc. Chem. Res. **39**(1), 29–36 (2006)
25.22 I. Casuso, F. Rico, S. Scheuring: Biological AFM: Where we come from–where we are–where we may go, J. Mol. Recogn. **24**(3), 406–413 (2011)
25.23 L.S. Dorobantu, M.R. Gray: Application of atomic force microscopy in bacterial research, Scanning **32**(2), 74–96 (2010)
25.24 D.J. Mueller, Y.F. Dufrene: Atomic force microscopy: A nanoscopic window on the cell surface, Trends Cell Biol **21**(8), 461–469 (2011)
25.25 F. Pillet, L. Chopinet, C. Formosa, E. Dague: Atomic force microscopy and pharmacology: From microbiology to cancerology, Biochim. Biophys. Acta Gen. Sub. **1840**(3), 1028–1050 (2014)
25.26 C. Lamprecht, P. Hinterdorfer, A. Ebner: Applications of biosensing atomic force microscopy in monitoring drug and nanoparticle delivery, Expert Opin. Drug. Deliv. **11**(8), 1237–1253 (2014)
25.27 A. Raab, W. Han, D. Badt, S.J. Smith-Gill, S.M. Lindsay, H. Schindler, P. Hinterdorfer: Antibody recognition imaging by force microscopy, Nat. Biotechnol. **17**(9), 901–905 (1999)
25.28 C.M. Stroh, A. Ebner, M. Geretschläger, G. Freudenthaler, F. Kienberger, A.S.M. Kamruzzahan, S.J. Smith-Gill, H.J. Gruber, P. Hinterdorfer: Simultaneous topography and recognition imaging using force microscopy, Biophys. J. **87**(3), 1981–1990 (2004)
25.29 M. Rangl, R. Nevo, I. Liashkovich, V. Shahin, Z. Reich, A. Ebner, P. Hinterdorfer: Stable, non-destructive immobilization of native nuclear membranes to micro-structured PDMS for single-molecule force spectroscopy, Chem. Phys. Chem. **10**(9/10), 1553–1558 (2009)
25.30 M. Mahut, M. Leitner, A. Ebner, M. Lämmerhofer, P. Hinterdorfer, W. Lindner: Time-resolved chloroquine-induced relaxation of supercoiled plasmid DNA, Anal. Bioanal. Chem. **402**(1), 373–380 (2012)
25.31 Y.J. Oh, Y. Cui, H. Kim, Y. Li, P. Hinterdorfer, S. Park: Characterization of curli a production on living bacterial surfaces by scanning probe microscopy, Biophys. J. **103**(8), 1666–1671 (2012)
25.32 S. Posch, I. Neundlinger, M. Leitner, P. Siostrzonek, S. Panzer, P. Hinterdorfer, A. Ebner: Activation induced morphological changes and integrin αIIbβ3 activity of living platelets, Methods **60**(2), 179–185 (2013)
25.33 P. Hansma, J. Cleveland, M. Radmacher, D. Walters, P. Hillner, M. Bezanilla, M. Fritz, D. Vie, H. Hansma, C. Prater: Tapping mode atomic force microscopy in liquids, Appl. Phys. Lett. **64**(13), 1738–1740 (1994)
25.34 R. Gracía: *Amplitude Modulation Atomic Force Microscopy* (Wiley, Weinheim 2010)

25.35 R. Garcia, E.T. Herruzo: The emergence of multifrequency force microscopy, Nat. Nanotechnol. **7**(4), 217–226 (2012)

25.36 J. Preiner, J. Tang, V. Pastushenko, P. Hinterdorfer: Higher harmonic atomic force microscopy: Imaging of biological membranes in liquid, Phys. Rev. Lett. **99**(4), 046102 (2007)

25.37 A. Cartagena, M. Hernando-Pérez, J.L. Carrascosa, P.J. de Pablo, A. Raman: Mapping in vitro local material properties of intact and disrupted virions at high resolution using multi-harmonic atomic force microscopy, Nanoscale **5**(11), 4729–4736 (2013)

25.38 A. Dulebo, J. Preiner, F. Kienberger, G. Kada, C. Rankl, L. Chtcheglova, C. Lamprecht, D. Kaftan, P. Hinterdorfer: Second harmonic atomic force microscopy imaging of live and fixed mammalian cells, Ultramicroscopy **109**(8), 1056–1060 (2009)

25.39 K. El Kirat, I. Burton, V. Dupres, Y. Dufrene: Sample preparation procedures for biological atomic force microscopy, J. Microsc. **218**(3), 199–207 (2005)

25.40 R. Schlapak, J. Danzberger, D. Armitage, D. Morgan, A. Ebner, P. Hinterdorfer, P. Pollheimer, H.J. Gruber, F. Schäffler, S. Howorka: Painting with biomolecules at the nanoscale: biofunctionalization with tunable surface densities, Nano Lett. **12**(4), 1983–1989 (2012)

25.41 P. Pollheimer, B. Taskinen, A. Scherfler, S. Gusenkov, M. Creus, P. Wiesauer, D. Zauner, W. Schöfberger, C. Schwarzinger, A. Ebner, R. Tampé, H. Stutz, V. Hytonen, H.J. Gruber: Reversible biofunctionalization of surfaces with a switchable mutant of avidin, Bioconjug. Chem. **24**(10), 1656–1668 (2013)

25.42 A. Ulman: *An Introduction to Ultrathin Organic Films: From Langmuir–Blodgett to Self-Assembly* (Academic Press, London 2013)

25.43 A. Ebner, L. Wildling, R. Zhu, C. Rankl, T. Haselgrubler, P. Hinterdorfer, H.J. Gruber: Functionalization of probe tips and supports for single-molecule recognition force microscopy, Top. Curr. Chem. **285**, 29–76 (2008)

25.44 M.J. Doktycz, C.J. Sullivan, P.R. Hoyt, D.A. Pelletier, S. Wu, D.P. Allison: AFM imaging of bacteria in liquid media immobilized on gelatin coated mica surfaces, Ultramicroscopy **97**(1-4), 209–216 (2003)

25.45 Y.F. Dufrene: Atomic force microscopy, a powerful tool in microbiology, J. Bacteriol. **184**(19), 5205–5213 (2002)

25.46 M. Raposo, Q. Ferreira, P. Ribeiro: A guide for atomic force microscopy analysis of soft-condensed matter, Modern Res, Edu. Topics Microsc. **1**, 758–769 (2007)

25.47 P. Bongrand, C. Capo, J. Mege, A. Benoliel: Use of hydrodynamic flows to study cell adhesion. In: *Physical Basis of Cell–Cell Adhesion*, ed. by P. Bongrand (CRC, Boca Raton 1988) pp. 125–156

25.48 J.N. Israelachvili: *Intermolecular and Surface Forces*, 3rd edn. (Academic Press, New York 2011)

25.49 J. Friedrichs, K.R. Legate, R. Schubert, M. Bharadwaj, C. Werner, D.J. Müller, M. Benoit: A practical guide to quantify cell adhesion using single-cell force spectroscopy, Methods **60**(2), 169–178 (2013)

25.50 W. Ott, M.A. Jobst, C. Schoeler, H.E. Gaub, M.A. Nash: Single-molecule force spectroscopy on polyproteins and receptor-ligand complexes: The current toolbox, J. Struct. Biol. **197**(1), 3–12 (2016)

25.51 A.R. Bizzarri, S. Cannistraro: The application of atomic force spectroscopy to the study of biological complexes undergoing a biorecognition process, Chem. Soc. Rev. **39**(2), 734–749 (2010)

25.52 M. Ritzefeld, V. Walhorn, D. Anselmetti, N. Sewald: Analysis of DNA interactions using single-molecule force spectroscopy, Amino Acids **44**(6), 1457–1475 (2013)

25.53 P.E. Marszalek, Y.F. Dufrêne: Stretching single polysaccharides and proteins using atomic force microscopy, Chem. Soc. Rev. **41**(9), 3523–3534 (2012)

25.54 A. Engel, H.E. Gaub: Structure and mechanics of membrane proteins, Annu. Rev. Biochem. **77**, 127–148 (2008)

25.55 A.M. Whited, P.S.-H. Park: Atomic force microscopy: A multifaceted tool to study membrane proteins and their interactions with ligands, Biochim. Biophys. Acta (BBA) Biomembr. **1838**(1), 56–68 (2014)

25.56 F. Kienberger, V.P. Pastushenko, G. Kada, H.J. Gruber, C. Riener, H. Schindler, P. Hinterdorfer: Static and dynamical properties of single poly (ethylene glycol) molecules investigated by force spectroscopy, Single Mol **1**(2), 123–128 (2000)

25.57 P. Hinterdorfer, F. Kienberger, A. Raab, H.J. Gruber, W. Baumgartner, G. Kada, C. Riener, S. Wielert-Badt, C. Borken, H. Schindler: Poly (ethylene glycol): An ideal spacer for molecular recognition force microscopy/spectroscopy, Single Mol **1**(2), 99–103 (2000)

25.58 H. Grubmüller, B. Heymann, P. Tavan: Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force, Science **271**(5251), 997–999 (1996)

25.59 G.I. Bell: Models for the specific adhesion of cells to cells, Science **200**(4342), 618–627 (1978)

25.60 E. Evans, K. Ritchie: Dynamic strength of molecular adhesion bonds, Biophys. J. **72**(4), 1541–1555 (1997)

25.61 S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten: Molecular dynamics study of unbinding of the avidin-biotin complex, Biophys. J. **72**(4), 1568–1581 (1997)

25.62 W. Baumgartner, P. Hinterdorfer, W. Ness, A. Raab, D. Vestweber, H. Schindler, D. Drenckhahn: Cadherin interaction probed by atomic force microscopy, Proc. Natl. Acad. Sci. **97**(8), 4005–4010 (2000)

25.63 J. Fritz, A.G. Katopodis, F. Kolbinger, D. Anselmetti: Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy, Proc. Natl. Acad. Sci. **95**(21), 12283–12288 (1998)

25.64 T. Strunz, K. Oroszlan, I. Schumakovitch, H. Guntherodt, M. Hegner: Model energy landscapes and the force-induced dissociation of ligand–

receptor bonds, Biophys. J. **79**(3), 1206–1212 (2000)
25.65 E. Evans, K. Ritchie: Strength of a weak bond connecting flexible polymer chains, Biophys. J. **76**(5), 2439–2447 (1999)
25.66 R. Merkel, P. Nassoy, A. Leung, K. Ritchie, E. Evans: Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy, Nature **397**(6714), 50–53 (1999)
25.67 M. Schlierf, M. Rief: Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape, Biophys. J. **90**(4), L33–5 (2006)
25.68 G. Neuert, C. Albrecht, E. Pamir, H.E. Gaub: Dynamic force spectroscopy of the digoxigenin-antibody complex, FEBS Lett **580**(2), 505–509 (2006)
25.69 R.W. Friddle, A. Noy, J.J. De Yoreo: Interpreting the widespread nonlinear force spectra of intermolecular bonds, Proc. Natl. Acad. Sci. **109**(34), 13573–13578 (2012)
25.70 O.K. Dudko, G. Hummer, A. Szabo: Intrinsic rates and activation free energies from single-molecule pulling experiments, Phys. Rev. Lett. **96**(10), 108101 (2006)
25.71 O.K. Dudko, G. Hummer, A. Szabo: Theory, analysis, and interpretation of single-molecule force spectroscopy experiments, Proc. Natl. Acad. Sci. **105**(41), 15755–15760 (2008)
25.72 D. Alsteens, M. Pfreundschuh, C. Zhang, P.M. Spoerri, S.R. Coughlin, B.K. Kobilka, D.J. Müller: Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape, Nature Methods **12**(9), 845–851 (2015)
25.73 S. Sekatskii, F. Benedetti, G. Dietler: Dependence of the most probable and average bond rupture force on the force loading rate: First order correction to the Bell–Evans model, J. Appl. Phys. **114**(3), 034701 (2013)
25.74 E.B. Walton, S. Lee, K.J. Van Vliet: Extending Bell's model: How force transducer stiffness alters measured unbinding forces and kinetics of molecular complexes, Biophys. J. **94**(7), 2621–2630 (2008)
25.75 A. Noy, R.W. Friddle: Practical single molecule force spectroscopy: How to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope, Methods **60**(2), 142–150 (2013)
25.76 F. Hane, S. Attwood, Z. Leonenko: Comparison of three competing dynamic force spectroscopy models to study binding forces of amyloid-β (1–42), Soft Matter **10**(12), 1924–1930 (2014)
25.77 T. Auletta, M.R. de Jong, A. Mulder, F.C.J.M. van Veggel, J. Huskens, D.N. Reinhoudt, S. Zou, S. Zapotoczny, H. Schönherr, G.J. Vancso, L. Kuipers: Beta-cyclodextrin host-guest complexes probed under thermodynamic equilibrium: Thermodynamics and AFM force spectroscopy, J. Am. Chem. Soc. **126**(5), 1577–1584 (2004)
25.78 A. Chen, V.T. Moy: Cross-linking of cell surface receptors enhances cooperativity of molecular adhesion, Biophys. J. **78**(6), 2814–2820 (2000)
25.79 A. Ebner, P. Hinterdorfer, H.J. Gruber: Comparison of different aminofunctionalization strategies for attachment of single antibodies to AFM cantilevers, Ultramicroscopy **107**(10/11), 922–927 (2007)
25.80 C.K. Riener, C.M. Stroh, A. Ebner, C. Klampfl, A.A. Gall, C. Romanin, Y.L. Lyubchenko, P. Hinterdorfer, H.J. Gruber: Simple test system for single molecule recognition force microscopy, Anal. Chim. Acta. **479**(1), 59–75 (2003)
25.81 R. Ros, F. Schwesinger, D. Anselmetti, M. Kubon, R. Schafer, A. Pluckthun, L. Tiefenauer: Antigen binding forces of individually addressed single-chain Fv antibody molecules, Proc. Natl. Acad. Sci. **95**(13), 7402–7405 (1998)
25.82 A. Ebner, L. Wildling, A.S.M. Kamruzzahan, C. Rankl, J. Wruss, C.D. Hahn, M. Holzl, R. Zhu, F. Kienberger, D. Blaas, P. Hinterdorfer, H.J. Gruber: A new, simple method for linking of antibodies to atomic force microscopy tips, Bioconjugate Chem **18**(4), 1176–1184 (2007)
25.83 L. Wildling, A. Ebner, A.S.M. Kamruzzahan, C.D. Hahn, F. Kienberger, G. Kada, P. Hinterdorfer, H.J. Gruber: Linking of underivatized antibody molecules to AFM tips via long, distensible poly(ethylene glycol) chains, Biophys. J. **88**(1), 154A–155A (2005)
25.84 A.S.M. Kamruzzahan, A. Ebner, L. Wildling, F. Kienberger, C.K. Riener, C.D. Hahn, P.D. Pollheimer, P. Winklehner, M. Holzl, B. Lackner, D.M. Schorkl, P. Hinterdorfer, H.J. Gruber: Antibody linking to atomic force microscope tips via disulfide bond formation, Bioconjugate Chem **17**(6), 1473–1481 (2006)
25.85 J. Tang, A. Ebner, B. Kraxberger, M. Leitner, A. Hykollari, C. Kepplinger, C. Grunwald, H.J. Gruber, R. Tampe, U.B. Sleytr, N. Ilk, P. Hinterdorfer: Detection of metal binding sites on functional S-layer nanoarrays using single molecule force spectroscopy, J. Struct. Biol. **168**(1), 217–222 (2009)
25.86 L. Wildling, B. Unterauer, R. Zhu, A. Rupprecht, T. Haselgrübler, C. Rankl, A. Ebner, D. Vater, P. Pollheimer, E.E. Pohl: Linking of sensor molecules with amino groups to amino-functionalized AFM tips, Bioconjugate Chem **22**(6), 1239–1248 (2011)
25.87 R. Barattin, N. Voyer: Chemical modifications of atomic force microscopy tips. In: *Atomic Force Microscopy in Biomedical Research*, ed. by P.C. Braga, D. Ricci (Humana, New York 2011) pp. 457–483
25.88 L. Lin, H. Wang, Y. Liu, H. Yan, S. Lindsay: Recognition imaging with a DNA aptamer, Biophys. J. **90**(11), 4236–4238 (2006)
25.89 M. Leitner, N. Mitchell, M. Kastner, R. Schlapak, H.J. Gruber, P. Hinterdorfer, S. Howorka, A. Ebner: Single-molecule AFM characterization of individual chemically tagged DNA tetrahedra, ACS Nano **5**(9), 7048–7054 (2011)
25.90 M. Ludwig, W. Dettmann, H. Gaub: Atomic force microscope imaging contrast based on molecular recognition, Biophys. J. **72**(1), 445 (1997)
25.91 O.H. Willemsen, M.M. Snel, K.O. Van Der Werf, B.G. De Grooth, J. Greve, P. Hinterdorfer, H.J. Gruber, H. Schindler, Y. Van Kooyk, C.G. Figdor: Simultaneous height and adhesion imaging of an-

tibody–antigen interactions by atomic force microscopy, Biophys. J. **75**(5), 2220–2228 (1998)
25.92 M. Grandbois, W. Dettmann, M. Benoit, H.E. Gaub: Affinity imaging of red blood cells using an atomic force microscope, J. Histochem. Cytochem. **48**(5), 719–724 (2000)
25.93 P. Lehenkari, G. Charras, A. Nykänen, M. Horton: Adapting atomic force microscopy for cell biology, Ultramicroscopy **82**(1), 289–295 (2000)
25.94 N. Almqvist, R. Bhatia, G. Primbs, N. Desai, S. Banerjee, R. Lal: Elasticity and adhesion force mapping reveals real-time clustering of growth factor receptors and associated changes in local cellular rheological properties, Biophys. J. **86**(3), 1753–1762 (2004)
25.95 I.H. Kim, M.N. Lee, S.H. Ryu, J.W. Park: Nanoscale mapping and affinity constant measurement of signal-transducing proteins by atomic force microscopy, Anal. Chem. **83**(5), 1500–1503 (2011)
25.96 D.H. Kim, J.-E. Lee, Z.-Y. Xu, K.R. Geem, Y. Kwon, J.W. Park, I. Hwang: Cytosolic targeting factor AKR2A captures chloroplast outer membrane-localized client proteins at the ribosome during translation, Nature Commun **6**, 6843 (2015)
25.97 C. Stroh, H. Wang, R. Bash, B. Ashcroft, J. Nelson, H. Gruber, D. Lohr, S. Lindsay, P. Hinterdorfer: Single-molecule recognition imaging microscopy, Proc. Natl. Acad. Sci. USA **101**(34), 12503–12507 (2004)
25.98 J. Adamcik, A. Berquand, R. Mezzenga: Single-step direct measurement of amyloid fibrils stiffness by peak force quantitative nanomechanical atomic force microscopy, Appl. Phys. Lett. **98**(19), 3701 (2011)
25.99 F. Rico, C. Su, S. Scheuring: Mechanical mapping of single membrane proteins at submolecular resolution, Nano Lett **11**(9), 3983–3986 (2011)
25.100 D. Alsteens, V. Dupres, S. Yunus, J.-P. Latgé, J.R.J. Heinisch, Y.F. Dufrêne: High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy, Langmuir **28**(49), 16738–16744 (2012)
25.101 C. Heu, A. Berquand, C. Elie-Caille, L. Nicod: Glyphosate-induced stiffening of HaCaT keratinocytes, a peak force tapping study on living cells, J. Struct. Biol. **178**(1), 1–7 (2012)
25.102 A. Ebner, F. Kienberger, G. Kada, C.M. Stroh, M. Geretschläger, A. Kamruzzahan, L. Wildling, W.T. Johnson, B. Ashcroft, J. Nelson: Localization of single avidin–biotin interactions using simultaneous topography and molecular recognition imaging, Chem. Phys. Chem. **6**(5), 897–900 (2005)
25.103 J. Preiner, A. Ebner, L. Chtcheglova, R. Zhu, P. Hinterdorfer: Simultaneous topography and recognition imaging: Physical aspects and optimal imaging conditions, Nanotechnology **20**(21), 215103 (2009)
25.104 J. Tang, A. Ebner, H. Badelt-Lichtblau, C. Völlenkle, C. Rankl, B. Kraxberger, M. Leitner, L. Wildling, H.J. Gruber, U.B. Sleytr: Recognition imaging and highly ordered molecular templating of bacterial S-layer nanoarrays containing affinity-tags, Nano Lett. **8**(12), 4312–4319 (2008)
25.105 A. Ebner, D. Nikova, T. Lange, J. Häberle, S. Falk, A. Dübbers, R. Bruns, P. Hinterdorfer, H. Oberleithner, H. Schillers: Determination of CFTR densities in erythrocyte plasma membranes using recognition imaging, Nanotechnology **19**(38), 384017 (2008)
25.106 T. Lange, P. Jungmann, J. Haberle, S. Falk, A. Duebbers, R. Bruns, A. Ebner, P. Hinterdorfer, H. Oberleithner, H. Schillers: Reduced number of CFTR molecules in erythrocyte plasma membrane of cystic fibrosis patients, Mol. Membr. Biol. **23**(4), 317–323 (2006)
25.107 C. Rankl, R. Zhu, G.S. Luengo, M. Donovan, N. Baghdadli, P. Hinterdorfer: Detection of corneodesmosin on the surface of stratum corneum using atomic force microscopy, Exp. Dermatol. **19**(11), 1014–1019 (2010)
25.108 L.A. Chtcheglova, J. Waschke, L. Wildling, D. Drenckhahn, P. Hinterdorfer: Nano-scale dynamic recognition imaging on vascular endothelial cells, Biophys. J. **93**(2), L11–L13 (2007)
25.109 L.A. Chtcheglova, L. Wildling, J. Waschke, D. Drenckhahn, P. Hinterdorfer: AFM functional imaging on vascular endothelial cells, J. Mol. Recogn. **23**(6), 589–596 (2010)
25.110 L. Lin, D. Hom, S.M. Lindsay, J.C. Chaput: In vitro selection of histone H4 aptamers for recognition imaging microscopy, J. Am. Chem. Soc. **129**(47), 14568–14569 (2007)
25.111 L.A. Chtcheglova, F. Atalar, U. Ozbek, L. Wildling, A. Ebner, P. Hinterdorfer: Localization of the ergtoxin-1 receptors on the voltage sensing domain of hERG K+ channel by AFM recognition imaging, Pflügers Archiv Eur. J. Phys. **456**(1), 247–254 (2008)
25.112 H. Wang, Y. Dalal, S. Henikoff, S. Lindsay: Single-epitope recognition imaging of native chromatin, Epigenetics Chromatin **1**(1), 1 (2008)
25.113 M. Koehler, G. Macher, A. Rupprecht, R. Zhu, H.J. Gruber, E.E. Pohl, P. Hinterdorfer: Combined recognition imaging and force spectroscopy: a new mode for mapping and studying interaction sites at low lateral density, Sci. Adv. Mater. **9**(1), 128–134 (2017)
25.114 S.F. Ahmad, L.A. Chtcheglova, B. Mayer, S.A. Kuznetsov, P. Hinterdorfer: Nanosensing of Fc gamma receptors on macrophages, Anal. Bioanal. Chem. **399**(7), 2359–2367 (2011)
25.115 R. Creasey, S. Sharma, J.E. Craig, C.T. Gibson, A. Ebner, P. Hinterdorfer, N.H. Voelcker: Detecting protein aggregates on untreated human tissue samples by atomic force microscopy recognition imaging, Biophys. J. **99**(5), 1660–1667 (2010)
25.116 R. Creasey, S. Sharma, C.T. Gibson, J.E. Craig, A. Ebner, T. Becker, P. Hinterdorfer, N.H. Voelcker: Atomic force microscopy-based antibody recognition imaging of proteins in the pathological deposits in pseudoexfoliation syndrome, Ultramicroscopy **111**(8), 1055–1061 (2011)
25.117 B. Wang, C. Guo, M. Zhang, B. Park, B. Xu: High-resolution single-molecule recognition imaging

of the molecular details of ricin-aptamer interaction, J. Phys. Chem. B **116**(17), 5316–5322 (2012)
25.118 M. Zhang, G. Chen, R. Kumar, B. Xu: Mapping out the structural changes of natural and pretreated plant cell wall surfaces by atomic force microscopy single molecular recognition imaging, Biotechnol. Biofuels **6**(1), 1 (2013)
25.119 R. Zhu, A. Rupprecht, A. Ebner, T. Haselgrübler, H.J. Gruber, P. Hinterdorfer, E.E. Pohl: Mapping the nucleotide binding site of uncoupling protein 1 using atomic force microscopy, J. Am. Chem. Soc. **135**(9), 3640–3646 (2013)
25.120 M. Duman, L.A. Chtcheglova, R. Zhu, B.L. Bozna, P. Polzella, V. Cerundolo, P. Hinterdorfer: Nanomapping of CD1d-glycolipid complexes on THP1 cells by using simultaneous topography and recognition imaging, J. Mol. Recogn. **26**(9), 408–414 (2013)
25.121 J. Zhang, L.A. Chtcheglova, R. Zhu, P. Hinterdorfer, B. Zhang, J. Tang: Nanoscale organization of human GnRH-R on human bladder cancer cells, Anal. Chem. **86**(5), 2458–2464 (2014)
25.122 B. Nimmervoll, L.A. Chtcheglova, K. Juhasz, N. Cremades, F.A. Aprile, A. Sonnleitner, P. Hinterdorfer, L. Vigh, J. Preiner, Z. Balogi: Cell surface localised Hsp70 is a cancer specific regulator of clathrin-independent endocytosis, FEBS Lett **589**(19), 2747–2753 (2015)
25.123 S. Posch: Collagen surface functionalization – A new strategy for molecular recognition force spectroscopy, http://www.imaging-git.com/science/scanning-probe-microscopy/collagen-surface-functionalization (2014)
25.124 C. Aponte-Santamaría, V. Huck, S. Posch, A.K. Bronowska, S. Grässle, M.A. Brehm, T. Obser, R. Schneppenheim, P. Hinterdorfer, S.W. Schneider: Force-sensitive autoinhibition of the von Willebrand factor is mediated by interdomain interactions, Biophys. J. **108**(9), 2312–2321 (2015)
25.125 S. Posch, C. Aponte-Santamaría, R. Schwarzl, A. Karner, M. Radtke, F. Gräter, T. Obser, G. König, M.A. Brehm, H.J. Gruber, R.R. Netz, C. Baldauf, R. Schneppenheim, R. Tampé, P. Hinterdorfer: Mutual a domain interactions in the force sensing protein von Willebrand factor, J. Struct. Biol. **197**(1), 57–64 (2017)
25.126 M. Klingenberg, K.S. Echtay: Uncoupling proteins: The issues from a biochemist point of view, Biochimica Biophysica Acta (BBA) Bioenerg **1504**(1), 128–143 (2001)
25.127 J. Nedergaard, D. Ricquier, L.P. Kozak: Uncoupling proteins: Current status and therapeutic prospects, EMBO Reports **6**(10), 917–921 (2005)
25.128 M.E. Lidell, S. Enerbäck: Brown adipose tissue – A new role in humans?, Nat. Rev. Endocrinol. **6**(6), 319–325 (2010)
25.129 D.G. Nicholls, V.S. Bernson, G.M. Heaton: The identification of the component in the inner membrane of brown adipose tissue mitochondria responsible for regulating energy dissipation. In: *Effectors of Thermogenesis*, ed. by L. Girardier, J. Seydoux (Springer, Basel 1978) pp. 89–93
25.130 C. Lin, M. Klingenberg: Isolation of the uncoupling protein from brown adipose tissue mitochondria, FEBS Lett **113**(2), 299–303 (1980)
25.131 D.G. Nicholls: Hamster brown-adipose-tissue mitochondria, Eur. J. Biochem. **62**(2), 223–228 (1976)
25.132 E. Urbánková, A. Voltchenko, P. Pohl, P. Ježek, E.E. Pohl: Transport kinetics of uncoupling proteins analysis of ucp1 reconstituted in planar lipid bilayers, J. Biol. Chem. **278**(35), 32497–32500 (2003)
25.133 V. Beck, M. Jabůrek, T. Demina, A. Rupprecht, R.K. Porter, P. Ježek, E.E. Pohl: Polyunsaturated fatty acids activate human uncoupling proteins 1 and 2 in planar lipid bilayers, FASEB J **21**(4), 1137–1144 (2007)
25.134 M.J. Berardi, W.M. Shih, S.C. Harrison, J.J. Chou: Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching, Nature **476**(7358), 109–113 (2011)
25.135 E. Pebay-Peyroula, C. Dahout-Gonzalez, R. Kahn, V. Trézéguet, G.J.-M. Lauquin, G. Brandolin: Structure of mitochondrial ADP–ATP carrier in complex with carboxyatractyloside, Nature **426**(6962), 39–44 (2003)
25.136 M. Modrianský, D.L. Murdza-Inglis, H.V. Patel, K.B. Freeman, K.D. Garlid: Identification by site-directed mutagenesis of three arginines in uncoupling protein that are essential for nucleotide binding and inhibition, J. Biol. Chem. **272**(40), 24759–24762 (1997)
25.137 A.S. Kristensen, J. Andersen, T.N. Jørgensen, L. Sørensen, J. Eriksen, C.J. Loland, K. Strømgaard, U. Gether: SLC6 neurotransmitter transporters: Structure, function, and regulation, Pharmacol. Rev. **63**(3), 585–640 (2011)
25.138 A. Yamashita, S.K. Singh, T. Kawate, Y. Jin, E. Gouaux: Crystal structure of a bacterial homologue of Na^+/Cl^- Dependent neurotransmitter transporters, Nature **437**(7056), 215–223 (2005)
25.139 A. Penmatsa, K.H. Wang, E. Gouaux: X-ray structure of dopamine transporter elucidates antidepressant mechanism, Nature **503**(7474), 85–90 (2013)
25.140 K.H. Wang, A. Penmatsa, E. Gouaux: Neurotransmitter and psychostimulant recognition by the dopamine transporter, Nature **521**(7552), 322–327 (2015)
25.141 P. Plenge, E.T. Mellerup: An affinity-modulating site on neuronal monoamine transport proteins, Pharmacol. Toxicol. **80**(4), 197–201 (1997)
25.142 P. Plenge, L. Shi, T. Beuming, J. Te, A.H. Newman, H. Weinstein, U. Gether, C.J. Loland: Steric hindrance mutagenesis in the conserved extracellular vestibule impedes allosteric binding of antidepressants to the serotonin transporter, J. Biol. Chem. **287**(47), 39316–39326 (2012)
25.143 R. Zhu, D. Sinwel, P.S. Hasenhuetl, K. Saha, V. Kumar, P. Zhang, C. Rankl, M. Holy, S. Sucic, O. Kudlacek: Nanopharmacological force sensing to reveal allosteric coupling in transporter binding sites, Angew. Chem. Int. **55**, 1719–1722 (2015)

25.144 V. Kumar, T. Rahbek-Clemmensen, C.B. Billesbølle, T.N. Jorgensen, U. Gether, A.H. Newman: Novel and high affinity fluorescent ligands for the serotonin transporter based on (S)-citalopram, ACS Med. Chem. Lett. **5**(6), 696–699 (2014)

25.145 L. Wildling, C. Rankl, T. Haselgrübler, H.J. Gruber, M. Holy, A.H. Newman, M.-F. Zou, R. Zhu, M. Freissmuth, H.H. Sitte: Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells, J. Biol. Chem. **287**(1), 105–113 (2012)

25.146 J. Monod, J.-P. Changeux, F. Jacob: Allosteric proteins and cellular control systems, J. Mol. Biol. **6**(4), 306–329 (1963)

25.147 J.A. Coleman, E.M. Green, E. Gouaux: X-ray structures and mechanism of the human serotonin transporter, Nature **532**(7599), 334–339 (2016)

25.148 G. Pfister, C.M. Stroh, H. Perschinka, M. Kind, M. Knoflach, P. Hinterdorfer, G. Wick: Detection of HSP60 on the membrane surface of stressed human endothelial cells by atomic force and confocal microscopy, J. Cell Sci. **118**(8), 1587–1594 (2005)

25.149 M. Duman, M. Pfleger, R. Zhu, C. Rankl, L.A. Chtcheglova, I. Neundlinger, B.L. Bozna, B. Mayer, M. Salio, D. Shepherd, P. Polzella, M. Moertelmaier, G. Kada, A. Ebner, M. Dieudonne, G.J. Schutz, V. Cerundolo, F. Kienberger, P. Hinterdorfer: Improved localization of cellular membrane receptors using combined fluorescence microscopy and simultaneous topography and recognition imaging, Nanotechnology **21**(11), 115504 (2010)

25.150 S. El-Kirat-Chatel, Y.F. Dufrene: Nanoscale imaging of the candida – Macrophage interaction using correlated fluorescence-atomic force microscopy, ACS Nano **6**(12), 10792–10799 (2012)

25.151 J. Madl, S. Rhode, H. Stangl, H. Stockinger, P. Hinterdorfer, G.J. Schutz, G. Kada: A combined optical and atomic force microscope for live cell investigations, Ultramicroscopy **106**(8/9), 645–651 (2006)

25.152 P. Lanzerstorfer, V. Stadlbauer, L.A. Chtcheglova, R. Haselgruebler, D. Borgmann, J. Wruss, P. Hinterdorfer, K. Schroeder, S.M. Winkler, O. Hoeglinger, J. Weghuber: Identification of novel insulin mimetic drugs by quantitative total internal reflection fluorescence (TIRF) microscopy, Br. J. Pharmacol. **171**(23), 5237–5251 (2014)

25.153 R. Govers: Molecular mechanisms of GLUT4 regulation in adipocytes, Diabetes Metab **40**(6), 400–410 (2014)

25.154 S. Bruzzone, P. Ameri, L. Briatore, E. Mannino, G. Basile, G. Andraghetti, A. Grozio, M. Magnone, L. Guida, S. Scarfi, A. Salis, G. Damonte, L. Sturla, A. Nencioni, D. Fenoglio, F. Fiory, C. Miele, F. Beguinot, V. Ruvolo, M. Bormioli, G. Colombo, D. Maggi, G. Murialdo, R. Cordera, A. De Flora, E. Zocchi: The plant hormone abscisic acid increases in human plasma after hyperglycemia and stimulates glucose consumption by adipocytes and myoblasts, FASEB J **26**(3), 1251–1260 (2012)

25.155 S. Kadan, B. Saad, Y. Sasson, H. Zaid: In vitro evaluations of cytotoxicity of eight antidiabetic medicinal plants and their effect on GLUT4 translocation, Evidence-Based Complement. Alternative Med. (2013) doi:10.1155/2013/549345

25.156 S. Kristiansen, E.A. Richter: GLUT4-containing vesicles are released from membranes by phospholipase D cleavage of a GPI anchor, Am. J. Physiol. Endocrinol. Metabolism **283**(2), E374–E382 (2002)

25.157 E. Carvalho, S.E. Schellhorn, J.M. Zabolotny, S. Martin, E. Tozzo, O.D. Peroni, K.L. Houseknecht, A. Mundt, D.E. James, B.B. Kahn: GLUT4 overexpression or deficiency in adipocytes of transgenic mice alters the composition of GLUT4 vesicles and the subcellular localization of GLUT4 and insulin-responsive aminopeptidase, J. Biol. Chem. **279**(20), 21598–21605 (2004)

25.158 Z.Y. Jiang, A. Chawla, A. Bose, M. Way, M.P. Czech: A phosphatidylinositol 3-kinase-independent insulin signaling pathway to N-WASP/Arp2/3/F-actin required for GLUT4 glucose transporter recycling, J. Biol. Chem. **277**(1), 509–515 (2002)

25.159 D. Baus, K. Heermeier, M. De Hoop, C. Metz-Weidmann, J. Gassenhuber, W. Dittrich, S. Welte, N. Tennagels: Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells, Cell. Signal. **20**(12), 2237–2246 (2008)

25.160 A. Persat, C.D. Nadell, M.K. Kim, F. Ingremeau, A. Siryaporn, K. Drescher, N.S. Wingreen, B.L. Bassler, Z. Gitai, H.A. Stone: The mechanical world of bacteria, Cell **161**(5), 988–997 (2015)

25.161 Y.J. Oh, Y. Cui, H. Kim, Y. Li, P. Hinterdorfer, S. Park: Characterization of curli a production on living bacterial surfaces by scanning probe microscopy, Biophys. J. **103**(8), 1666–1671 (2012)

25.162 Y.F. Dufrêne: Sticky microbes: Forces in microbial cell adhesion, Trends Microbiol **23**(6), 376–382 (2015)

25.163 Y.J. Oh, M. Hubauer-Brenner, H.J. Gruber, Y. Cui, L. Traxler, C. Siligan, S. Park, P. Hinterdorfer: Curli mediate bacterial adhesion to fibronectin via tensile multiple bonds, Sci. Rep. **6**, 33909 (2016)

25.164 J. Tang, A. Ebner, H. Badelt-Lichtblau, C. Völlenkle, C. Rankl, B. Kraxberger, M. Leitner, L. Wildling, H.J. Gruber, U.B. Sleytr, N. Ilk, P. Hinterdorfer: Recognition imaging and highly ordered molecular templating of bacterial S-layer nanoarrays containing affinity-tags, Nano Lett. **8**(12), 4312–4319 (2008)

25.165 M. Chalabi, F. Rezaie, S. Moghim, A. Mogharehabed, M. Rezaei, B. Mehraban: Periodontopathic bacteria and herpesviruses in chronic periodontitis, Mol. Oral Microbiol. **25**(3), 236–240 (2010)

25.166 G. Sekot, G. Posch, P. Messner, M. Matejka, X. Rausch-Fan, O. Andrukhov, C. Schaffer: Potential of the tannerella forsythia S-layer to delay the immune response, J. Dent. Res. **90**(1), 109–114 (2011)

25.167 G. Sekot, G. Posch, Y.J. Oh, S. Zayni, H.F. Mayer, D. Pum, P. Messner, P. Hinterdorfer, C. Schaffer: Analysis of the cell surface layer ultrastructure of the oral pathogen Tannerella forsythia, Arch. Mi-

crobiol. **194**(6), 525–539 (2012)
25.168 Y.J. Oh, G. Sekot, M. Duman, L. Chtcheglova, P. Messner, H. Peterlik, C. Schäffer, P. Hinterdorfer: Characterizing the S-layer structure and anti-S-layer antibody recognition on intact tannerella forsythia cells by scanning probe microscopy and small angle x-ray scattering, J. Mol. Recogn. JMR **26**(11), 542–549 (2013)
25.169 A. Touhami, B. Nysten, Y.F. Dufrêne: Nanoscale mapping of the elasticity of microbial cells by atomic force microscopy, Langmuir **19**(11), 4539–4543 (2003)
25.170 Y.-Y. Chen, C.-C. Wu, J.-L. Hsu, H.-L. Peng, H.-Y. Chang, T.-R. Yew: Surface rigidity change of escherichia coli after filamentous bacteriophage infection, Langmuir **25**(8), 4607–4614 (2009)
25.171 T.G. Kuznetsova, M.N. Starodubtseva, N.I. Yegorenkov, S.A. Chizhik, R.I. Zhdanov: Atomic force microscopy probing of cell elasticity, Micron **38**(8), 824–833 (2007)
25.172 M. Arnoldi, M. Fritz, E. Bauerlein, M. Radmacher, E. Sackmann, A. Boulbitch: Bacterial turgor pressure can be measured by atomic force microscopy, Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics **62**(1 Pt B), 1034–1044 (2000)
25.173 S. Alexander: Adsorption of chain molecules with a polar head a scaling description, J. Phys. France **38**(8), 983–987 (1977)
25.174 T.A. Camesano, B.E. Logan: Probing bacterial electrosteric interactions using atomic force microscopy, Environ. Sci. Technol. **34**(16), 3354–3362 (2000)
25.175 P.G. de Gennes: Polymers at an interface; A simplified view, Adv. Colloid Interface Sci. **27**(3), 189–209 (1987)
25.176 J. Strauss, N.A. Burnham, T.A. Camesano: Atomic force microscopy study of the role of LPS O-antigen on adhesion of E. coli, J. Mol. Recogn. **22**(5), 347–355 (2009)
25.177 K. El Kirat, I. Burton, V. Dupres, Y.F. Dufrene: Sample preparation procedures for biological atomic force microscopy, J. Microsc. **218**(3), 199–207 (2005)
25.178 Y.F. Dufrene, E. Evans, A. Engel, J. Helenius, H.E. Gaub, D.J. Muller: Five challenges to bringing single-molecule force spectroscopy into living cells, Nature Methods **8**(2), 123–127 (2011)
25.179 A. Noy: Force spectroscopy 101: How to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment, Curr. Opin. Chem. Biol. **15**(5), 710–718 (2011)
25.180 A. Ebner, L. Wildling, R. Zhu, C. Rankl, T. Haselgrubler, P. Hinterdorfer, H.J. Gruber: Functionalization of probe tips and supports for single-molecule recognition force microscopy. In: *STM and AFM Studies on (Bio)molecular Systems: Unravelling the Nanoworld*, ed. by P. Samori (Springer, Berlin 2008) pp. 29–76
25.181 J. te Riet, A.J. Katan, C. Rankl, S.W. Stahl, A.M. van Buul, I.Y. Phang, A. Gomez-Casado, P. Schön, J.W. Gerritsen, A. Cambi, A.E. Rowan, G.J. Vancso, P. Jonkheijm, J. Huskens, T.H. Oosterkamp, H. Gaub, P. Hinterdorfer, C.G. Figdor, S. Speller: Interlaboratory round robin on cantilever calibration for AFM force spectroscopy, Ultramicroscopy **111**(12), 1659–1669 (2011)
25.182 K. Poole, D. Meder, K. Simons, D. Muller: The effect of raft lipid depletion on microvilli formation in MDCK cells, visualized by atomic force microscopy, FEBS Lett **565**(1–3), 53–58 (2004)
25.183 R. Kassies, K.O. van der Werf, A. Lenferink, C.N. Hunter, J.D. Olsen, V. Subramaniam, C. Otto: Combined AFM and confocal fluorescence microscope for applications in bio-nanotechnology, J. Microsc. **217**(1), 109–116 (2005)
25.184 S. Doak, D. Rogers, B. Jones, L. Francis, R.S. Conlan, C. Wright: High-resolution imaging using a novel atomic force microscope and confocal laser scanning microscope hybrid instrument: Essential sample preparation aspects, Histochem. Cell Biol. **130**(5), 909–916 (2008)
25.185 A.B. Mathur, G.A. Truskey, W.M. Reichert: Atomic force and total internal reflection fluorescence microscopy for the study of force transmission in endothelial cells, Biophys. J. **78**(4), 1725–1735 (2000)
25.186 A. Trache, S.M. Lim: Integrated microscopy for real-time imaging of mechanotransduction studies in live cells, J. Biomed. Opt. **14**(3), 034024 (2009)
25.187 J.V. Chacko, F.C. Zanacchi, A. Diaspro: Probing cytoskeletal structures by coupling optical superresolution and AFM techniques for a correlative approach, Cytoskeleton **70**(11), 729–740 (2013)
25.188 I. Casuso, F. Rico, S. Scheuring: Biological AFM: Where we come from – Where we are – Where we may go, J. Mol. Recogn. **24**(3), 406–413 (2011)
25.189 M. Shibata, H. Yamashita, T. Uchihashi, H. Kandori, T. Ando: High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin, Nat. Nano **5**(3), 208–212 (2010)
25.190 T. Ando: High-speed atomic force microscopy coming of age, Nanotechnology **23**(6), 062001 (2012)
25.191 A. Karner, B. Nimmervoll, B. Plochberger, E. Klotzsch, A. Horner, D.G. Knyazev, R. Kuttner, K. Winkler, L. Winter, C. Siligan, N. Ollinger, P. Pohl, J. Preiner: Tuning membrane protein mobility by confinement into nanodomains, Nat. Nanotechnol. **12**(3), 260–266 (2017)

26. Superresolution Microscopy

Tom D. Milster

There are many types of optical microscopy systems that produce superresolution. This discussion centers on optical microscopy techniques that have the potential to extract features from objects that are one the scale of 100 nm or less, which is much smaller than what can be achieved with a classical optical microscope.

In order to achieve resolution on the order of 100 nm with visible-light photons, more information must be obtained from the system than a single image using classical illumination can produce. In some cases, the extra information is in the form of a series of images with a customized illumination pattern. In other systems, the object displays response characteristics that effectively reduce the size of the scanning laser spot used to illuminate it. In yet other systems, light is forced through a nanosized aperture and then scanned over the object.

None of the systems described in this chapter actually change the characteristics or physics of the optical systems used to collect photons. Instead, classical optical systems are cleverly combined with advanced illumination techniques and postprocessing that produce superresolution images.

In the introduction, basic concepts regarding classical resolution are reviewed, and terms are defined that are important with respect to understanding how superresolution microscopy works. Subsequent sections describe superresolution techniques, including scanning aperture techniques, 4-Pi microscopy, enhancement/depletion techniques, photoactivated localization, lattice light-sheet microscopy, and structured illumination. A short comparison of techniques for live-cell imaging is also provided. Although examples of the techniques are given, this chapter is not intended to be a state-of-the-art review inclusive of all variations. Instead, the intent is to provide a basic understanding of the primary classes of superresolution microscopy techniques.

It is notable that work in this area has generated several recent Nobel prizes, because of the importance to science of being able to resolve structures and physiology at the nanoscale [26.1].

26.1 **Overview** 845
26.1.1 Review of Classical Resolution 845
26.1.2 Point Spread Function (PSF) and Frequency Response 847
26.2 **Scanning Aperture Techniques** 849
26.3 **4-Pi Microscopy** 853
26.4 **Enhancement/Depletion Techniques** 854
26.5 **Photoactivated Localization** 856
26.6 **Structured Illumination** 858
26.7 **Light-Sheet Microscopy** 862
26.8 **Comparison of Techniques for Live-Cell Imaging** 863
26.9 **Summary** 864
References 864

26.1 Overview

26.1.1 Review of Classical Resolution

It is well known that the physical effects of diffraction limit the image size of a point source object, as shown in Fig. 26.1. When a monochromatic point source is imaged by an optical system with a limiting aperture of radius a in the space of the image, the Airy disk pattern is observed at the image plane with an irradiance

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_26

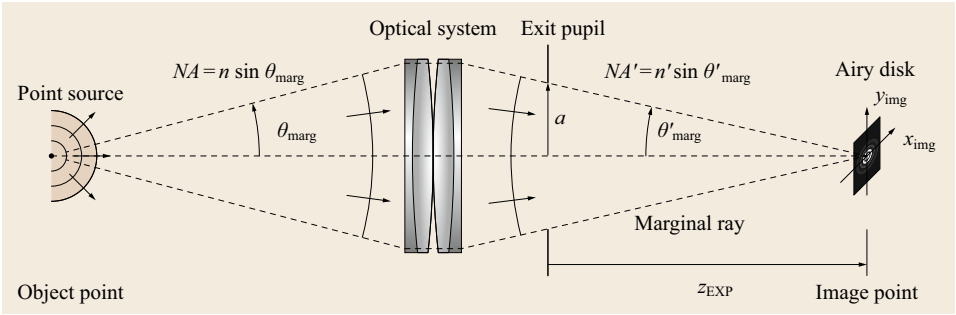


Fig. 26.1 Imaging a point source. The image of an expanding point source in an ideal classical optical system is an Airy disk

distribution

$$I_{img}\left(x_{img}, y_{img}\right) = I_0 \,\mathrm{somb}^2\left(\frac{2an'\rho_{img}}{\lambda z_{EXP}}\right) = I_0 \,\mathrm{somb}^2\left(\frac{2NA'\rho_{img}}{\lambda}\right), \quad (26.1)$$

where $\left(x_{img}, y_{img}\right)$ are coordinates of the image plane with $\rho_{img} = \sqrt{x_{img}^2 + y_{img}^2}$, z_{EXP} is the distance between the limiting aperture and the image plane, I_0 is the peak irradiance, n' is the refractive index of the image space, and λ is the source wavelength. In (26.1), somb(r) is the sombrero function, which is defined by

$$\mathrm{somb}(r) = \frac{2J_1(\pi r)}{\pi r}, \quad (26.2)$$

and J_1 is a first-order Bessel function. The limiting aperture could be in the space of the image, as shown in Fig. 26.1, or it could be internal to the optical/lens system. In either case, the image of the limiting aperture in image space is the *exit pupil*, which defines the size of the Airy disk on the image detector.

It is more common in microscope systems to consider the *numerical apertures* in object and image spaces, rather than the physical aperture diameter a. As shown in Fig. 26.1, the numerical aperture in image space is defined by

$$NA' = n' \sin \theta'_{marg}, \quad (26.3)$$

where θ'_{marg} is the marginal ray angle that passes through the edge of the exit pupil, and n' is the refractive index. A similar definition for numerical aperture is used in object space, where the marginal ray angle and NA are used without primes ($'$). In well-behaved optical imaging systems, NA and NA' are related to each other by the transverse magnification m_T, where

$$NA' = \frac{NA}{m_T}. \quad (26.4)$$

The Airy disk is displayed in Fig. 26.2. Interestingly, *nanometer-size sources of light, like fluorescent molecules, effectively produce the same Airy disk in the image plane as an ideal point source.* A Gaussian approximation for the Airy disk profile is also shown with

$$I_{Gaus}\left(\rho_{img}\right) = I_0 \,\mathrm{gaus}\left(\frac{2NA'\rho_{img}}{\lambda}\right) = I_0 \exp\left[-\pi\left(\frac{2NA'\rho_{img}}{\lambda}\right)^2\right]. \quad (26.5)$$

The Gaussian approximation is important, because it simplifies development of superresolution algorithms.

Now consider two points in object space that are incoherent point emitters, like those generated from fluorophore molecules in a biological experiment. The emitters themselves are much smaller than the wavelength of light emitted from them, and there is no correlation in time between the wavefronts generated by the two emitters. Figure 26.3 shows how the two points are imaged onto a camera. Irradiances from the two Airy disks combine to produce the total image. The effect is a blurring between image points. If the object points are too close together, it is difficult to tell from the image distribution if there is one point or two. At the *Rayleigh criterion*, image separation is such that the first zero of one Airy disk overlaps at the peak of the second Airy disk. At the Rayleigh separation shown in Fig. 26.3, a small dip is present between the two peaks. Mathematically, the Rayleigh criterion specifies that two points are resolved if their separation is

$$\rho_{RES} = 0.61\frac{\lambda}{NA'} = 0.61\, m_T \frac{\lambda}{NA}. \quad (26.6)$$

When imaging an extended object with a classical optical system, it is difficult to discern features separated in the object by distances less than ρ_{RES}/m_T. For example, an ideal $NA = 1.2$ water immersion microscope using $\lambda = 400\,\mathrm{nm}$ light produces a Rayleigh resolution

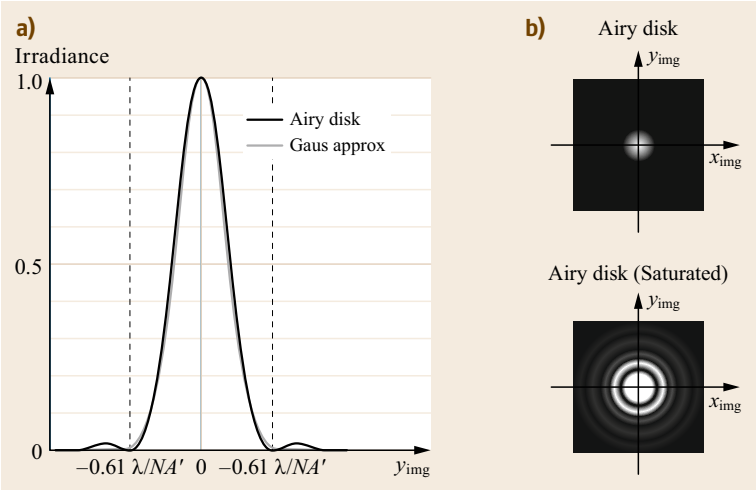


Fig. 26.2 **(a)** Focused laser spot profiles, including the Airy disk and a Gaussian approximation to the Airy disk. **(b)** Airy disk as it appears in the image plane with a linear irradiance scale, and Airy disk with 50X saturation, where the plot is increased in scale by a factor of 50. The saturated Airy disk image highlights the rings present in the Airy disk, which are not present with the Gaussian approximation

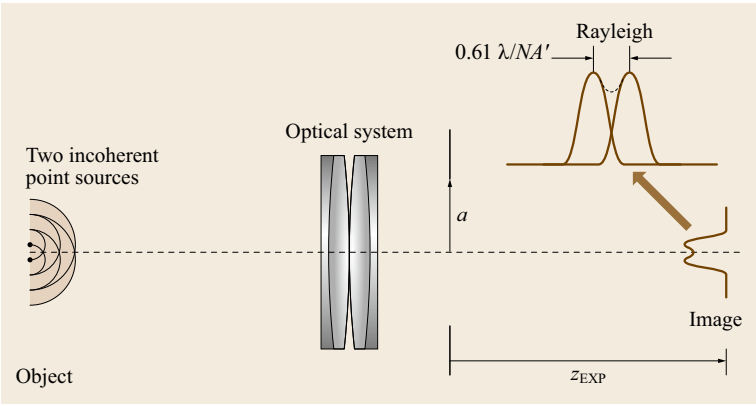


Fig. 26.3 Rayleigh resolution. Two points are said to be just resolved according to the Rayleigh criterion if the separation of their Airy disks in the image plane is such that the peak of one Airy disk falls on the zero of the adjacent disk, as shown in the inset figure. At this separation, the combined irradiance of the two Airy disks has a slight dip in the center

of 203 nm, and the width of the Airy disk in object space is about 407 nm. *The goal of superresolution microscopy is to improve the imaging-system resolving capability to orders of magnitude smaller than the classical Rayleigh criterion.*

26.1.2 Point Spread Function (PSF) and Frequency Response

If the object structure under investigation is given by the function $o\left(x_{\text{obj}}, y_{\text{obj}}\right)$, uniform incoherent illumination produces an irradiance distribution at the object given by

$$I_{\text{obj}}\left(x_{\text{obj}}, y_{\text{obj}}\right) = I_{\text{ill}} o\left(x_{\text{obj}}, y_{\text{obj}}\right) \ , \tag{26.7}$$

where I_{ill} is a constant irradiance value in W/m^2. Under these conditions, the object irradiance is considered as a continuous collection of incoherent point emitters, where the object irradiance function I_{obj} and image irradiance function I_{img} are related by

$$I_{\text{img}}\left(x_{\text{img}}, y_{\text{img}}\right) = CI_{\text{obj}}\left(\frac{x_{\text{img}}}{m_{\text{T}}}, \frac{y_{\text{img}}}{m_{\text{T}}}\right) **h_{\text{incoh}}\left(x_{\text{img}}, y_{\text{img}}\right) \ . \tag{26.8}$$

The *incoherent point spread function* (PSF) of the system is $h_{\text{incoh}}\left(x_{\text{img}}, y_{\text{img}}\right)$, where $(**)$ represents convolution of the two functions, C is a constant, and it is assumed that the optical system is linear and shift invariant. Equation 26.8 also represents the relationship between object and image if the object is self-luminous, like a fluorescent biological sample, where $o\left(x_{\text{obj}}, y_{\text{obj}}\right)$ is the object function that describes the distribution of fluorophores throughout the object. In optical systems with a round exit pupil, like most microscopes, the Airy disk in (26.1) divided by I_0 is the incoherent point spread function h_{incoh}.

Application of the properties of Fourier transforms to (26.8) allows the *frequency response* of the system to

be written as [26.2]

$$\mathcal{I}\left(\xi',\eta'\right)=\mathcal{O}\left(\xi',\eta'\right)\mathcal{H}_{\mathrm{OTF}}\left(\xi',\eta'\right)\ , \tag{26.9}$$

where (ξ',η') are spatial frequency variables that correspond to the $\left(x_{\mathrm{img}},y_{\mathrm{img}}\right)$ directions, $\mathcal{I}\left(\xi',\eta'\right)$ is the Fourier transform of I_{img}, $\mathcal{O}\left(\xi',\eta'\right)$ is the Fourier transform of I_{obj} scaled with the transverse magnification to the image plane, and $\mathcal{H}_{\mathrm{OTF}}\left(\xi',\eta'\right)$ is the Fourier transform of h_{incoh}. $\mathcal{H}_{\mathrm{OTF}}\left(\xi',\eta'\right)$ is called the *optical transfer function* (OTF), because it relates the spatial frequency content of the image to the spatial frequency content of the object. An ideal classical optical system with a round exit pupil and an Airy disk point spread function exhibits an OTF like that shown in Fig. 26.4.

Notice that improving resolution defined by (26.6) is obtained by increasing NA, which also increases the maximum incoherent spatial frequency that can be passed through the optical system. The classical limit for the highest incoherent spatial frequency that can be passed through the system is called the *incoherent cutoff frequency*, which is given by $\xi'_{\mathrm{incoh}}=2NA'/\lambda$. In terms of object spatial frequencies, the incoherent cutoff frequency is $\xi_{\mathrm{incoh}}=2NA/\lambda$.

Several superresolution microscopy techniques implement illumination conditions that transmit higher spatial frequencies through the optical system than what can be achieved with uniform incoherent illumination. An appropriate linear systems mathematical model for irradiance generated at the object is

$$I_{\mathrm{obj}}\left(x_{\mathrm{obj}},y_{\mathrm{obj}}\right)= I_{\mathrm{ill}}\left(x_{\mathrm{obj}}-x_{\mathrm{src}},y_{\mathrm{obj}}-y_{\mathrm{src}}\right)o\left(x_{\mathrm{obj}},y_{\mathrm{obj}}\right)\ , \tag{26.10}$$

where I_{ill} is now a function of position and is shifted by the scan values $(x_{\mathrm{src}},y_{\mathrm{src}})$. Information is detected from multiple scan values to reconstruct an image with high resolution.

In terms of spatial frequencies at the objects, the Fourier transform of (26.10) is $\mathcal{O}\left(\xi,\eta\right)$, which is expressed by

$$\mathcal{O}\left(\xi,\eta\right)=S\left(\xi,\eta;x_{\mathrm{src}},y_{\mathrm{src}}\right)**\tilde{o}\left(\xi,\eta\right)\ , \tag{26.11}$$

where $S\left(\xi',\eta;x_{\mathrm{src}},y_{\mathrm{src}}\right)$ is the frequency spectrum of the illumination and is the Fourier transform of $I_{\mathrm{ill}}\left(x_{\mathrm{obj}}-x_{\mathrm{src}},y_{\mathrm{obj}}-y_{\mathrm{src}}\right)$. $\tilde{o}\left(\xi,\eta\right)$ is the frequency spectrum of the object and is the Fourier transform of $o\left(x_{\mathrm{obj}},y_{\mathrm{obj}}\right)$. Ideally, all object spatial frequencies would be recovered by the imaging system in order to completely reconstruct $o\left(x_{\mathrm{obj}},y_{\mathrm{obj}}\right)$.

For uniform incoherent illumination described by (26.7), I_{ill} is a constant and $S\left(\xi,\eta;x_{\mathrm{src}},y_{\mathrm{src}}\right)=\delta\left(\xi,\eta\right)$, where $\delta\left(\xi,\eta\right)$ is a delta function. Evaluation of the convolution in (26.10) yields

$$\mathcal{O}\left(\xi,\eta\right)=\tilde{o}\left(\xi,\eta\right)\ . \tag{26.12}$$

The spatial frequency content transmitted to the image plane is governed by the OTF through (26.9) with the scaling relationships $\delta'=\xi/m_{\mathrm{T}}$ and $\eta'=\eta/m_{\mathrm{T}}$. For example, in ideal $NA=1.2$ water immersion microscope using $\lambda=400\,\mathrm{nm}$ would not be able to transmit periodic object structures with periods smaller than $\lambda/2NA=167\,\mathrm{nm}$, which is slightly smaller than the Rayleigh resolution of 203 nm.

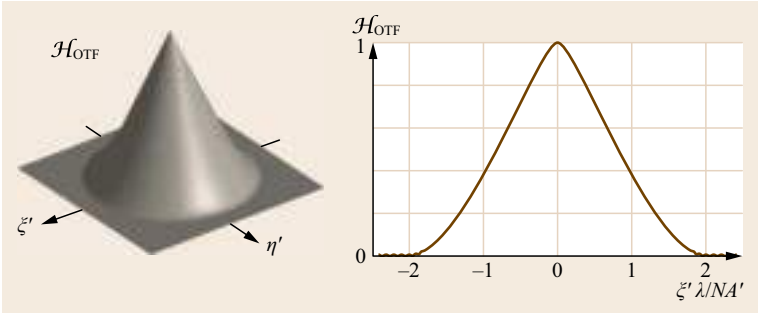


Fig. 26.4 The optical transfer function (OTF) relates the incoherent spatial frequency content of the object to the spatial frequency content of the image. In an ideal classical optical system with a round exit pupil, the OTF is the Fourier transform of the Airy disk

26.2 Scanning Aperture Techniques

There are several superresolution microscopy techniques that rely on limiting the illumination function I_{ill} to a nanoscopic size and scanning the illumination over the sample to collect information for reconstructing the object. One way to reduce the size of the illumination function is to force light through a small hole [26.3]. In practice, most aperture-type scanning instruments use an optical fiber that has been drawn to a small diameter and coated with a metal like aluminum on all but a very small diameter of the tip, as shown in Fig. 26.5 [26.4]. This type of scanning microscopy is called scanning near-field optical microscopy (SNOM) or near-field scanning optical microscopy (NSOM). Its main attraction is the ability to probe optical characteristics of a sample on the nanoscale.

Ideally, the illumination function is a point, where

$$I_{\mathrm{ill}}\left(x_{\mathrm{obj}}-x_{\mathrm{src}}, y_{\mathrm{obj}}-y_{\mathrm{src}}\right) = \delta\left(x_{\mathrm{obj}}-x_{\mathrm{src}}, y_{\mathrm{obj}}-y_{\mathrm{src}}\right) . \tag{26.13}$$

Consider a grating-like object simplified to one dimension with

$$o\left(x_{\mathrm{obj}}\right) = \frac{1}{2}+\frac{1}{2}\cos\left(\frac{2\pi x_{\mathrm{obj}}}{T}\right) , \tag{26.14}$$

where T is the grating period. Application of (26.10) and (26.11) results in

$$\mathcal{O}\left(\frac{1}{T}\right) = \frac{1}{2}+\frac{1}{2}\cos\left(\frac{2\pi x_{\mathrm{src}}}{T}\right) , \tag{26.15}$$

which is only a function of the scan coordinate x_{src} and the grating period T. *The OTF does not affect the signal*, other than to increase the number of photons for improved signal-to-noise ratio, because the data signal is only a function of x_{src}. In fact, a simple photomultiplier tube placed close to the object is often used for data collection, rather than a more complicated optical system. Also, there is no conceptual lower limit on the grating period T, which can be much smaller that the Rayleigh resolution from (26.6). Of course, the time required to obtain enough data to reconstruct an image depends on the scanning speed of the illumination function.

In practice, nanoscopic scanning light sources are of finite width. Due to the skin depth and other electromagnetic properties of the metal cladding, the geometric form of the illumination function is sometimes approximated by a Gaussian distribution. In terms of the one-dimensional grating example of (26.14), a Gaussian illumination function of width d_{ap}, where $d_{\mathrm{ap}} \ll \lambda/NA$, results in an object frequency spectrum

$$\begin{aligned}\mathcal{O}(\xi) \propto\ & \mathrm{gaus}\left(d_{\mathrm{ap}}\xi\right) \\ & + \mathrm{gaus}\left[d_{\mathrm{ap}}\left(\xi-\frac{1}{T}\right)\right]\exp\left(\frac{\mathrm{i}2\pi x_{\mathrm{src}}}{T}\right) \\ & + \mathrm{gaus}\left[d_{\mathrm{ap}}\left(\xi+\frac{1}{T}\right)\right]\exp\left(-\frac{\mathrm{i}2\pi x_{\mathrm{src}}}{T}\right) ,\end{aligned} \tag{26.16}$$

which is shown conceptually in Fig. 26.6 with fre-

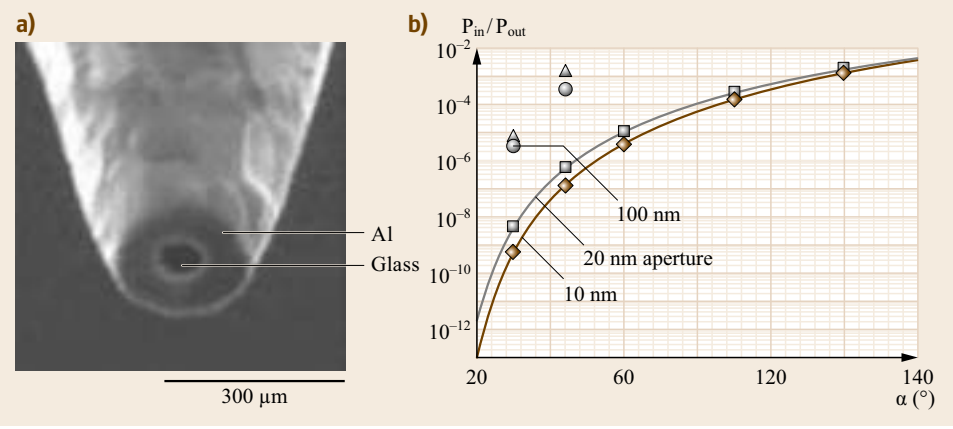


Fig. 26.5a,b SNOM aperture tip fabricated on the end of an optical fiber. **(a)** Construction of the tip involves drawing an optical fiber and partially coating the end with a metal, like aluminum. The diameter of the glass opening roughly corresponds to d_{ap}. Scale bar is 300 nm. **(b)** The power throughput of a SNOM tip is limited, because the aperture is much smaller than the wavelength of light. This plot shows the calculated ratio of input to output power for 10 and 20 nm diameter apertures, along with scaling of these results to a 100 nm aperture. The horizontal axis is the taper angle of the tip. Reprinted from [26.4] with the permission of AIP Publishing

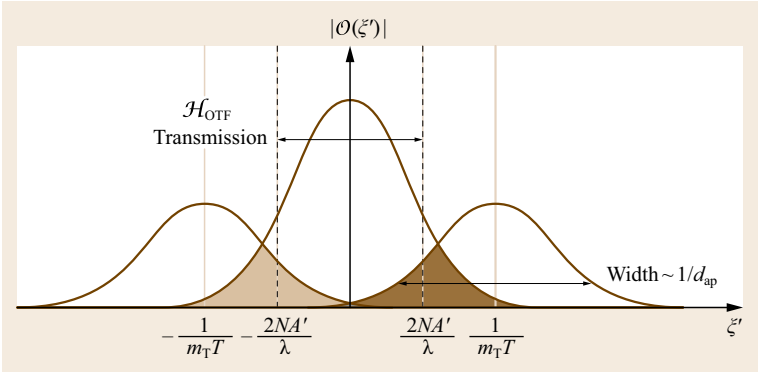


Fig. 26.6 SNOM signal generation. For a cosine object and a small Gaussian aperture of width d_{ap}, object frequencies distribution consist of three Gaussian distributions, as described by (26.16). Two Gaussians distributions are offset from the center by $\pm 1/m_{\mathrm{T}}T$. As the aperture is scanned, modulation between the central Gaussian distribution and the shifted Gaussian distributions occurs in the regions of overlap due to the exponential phase factors. Modulation below the incoherent cutoff frequency $2NA'/\lambda$ is transmitted to the detector as the scanning signal

quency variables scaled by $1/m_{\mathrm{T}}$ for image space. In (26.16), $\mathrm{i} = \sqrt{-1}$, and the second and third terms contain complex phase exponentials that depend on the scan coordinate of the illumination function. The three Gaussian distributions in (26.16) are separated in the frequency domain according to $1/T$. As T decreases, the separation becomes larger. Gaussian distribution widths are proportional to $1/d_{\mathrm{ap}}$, so small illumination functions produce wide Gaussian distributions. As the illumination function is scanned in x_{src}, interference between the exponential phase terms in (26.16) and the central Gaussian distribution produces modulation in the region below the incoherent cutoff frequency $\pm 2\lambda/NA$. That is, the *shaded regions* in Fig. 26.6 become brighter or darker as the aperture is scanned over the object. The integrated signal transmitted through $\mathcal{H}_{\mathrm{OTF}}$ contains this modulation, and the image is reconstructed from the accumulated signal. The smallest period that can be resolved depends on the width of the Gaussian illumination function. Smaller d_{ap} produces finer superresolution, because the amount of overlap increases for large $1/\mathrm{T}$ as d_{ap} becomes smaller.

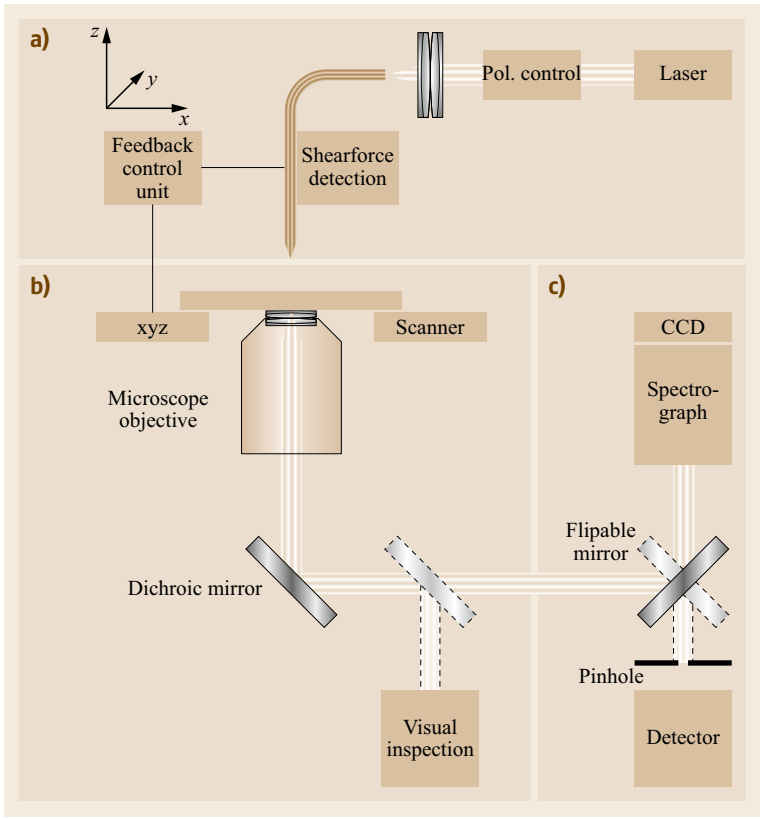


Fig. 26.7a–c A typical SNOM system (after [26.4]) **(a)** illumination unit; **(b)** collection and redistribution unit; and **(c)** detection module

One complication of the small tip diameter is that the transmission of light through holes with diameters less than the wavelength of light is very low. As shown in Fig. 26.5b, a 100 nm diameter tip with a 30° cone angle only emits one photon for every 10^6 photons incident upon the tip through the fiber [26.4]. In practice, d_{ap} is limited to a minimum of 80–100 nm, because of the poor signal-to-noise ratio and long scan times required for smaller apertures. Unfortunately, increasing the incident laser power level to improve signal-to-noise ratio has an upper limit, due to ablation of the metal tip where most of the incident photons are absorbed, so systems using d_{ap} smaller than about 80 nm are not practical.

Another complication with scanning optical probe techniques is that the illumination function needs to be in near contact with the object. Techniques similar to those used for atomic force microscopy are used to successfully regulate SNOM tip position above the sample. A typical arrangement for a SNOM device is shown in Figure 26.7 [26.4]. The illumination unit (a) directs the laser beam through polarization control into the proximal end of the fiber. The fiber tip is placed in proximity to the sample, and shear-force detection is used

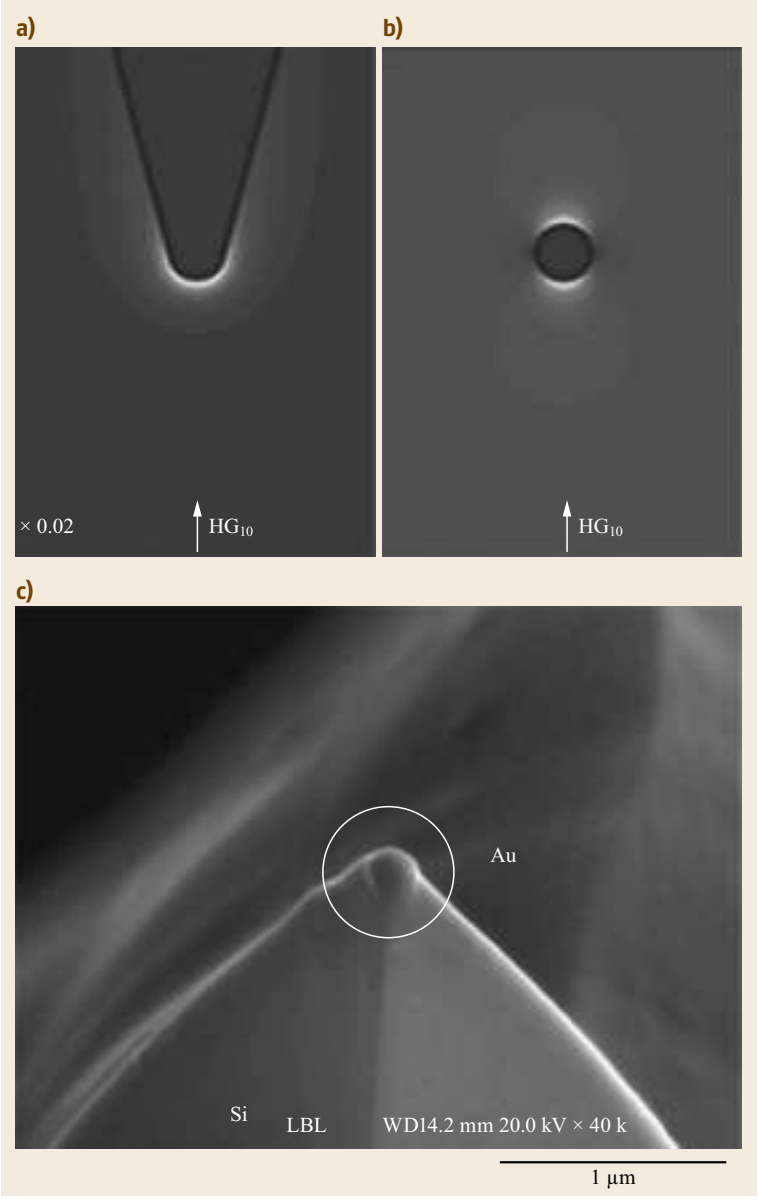


Fig. 26.8a–c Aperturless scanning near-field optical microscopy (ASNOM), where lighter shading indicates higher field strength; **(a)** field strength of a tapered probe; **(b)** field length of an isolated gold nanosphere excited by a vertically polarized laser beam as indicated by the *small arrow*; **(c)** fabrication of an ASNOM tip using a 50 nm gold nanosphere attached to a tapered silicon probe. **(a,b)** from [26.5], copyright Lukas Novotny and Bert Hecht 2012, published by Cambridge University Press. **(c)** reprinted with permission from [26.6], copyright 2006 American Chemical Society

to regulate the tip-to-sample distance through a feedback control unit [26.7]. Collection is in transmission through the object with a standard microscope objective lens in the collection and redistribution unit (b), which also contains the mechanical scanner for the sample stage. The detection module (c) can contain spectrographic and photon-counting detectors. Other modes of operation include using the tip in *collection* mode, where the tip is scanned over a light field and photons are collected through the fiber onto a detector, and *illumination/collection* mode, where the tip both illuminates the object and collects photons back through it.

An alternative to launching light into the fiber and forcing photons through the aperture is to illuminate a sharpened tip externally with a laser beam, as shown in Fig. 26.8. In Fig. 26.8a, excitation of a sharpened gold tip with an end radius of 10 nm is compared to excitation of a gold nanosphere [26.5]. The excitation is polarized in the vertical direction, and the gray scale indicates field strength. Clearly, fields are well confined near the tip. Maximum field strength near the tapered tip is 50 times the strength of the isolated gold nanosphere, because the tapered tip acts like an antenna for the electric field. The use of the externally activated tip to scan a surface and collect image information is called *apertureless scanning near-field optical microscopy* (ASNOM). The primary advantage to ASNOM over SNOM is that field confinement depends on the tip radius, which can produce optical resolution in the range of 30 nm. In Fig. 26.8b, a 50 nm gold nanosphere is attached to a tapered silicon probe for ASNOM studies [26.6].

Examples of SNOM studies are shown in Fig. 26.9 that illustrate the diversity of applications [26.8].

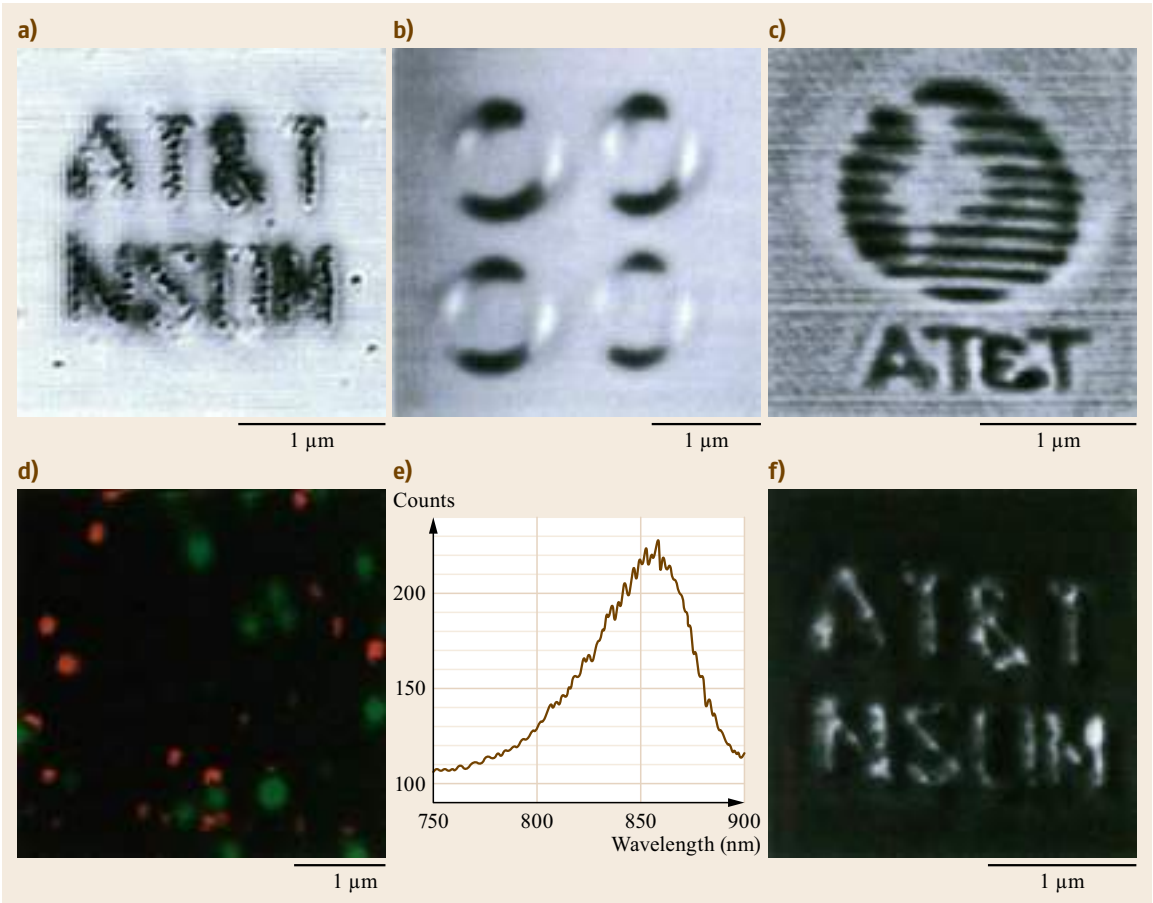


Fig. 26.9a–f NSOM contrast mechanisms (after [26.8]): **(a)** absorption; **(b)** polarization; **(c)** refractive index; **(d)** two-color fluorescence; **(e)** spectroscopy; **(f)** reflectivity

26.3 4-Pi Microscopy

4-Pi microscopy is based on the discovery that the three-dimensional PSF of a confocal microscope is improved by interference of two coherent wavefronts at the focal plane of the microscope [26.9]. As shown in Fig. 26.10, two microscope objectives (OL1 and OL2) illuminate a sample from both sides with coherent light at a common focal plane in the sample. Light from OL1 and OL2 interferes in the focus volume and produces a standing wave illumination pattern. Fluorescent emission is received by a detector that includes a pinhole conjugate to the focal plane through both Paths 1 and 2. (The additional relay optics are not shown in Fig. 26.10.) The axial extent of the effective confocal point-spread function is significantly reduced, which is equivalent to an improved axial resolution.

The effective PSF is given by the product of the illumination PSF and the PSF of the pinhole detector, where

$$h_{4\text{-Pi}} = h_{\text{ill}} h_{\text{det}} \,. \tag{26.17}$$

The function $h_{4\text{-Pi}}$ is proportional to the probability of an illumination photon both exciting a fluorophore in the vicinity of focus and detecting a photon emitted from the same position. Depending on the choice of pinhole size, $h_{4\text{-Pi}}$ produces better resolution in the transverse plane when compared to standard illumination and Rayleigh resolution, and the axial resolution is significantly improved.

4-Pi confocal fluorescence microscopy is used primarily in two different configurations [26.10]. The Type A configuration utilizes interference of the laser in the focal volume and only collects fluorescence from Path 1. The Type C configuration also utilizes interference of the laser from Paths 1 and 2, but it collects fluorescence from both paths at the detector pinhole. When arranged properly, the combination of fluorescence from both paths also creates an interference signal at the pinhole. Due to the combination of the interference in excitation and detection, the Type C configuration exhibits lower side lobes in the axial direction

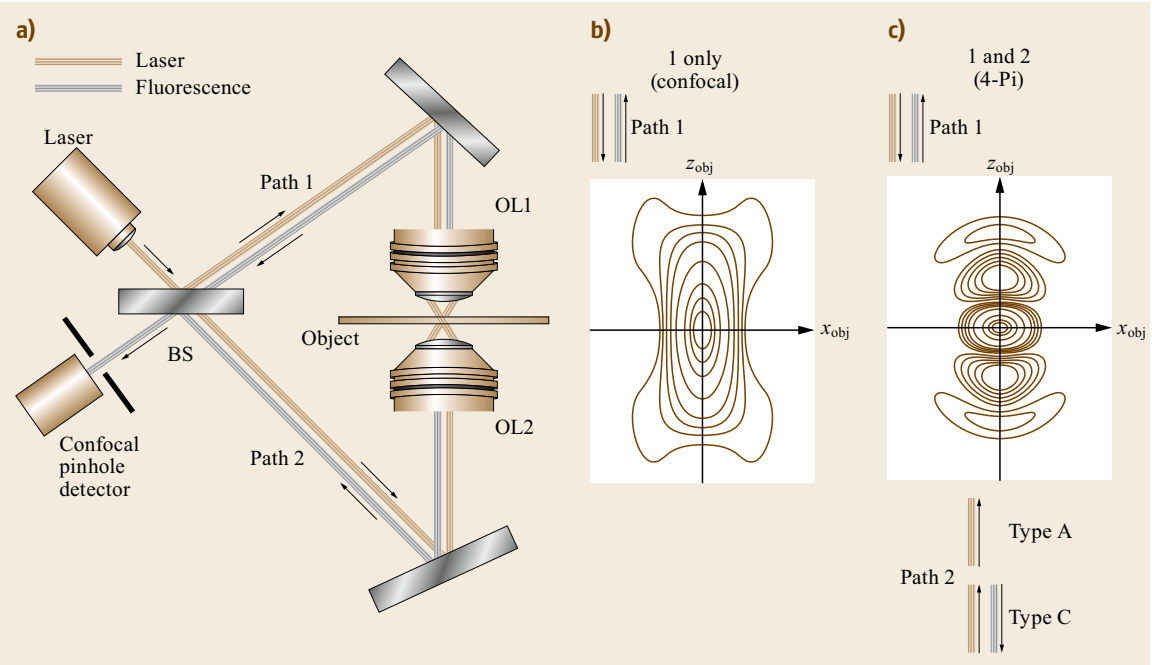


Fig. 26.10a,b,c 4-Pi Microscopy. **(a)** A confocal microscope is modified to use two counter-opposed illumination objective lenses (OL1 and OL2). A laser beam is directed into each lens, and the combination of the focused spots creates a standing wave pattern that dramatically reduces the axial extent of the illumination volume. Effective PSFs from [26.9] are shown in **(b)** for a confocal arrangement (*Path 1* only) and in **(c)** for the 4-Pi arrangement using *Path 1* and *Path 2*. Type A 4-Pi microscopy uses only the fluorescence signal from *Path 1*, while Type C 4-Pi microscopy uses the fluorescence signal from both paths. Interference fringes are clearly visible in the effective PSF for the 4-Pi geometry

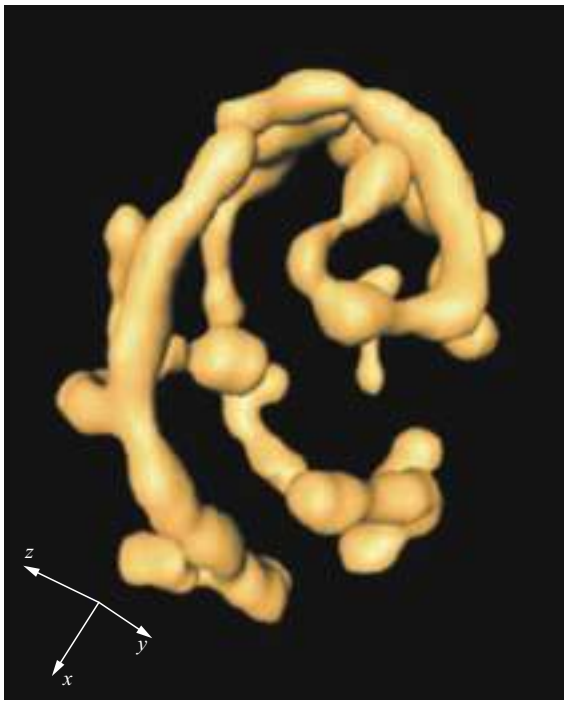


Fig. 26.11 3-D image of the mitochondrial compartments of a yeast cell taken with a Type C 4-Pi microscope. The extent of the cell is ≈ 5 μm (after [26.10]) ◀

of the effective PSF and increased axial resolution compared to Type A.

Figure 26.11 displays a three-dimensional (3-D) image of a living yeast cell labeled with green fluorescent proteins that targeted the mitochondrial compartments in the cell [26.10]. A series of scanned image planes was obtained with a Type C 4-Pi confocal microscope, and then postprocessing was applied to render the three-dimensional distribution of fluorescence above a threshold value.

Two-photon 4-Pi fluorescence microscopy can improve axial resolution to 100 nm, because the energy in the first higher and lower axial lobes is insufficient to produce two-photon fluorescence [26.11]. Both axial and lateral resolution can be additionally improved to as little as 75 nm using image restoration techniques [26.12].

26.4 Enhancement/Depletion Techniques

The concept behind enhancement/depletion techniques is very similar to the idea of SNOM/NSOM, except the size of the illumination function is reduced photochemically in the sample, rather than with a physical aperture. Optically, information is transmitted through the microscope using an effectively nanoscopic illumination function, as shown in Fig. 26.6. Stimulated emission depletion fluorescence microscopy (STED) uses two lasers focused to the same location in the sample, where the focused spot from one laser excites fluorescence, and the second laser inhibits spontaneous fluorescence, as shown in Fig. 26.12 [26.13, 14]. Clever engineering of the spot shapes allows superresolution to be achieved. For two-dimensional (2-D) STED, the first laser pulse (shown in green in Fig. 26.12) generates an unaltered Airy or Gaussian focus distribution, which excites fluorophore molecules in the focus area. After a short time delay Δt, the second laser pulse (shown in red) generates a donut-shaped focus distribution due to the 0 to 2π spiral phase optical filter. The second laser beam depletes the carriers through stimulated emission or transition to a nonradiative mode before excited carriers decay with fluorescence. The remaining distribution of fluorescent molecules are contained in an area that is effectively the subtraction of the two laser focus distributions, which is much smaller than the Airy disk. This effective PSF contains fluorescent molecules that emit at wavelength λ_{em}, which is longer that the excitation laser wavelength. By increasing the power of the STED laser beam, the effective PSF becomes smaller. Lateral resolutions of 25 nm and below have been demonstrated [26.15, 16].

Applied to three dimensions, STED creates a 3-D PSF that localizes fluorescence information to the nanoscale in x, y and z dimensions, as shown in Fig. 26.12 [26.17]. In this case, the optical phase filter consists of a small, circular π phase retarder in the illumination path of the STED laser, which produces zero optical field at the center of the 3-D distribution. A measured 3-D STED PSF is shown in the lower-right portion of Fig. 26.12. Experimental localization is nearly 100 nm in all three dimensions and is significantly smaller than the effective PSF of the same microscope in a confocal configuration.

STED is a type of reversible saturable optical fluorescence transition (RESLOFT) microscopy, where resolution is given by

$$\rho_{\mathrm{RESLOFT}} \approx \frac{\lambda_{\mathrm{exc}}}{2NA\sqrt{1+\zeta}}\,, \tag{26.18}$$

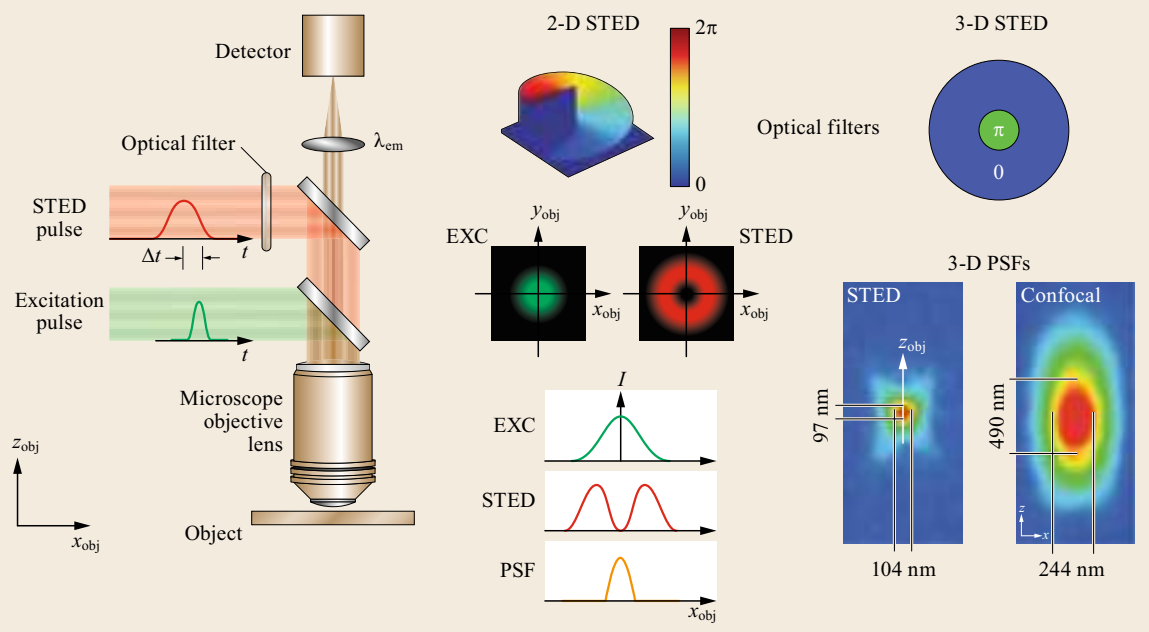


Fig. 26.12 Stimulated emission depletion superresolution microscopy. Two laser beams of different colors are passed through a microscope objective lens and focused onto the sample. The 2-D STED laser beam is modified with an optical filter to produce a donut-shaped focus irradiance. Due to properties of the fluorescent molecules, the excitation laser beam excites carriers that are depleted with the STED beam before they fluoresce in regions where the STED beam irradiance is high enough. The net result is a PSF that is the subtraction of the excitation (EXC) beam and the STED beam, which is considerably smaller than either the EXC or STED profiles. If the optical filter is changed to a simple circular π phase delay, the STED focus in 3-D contains a zero in the center and can deplete fluorescence except in the very center of the volume. Lower-right effective PSF distributions derived from [26.17]

where λ_{exc} is the excitation wavelength and ζ is the saturation factor given by

$$\zeta = \frac{I_{max}}{I_{sat}} , \qquad (26.19)$$

I_{sat} is the saturation power per unit area of the fluorophore, and I_{max} is the maximum value of the irradiance in the STED laser spot [26.18]. Equation (26.18) is smaller than the Rayleigh resolution in (26.6) by a factor of $\sqrt{1+\zeta}$. As STED-beam laser power I_{max} increases compared to the saturation irradiance, the resolution improves.

STED fluorophores exhibit a lifetime in the excited state of only a few nanoseconds, so the probability of depleting the fluorescent transition is relatively low with a correspondingly high value of I_{sat} required for saturation. Typically, picosecond lasers are used in STED with power densities of hundreds of MW/cm^2. In order to increase the probability of depletion, different fluorophores with a triplet state between the ground state and the excited state are utilized, which last for periods of microseconds, thus lowering I_{sat} by a factor of 1000. This variation of STED is called ground-state depletion (GSD) microscopy, and it is advantageous for live-cell imaging due to the dramatically lower power levels required for I_{max} [26.19]. STED can also be carried out with continuous wave (CW) depletion lasers, but the laser power needed for reaching the same resolution is much higher with CW STED compared to pulsed STED [26.20]. Time-gated detection, which precisely times detection with duration of the STED effect to achieve better rejection on spontaneous emission, can greatly improve the resolution under the same I_{sat} value [26.21].

26.5 Photoactivated Localization

Localization microscopy uses a classical microscope system with a special illumination system that limits the number of emitting fluorescent molecules during each frame of the imaging camera. Individual fluorescent molecules are so small that an image of a molecule is essentially the width of the microscope PSF. For each frame, pulses of short wavelength excitation light are used to photochemically activate molecules from a dark state to a bright longer-wavelength fluorescent state. The irradiance of the switching pulse is adjusted so that the fraction of molecules in the bright state is relatively small, and therefore the average spacing between molecules in the bright state is much larger than the width of the PSF in a single frame. There are many variations of localization microscopy, which mainly have to do with the type fluorophore and illumination scheme.

The concept of localization microscopy is shown in Fig. 26.13. Fluorophore molecules are distributed in the object plane, where six individual molecules are identified in this example. In Frame 1, molecules 2 and 6 are actively emitting photons, as indicated by the *open circles*. Molecule 1, and molecules 3–5 are inactive (*dark*). PSF images of molecules 2 and 6 are measured on the camera, where about 1000 detected photons comprise each PSF. The image is processed by fitting the Gaussian approximation of (26.5) to determine centroids of the component PSFs. Assuming that the image is only limited by photon noise, the precision of localizing PSF centroids is approximately [26.22]

$$\sigma_{xy} \approx \frac{\sigma_{\mathrm{PSF}}}{\sqrt{N}} , \tag{26.20}$$

where N is the number of photons in the PSF and σ_{PSF} is the width of the PSF.

Processing continues by replacing each PSF with a smaller Gaussian distribution of width σ_{xy} at the calculated centroid positions. The processed image frame is complete when all PSFs are replaced with their narrower versions. With $N = 1000$, the localization precision is smaller than the PSF by a factor of nearly 32. That is, a good quality microscope with a Rayleigh resolution of 279 nm would exhibit a theoretical processed resolution of 8.8 nm.

After Frame 1 is collected, an illumination mechanism is introduced that switches off molecules 2 and 6 and allows a few inactive molecules, which were dark in Frame 1, to actively emit photons. In Frame 2 of Fig. 26.13, molecules 1 and 5 are emitting photons. The Frame 2 image data are processed in a similar way, and reduced-width Gaussian distributions are positioned according to the component PSF centroids. The processed Frame 2 is then added to the processed Frame 1 to produce a composite superresolved image. Notice that the separation of molecules 5 and 6 would be indistinguishable with standard incoherent illumination. Molecules 1 and 2 would also be indistinguishable with uniform incoherent illumination,

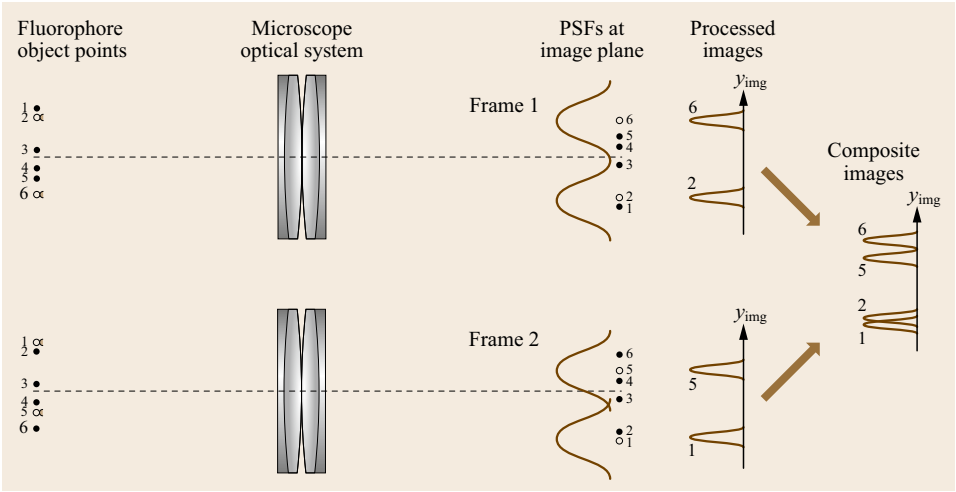


Fig. 26.13 Localization microscopy. Illumination of a distribution of fluorophores in an object is tailored to active emission from only a few molecules at a time, such that the image-lane PSFs are clearly separated. The center of each PSF is determined and replaced by a Gaussian distribution that represents uncertainty of the centroid. Successive frames accumulate the processed images to obtain a complete composite image

and the processed composite image exhibits these two reduced-width Gaussian distributions at the superresolved resolution limit. Consecutive image frames continue with the same process steps until enough data are acquired.

A characteristic of localization microscopy is that each image frame contains only a small number of localizations, so a large number of image frames are required to collect enough data to be scientifically interesting. With an individual frame integration time of 0.5 s, it is not unusual to collect data for one sample in 10^4–10^5 frames over a few hours or even up to twelve hours, depending on the number of frames and fluorophore emission.

For localization microscopy to work, the fluorophores being imaged must exhibit a contrast in photon emission between the *on* (actively emitting photons) and *off* states. This contrast is critical to the performance of the technique. If the off state produces too many photons, the density of fluorophores could be limited, due to fluorescence background noise. Consideration of resolving image structures from the Nyquist sampling theorem [26.23] dictates that the smallest resolvable structure Δ_{Nyquist} is defined by

$$\Delta_{\text{Nyquist}} = \frac{2}{\Sigma^{1/D}} , \qquad (26.21)$$

where Σ is the density of fluorophores and D is the dimension of the structure. For example, $\Delta_{\text{Nyquist}} = 20\,\text{nm}$ for a two-dimensional object requires $\Sigma = 10^4$ per μm^2, in general. However, due to the heterogeneity of biological structures, a high local density is often observed with a low overall areal and/or volumetric density. The types of fluorophores used for localization microscopy are limited by the contrast they provide and by the switching mechanism between on and off states. The following paragraphs briefly describe the differences between a few of the major localization microscopy techniques.

Photoactivated localization microscopy (PALM) is a technique using photoactivatable protein fluorophores (PA-PFs), like photoactivatable green fluorescent protein (PA-GFP) [26.25]. Proteins are continuously excited by a laser, with a power level that allows frames to contain sparse fields with individually resolvable single molecules on a weaker background. Additional frames are captured until single molecule bleaching results in a mean molecular separation that is much larger than that required for isolation. Then, a shorter wavelength laser is pulsed to activate remaining inactive PA-FP molecules at a duration and intensity chosen so that overall density of PA-FP molecules is increased to a higher, but still resolvable level. The fluorescent photoactivated localization microscopy (fPALM) technique uses a similar protocol, but PALM uses total-internal-reflection fluorescence (TIRF) illumination, while fPALM uses a standard wide-field illumination geometry [26.26, 27].

Like with STED microscopy, the invention of PALM/fPALM and related techniques created an abundance of interest in the scientific community. Although many significant advances with this technology have been documented, there is not space in this article to review them in detail. However, it is worth noting some specific examples, at the least for historical reasons. For example, Fig. 26.14 shows an early comparison between TIRF and PALM for clus-

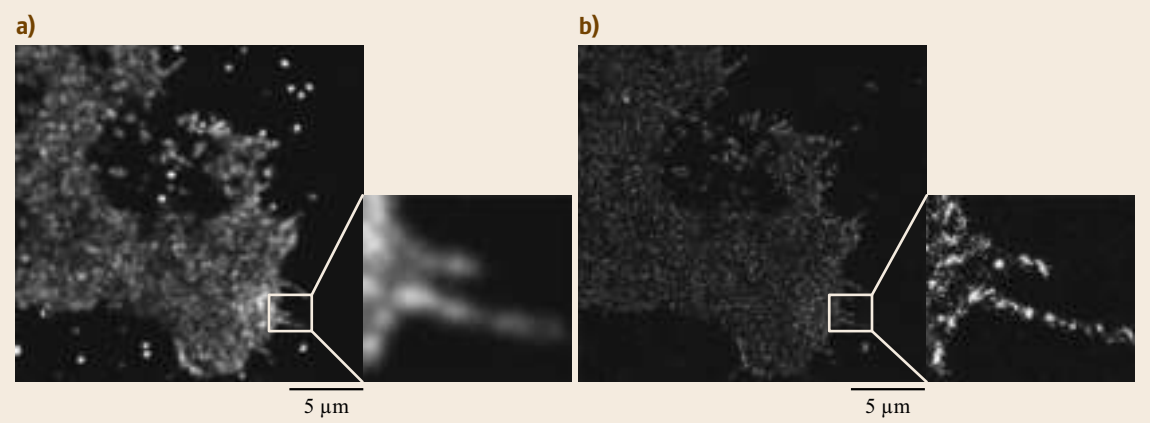


Fig. 26.14a,b Comparison of TIRF and PALM for cluster analysis of LckN10-tdEos. **(a)** Conventional TIRF image; **(b)** PALM image of the same section. Zoomed regions of the images are indicated with *boxes*. Reprinted with permission from [26.24]

ter analysis of LckN10-tdEos [26.24]. PALM localizes peak emission of molecules to about 20 nm. PALM has also been applied to single-particle tracking (spt-PALM) [26.28] and centroid localization techniques have incorporated optimized PSF and advanced statistical methods [26.29].

Stochastic optical reconstruction microscopy (STORM) is a localization microscopy technique that uses pairs of photoswitchable dyes, like a *switch* pair of Cy3 and Cy5 dyes [26.30]. The Cy5 dye can be switched between the on and off states in a controlled and reversible manner by using light of different wavelengths. For example, red light that produces emission from Cy5 dye can also be used to switch the dye to an off state. Exposure to green light converts the Cy5 dye back to an on state, but the recovery rate critically depends on proximity of the Cy5 molecule to a Cy3 secondary dye molecule. When attached to nucleic acids or proteins, illumination conditions appropriate for localization microscopy allow the switch to be cycled on and off hundreds of times before permanent photobleaching occurs.

Direct stochastic optical reconstruction microscopy (dSTORM) uses conventional, standalone carbocyanine dyes, like Cy5 and AlexaFluor 647, that can be switched on and off with a single laser [26.31]. In dSTORM, PA-FPs are genetically fused to target proteins and endogenously expressed in cells and organelles. dSTORM is not limited to biological samples, so it can be used to study polymers or nonbiological optically transparent samples.

Superresolution optical fluctuation imaging (SOFI) is a localization microscopy technique that uses a conventional wide-field microscope and takes advantage of the blinking or intermittency of the emission from fluorophores [26.32]. Like other localization techniques, SOFI requires two states of emission from the fluorophore. Emitters must switch between states repeatedly and independently in a stochastic way. Image acquisition is made with rapid successive frames over a period of time, where emitters can fluctuate over many cycles. To process data, second-order correlation of the dataset provides localization data. Higher-order cumulant calculations provide greatly improved localization precision, including three-dimensional information, and resolution improvement by a factor of five has been achieved.

26.6 Structured Illumination

In structured illumination microscopy, the illumination function I_{ill} in (26.10) is a cosine function. By acquiring a series of images and processing data, improvements of a factor of two in resolution can be obtained. If nonlinear excitation is used, the resolution improvement factor can be several times higher. An advantage to using structured illumination over localization microscopy techniques is that the energy dose for live cell imaging is much lower, so the potential for photodamage is reduced. In the following paragraphs, the experimental configuration, theory and algorithm for structured illumination microscopy are reviewed, and a few examples are provided.

The basic concept for structured illumination is shown in Fig. 26.15. A cosine function I_{ill} illuminates the object for Frame 1 of the image sequence. In biological applications, a distribution of fluorescent molecules defines the object function $o\left(x_{\text{obj}}, y_{\text{obj}}\right)$. Fluorophore molecules are bright, dark or in between as a function of position due to the cosine. The cosine pattern can be generated by imaging a simple grating onto the object, or more complicated holographic methods may be used. The image is collected with a standard high-NA microscope system onto a good-quality camera. For Frame 2, the cosine is shifted slightly by an amount x_{src}, which can be accomplished with a galvanometer mirror. The distribution of brightness from the fluorescent molecules changes, due to the shifted cosine illumination pattern. Several images in a series with small shifts between them are collected. The cosine illumination pattern is rotated in the $\left(x_{\text{obj}}, y_{\text{obj}}\right)$ plane at least two times, with a similar series of images collected for each rotation. The image sequences are then processed in the spatial frequency domain, and a superresolved image is reconstructed.

The illumination function is written

$$I_{\text{ill}}\left(x_{obj} - x_{src}\right) = 0.5 I_0 \left\{1 + \cos\left[\frac{2\pi\left(x_{obj} - x_{src}\right)}{T}\right]\right\}, \tag{26.22}$$

where T is the period of the illumination in the object plane. The frequency spectrum of the illumination is the Fourier transform of the illumination, where

$$\begin{aligned} S\left(\xi, \eta; x_{src}\right) = 0.5 I_0 \{ & \delta\left(\xi\right) \\ & + 0.5\left[\delta\left(\xi - \xi_0\right)\exp(-j\phi)\right. \\ & \left. + \delta\left(\xi + \xi_0\right)\exp(j\phi)\right]\}\delta(\eta) \end{aligned} \tag{26.23}$$

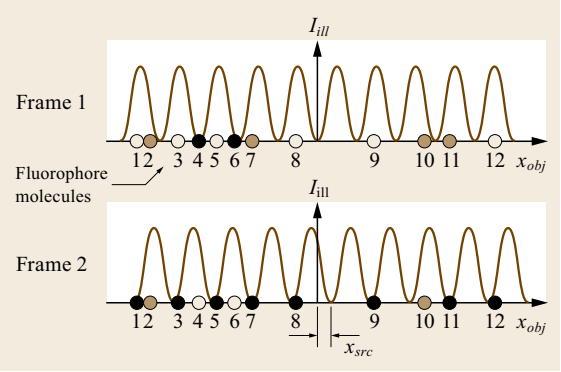


Fig. 26.15 Basic concept of structured illumination imaging. A cosine irradiance illumination pattern is imaged onto the object that excites fluorophores in the tissue. The resulting emission is imaged by a standard microscope onto a good-quality camera. Successive images are acquired, where the cosine illumination is shifted between them. After at least two rotations of the illumination pattern and a shifted image sequence for each rotation, the entire image data set is processed to produce a superresolved image

in which $\phi = 2\pi x_{src}/T$, and $\xi_0 = 1/T$. Application of (26.11) then gives

$$\mathcal{O}(\xi,\eta) = 0.5I_0\{\tilde{o}(\xi,\eta) + 0.5\left[\tilde{o}(\xi-\xi_0,\eta)\exp(-j\phi) + \tilde{o}(\xi-\xi_0,\eta)\exp(j\phi)\right]\}. \quad (26.24)$$

That is, the incoherent spatial frequency content of the light transmitted through the object consists of a central order, which is the object spatial frequency distribution $\tilde{o}(\xi,\eta)$, and two shifted versions of the same distribution. The shifted orders exhibit a phase shift ϕ that is proportional to the grating shift. A graphical interpretation of (26.24) scaled to image frequency space is shown in Fig. 26.16, where the illumination cosine distribution is modified to have components along both unit Cartesian unit vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ defined by the orientation of vector $\boldsymbol{\xi}_0' = \xi_0'\hat{\boldsymbol{x}} + \eta_0'\hat{\boldsymbol{y}}$. The three orders are shown in separate subfigures, although they are all present in the total frequency distribution. The OTF diameter of the microscope imaging system is shown as a *dashed circle*, where *light shading* represents spatial frequencies that are transmitted through the exit pupil. The transmitted portion of the central order is limited by the OTF and is the same content that would be transmitted from uniform incoherent illumination described by (26.9). The upshifted order transmits a small portion of the spatial frequency spectrum that is not transmitted by the central order, as indicated by the *darker shading*. Note that this portion contains higher spatial frequency content than is transmitted by uniform incoherent illumination. The downshifted order also transmits spatial frequencies that are not transmitted with uniform incoherent illumination.

Successive images taken at different (x_{src}, y_{src}) shifts contain overlapped frequency information and must be processed. Conceptually, the algorithm is shown in Fig. 26.17. In a computer, an image is Fourier transformed to produce its spatial frequency distribution, which is inside the OTF. Information is divided according to the grating orientation and period, and higher spatial frequency information is moved to appropriate regions in the frequency plane. Multiple images are obtained with different shifts of the illumination grating to change ϕ in (26.24), and the images are processed to retrieve correct phase values for the translated frequency information. Multiple orientations of the grating are used to fill in other portions of the frequency plane, and normalization of the OTF shape shown in Fig. 26.4 is included in the processing algorithm.

A representative experimental demonstration is shown in Fig. 26.18, where three orientations of the grating are used to produce a total of six shifted components [26.33]. Figure 26.18a shows the incoherent

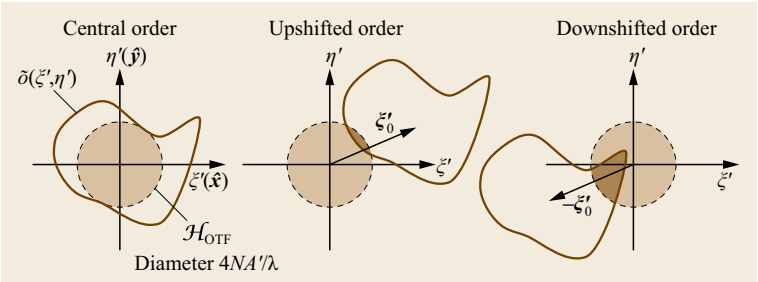


Fig. 26.16 Conceptual distribution of incoherent spatial frequency information with cosine structured illumination. The object's spatial frequency content $\tilde{o}(\xi',\eta')$ is distributed in three orders. The boundary of $\mathcal{H}_{\mathrm{OTF}}(\xi',\eta')$ is indicated by the *dashed circle* and *light shading*. Portions of upshifted and downshifted object spatial frequency distributions are transmitted through $\mathcal{H}_{\mathrm{OTF}}(\xi',\eta')$, as indicated by the *darker shading*

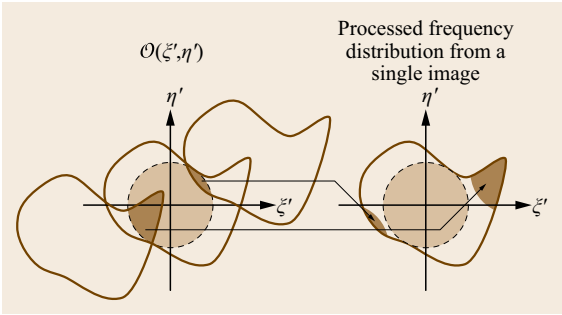


Fig. 26.17 Algorithm of structured illumination. The incoherent object frequency spectrum $\mathcal{O}(\xi', \eta')$ is the combination of the central and shifted orders. For each image, overlapped frequency components are moved to appropriate locations in the frequency plane. Multiple shifts of the illumination cosine provide the phase information for these components. Several orientations of the cosine fill in more space in the frequency plane, and the Fourier transform of the frequency plane is the enhanced-resolution reconstructed image

spatial frequency plane with standard incoherent illumination. Figure 26.18b shows the frequency plane using the illumination cosine at one orientation and shift. Note that the maximum spatial frequency transmitted by the optical system is the same in each case, but the illumination cosine distribution contains information that has been upshifted and downshifted, as indicated by the small arrows. This information is separated and moved in a computer, and it is used to form two of the additional shifted components in Fig. 26.18c, which are then combined to produce the enhanced frequency plane pattern shown in Fig. 26.18d. It is this final enhanced frequency plane pattern that is Fourier transformed to produce the enhanced-resolution image. Resolution enhancement is achieved because the frequency content of the image is effectively doubled. Conceptually, the width of the OTF is doubled and results in an improvement in resolution by a factor of two.

Images are shown in Fig. 26.19 for uniform incoherent illumination and for cosine structured illumination [26.33]. The object is a sample of a HeLa cell, and the image illustrates the actin cytoskeleton at the edge of the cell.

Cosine structured illumination resolution enhancement is limited to a factor of two. However, nonlinear illumination can produce an illumination function resembling

$$I_{\text{ill}}\left(y_{\text{obj}} - y_{\text{src}}\right) = 0.5 I_0 \left(1 + \cos^p \times \left[2\pi \frac{\left(y_{\text{obj}} - y_{\text{src}}\right)}{T}\right]\right), \quad (26.25)$$

where the factor p is the nonlinear enhancement factor. For example, $p = 2$ for a second-order response, like two-photon emission, and $p = 3$ for a third-order response, like three-photon emission. Equation 26.25 now contains harmonic components that expand shifts in the frequency plane, as shown in Fig. 26.20 [26.34]. If $p = 2$, the resolution improvement is a factor of three. If $p = 3$, the resolution improvement is a factor of four. Conceptually, large p could provide extremely high nonlinear enhancement factors. However, the diffraction efficiency from practical illumination patterns limits the number of harmonic components that can be used.

As shown in Fig. 26.20d, the overlapped orders of a nonlinear structured illumination image extend much further into the frequency plane than with cosine structured illumination [26.34]. The processing algorithm is much more complicated, but similar in concept, to the ideas shown in Fig. 26.17. Figure 26.20e shows the frequency space coverage for $p = 4$ with 12 grating

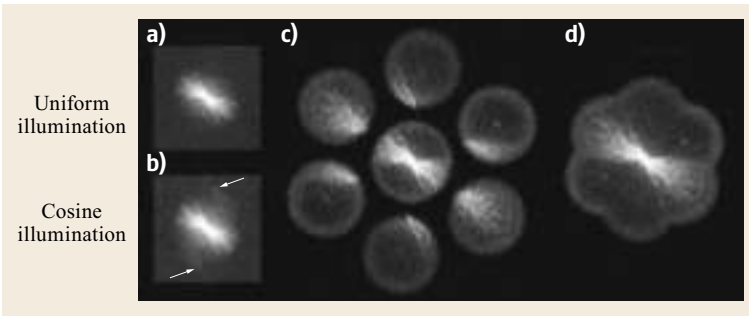


Fig. 26.18a–d Example of frequency-space processing for structured illumination. Reprinted with permission from [26.33]. **(a)** $\mathcal{O}(\xi', \eta')$ using standard uniform incoherent illumination, which is equal to $\tilde{o}(\xi', \eta')$ as described by (26.11); **(b)** $\mathcal{O}(\xi', \eta')$ using cosine structured illumination; **(c)** the central order and separated upshifted and downshifted frequency components from three orientations of the illumination cosine; **(d)** combined frequency information for structured illumination

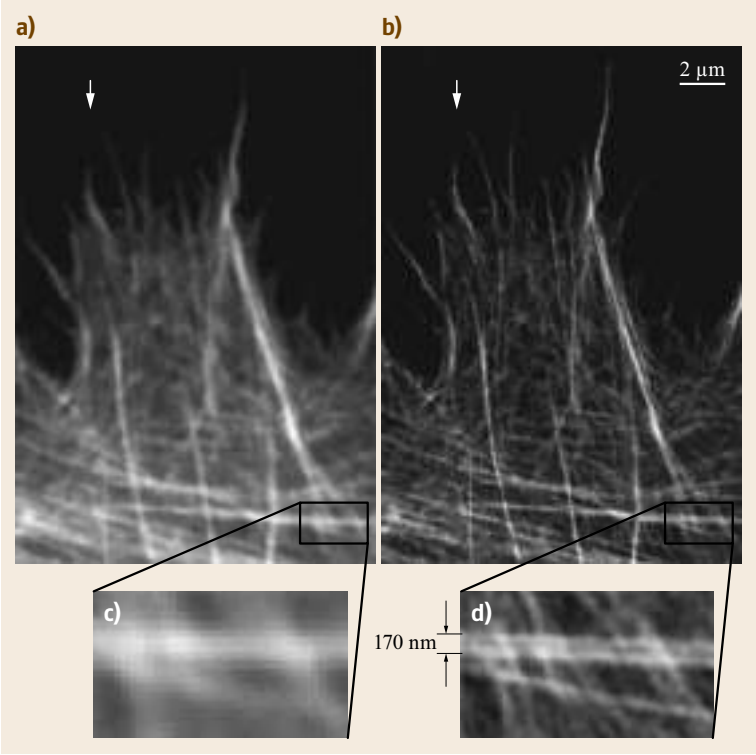


Fig. 26.19a–d Actin cytoskeleton at the edge of a HeLa cell. Reprinted with permission from [26.33]. **(a)** Image with standard uniform incoherent illumination; **(b)** image with structured illumination. **(c)** and **(d)** show fibers separated by less than the resolution limit of the conventional microscope used in the experiment

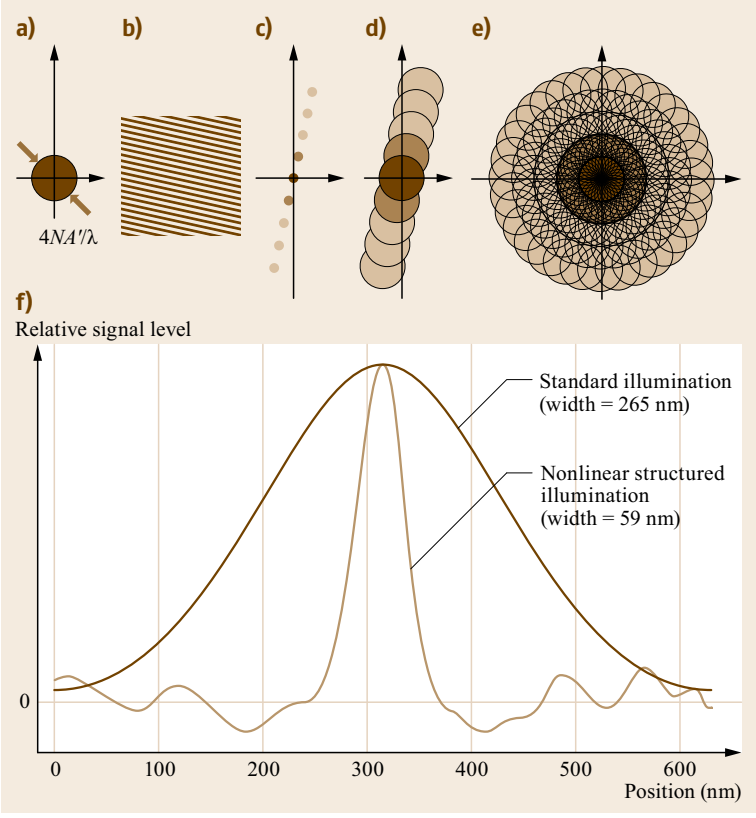


Fig. 26.20a–f Nonlinear structured illumination [26.34]. **(a)** Frequency plane showing area transmitted by $\mathcal{H}_{\mathrm{OTF}}(\xi', \eta')$ as a dark circle; **(b)** nonlinear cosine illumination with $p = 4$; **(c)** frequency plane illustration of $S(\xi', \eta')$ with $p = 4$; **(d)** $\mathcal{O}(\xi', \eta')$ for one nonlinear image; **(e)** the processed frequency plane coverage for $p = 4$; **(f)** line profiles from images of a 50 nm-diameter bead with standard uniform illumination and with $p = 4$ nonlinear illumination

positions that would theoretically provide a resolution improvement factor of five. Figure 26.20f shows line-profile images from a nonlinear structured illumination experiment. The reconstruction from a 50 nm-diameter bead using the nonlinear technique is compared to the profile from a bead using standard incoherent illumination. The measured resolution improvement factor for this example is $265/59 = 4.5$.

26.7 Light-Sheet Microscopy

In several of the superresolution techniques described above, out-of-plane emitters are present that produce a background signal that limits the achievable signal-to-noise ratio [26.35]. Light sheet microscopy, also known as selective plane illumination microscopy (SPIM) [26.36], is a technique for increasing the contrast of points in a thin layer of a sample by illuminating it with a thin in z and wide in xy light beam that is projected through the sample at an angle, as shown in Fig. 26.21. Regardless of the imaging technique, light-sheet illumination effectively increases the contrast of fluorescent emitters within a given light sheet by eliminating out-of-plane emitters [26.37]. In biological applications, SPIM provides a way to optically section living tissue [26.36].

The use of SPIM does not give rise to superresolution by itself. However, when used with techniques like structured illumination [26.38] object features can be distinguished with superresolution deep within the object volume. SPIM is used with one or more fluorophores. The illumination typically uses wavelengths of 532 and 488 nm, which selectively excite fluorophores of interest [26.39]. The molecules to which the fluorophores are attached are distinguished by separation or color of the fluorescence, or both.

In addition to the higher signal-to-noise ratio achieved with light sheet microscopy and the inherent optical sectioning in biological applications, two additional compelling advantages of light-sheet microscopy in biological applications are (1) that broad-field imaging may be used to acquire large datasets simultaneously, thereby allowing high-speed data acquisition of live samples; and (2) avoiding illumination of sections of the sample that are not actively being imaged minimizes photobleaching and phototoxicity in live samples [26.40].

One common way to accomplish light-sheet illumination is to illuminate the sample with a Gaussian beam, as shown in Fig. 26.21a, whose z boundary $w\left(y_{\text{obj}}\right)$ is a function of distance from the focus at $y_{\text{obj}} = 0$ and is described by the hyperbolic function

$$w^2\left(y_{\text{obj}}\right) = \left(\frac{\lambda_{\text{exc}}}{NA_{\text{ill}}}\right)^2 + \left(\frac{NA_{\text{ill}}}{\pi}\right)^2 y_{\text{obj}}^2 , \qquad (26.26)$$

where λ_{exc} is the excitation wavelength and NA_{ill} is the numerical aperture of the illumination [26.38]. This illumination pattern substantially limits the effective z thickness of the sample over the width $w\left(y_{\text{obj}}\right)$ of the beam waist. The xy field is limited to the length of the waist in y_{obj} before it expands to the point where out-of-plane emitters reduce signal-to-noise ratio to an unacceptable level and by the width of the waist produced by the illumination system in the x_{obj} direction. In practice, the Gaussian beam light sheet may be produced by a laser source and a cylindrical lens, or it may be produced by laterally scanning a symmetric Gaussian laser beam.

Transverse spatial resolution of Gaussian SPIM is determined by the NA of the microscope objective lens,

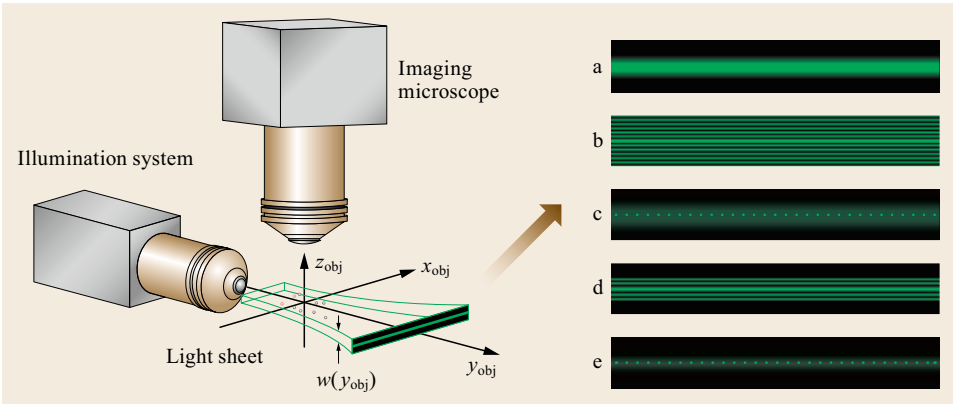


Fig. 26.21 Lattice light-sheet microscopy. A shaped laser beam illuminates a sample volume to excite fluorophores in a thin section that is then imaged by the microscope. Several types of light-sheet illumination patterns are used, which include combining the light sheet sectioning capability with techniques like SIM

Table 26.1 Attributes of advanced light-sheet microscopy techniques

Attribute	Bessel beam plane illumination	Bessel SIM	Two-photon Bessel beam plane illumination	Two-photon Bessel SIM
Typical spatial resolution	$x \approx 230$ nm $y \approx 230$ nm $z \approx 600$ nm	$x \approx 180$ nm $y \approx 230$ nm $z \approx 350$ nm	$x \approx 230$ nm $y \approx 230$ nm $z \approx 350$ nm	$x \approx 180$ nm $y \approx 230$ nm $z \approx 450$ nm
Optical sectioning capability	Fair	Good	Excellent	Excellent
Typical imaging speed	≈ 40 planes/sec	10–20 planes/sec	≈ 40 planes/sec	10–20 planes/sec
Multicolor imaging	Yes	Yes	No	No
Fluorophore brightness requirement	Low	Low	High	High
Photodamage	Low	Fair	Low	High

and is the same as the classical diffraction-limited resolution limit of a wide-field microscope described by (26.6). Axial resolution is related to both the thickness of the illumination light sheet $w\left(y_{\text{obj}}\right)$ and the axial resolution of the detection objective $\lambda_{\text{em}}/n\left(1-\cos\theta_{\text{marg}}\right)$, where n is refractive index of the sample, λ_{em} is the emission wavelength, and θ_{marg} is the marginal ray angle.

A way to accomplish light-sheet illumination over a longer y_{obj} distance is to generate a thin Bessel beam. An ideal Bessel beam propagates indefinitely with no change in its cross-sectional intensity profile $I(r) \propto |J_0(\alpha r)|^2$, where $J_0(\alpha r)$ is a zero-order Bessel function of the first kind. However, because an ideal Bessel beam is created by a spherical lens illuminated at its back focal plane with an infinitesimal annulus of light, an ideal Bessel beam cannot be realized, and a virtual Bessel light sheet must be created by scanning a symmetric Bessel beam. When the scan in the x_{obj} direction is synchronized with the recording camera, the result is an effective Bessel light sheet, as shown in Fig. 26.21b. The Bessel beam is typically produced by an axicon (conically shaped prism) located at the back focal plane of the objective lens. The axicon in combination with scanning produces a light sheet whose width remains constant as it propagates through a substantial distance in the y_{obj} direction, which increases the y_{obj} distance over which individual objects may be distinguished. In reality, an ideal symmetric Bessel beam cannot be achieved due to the limited overlap of the component waves from the axicon and scattering in the sample, so in practice the goal is to produce a beam that is only as long as is needed to cover the field of view. However, a Bessel beam includes concentric rings surrounding the central lobe that also excite molecules to fluoresce, which reduces signal-to noise ratio due to out-of-plane emitters.

Combination of a Bessel beam array, as shown in Fig. 26.21c, with structured illumination (SIM) provides superresolution in the transverse plane, high signal-to-noise ratio and optical sectioning. The array can be formed with a Dammann-type grating in combination with the axicon. In operation, the array is scanned in discrete steps in the x_{obj} direction between each image frame, and images are processed, as discussed in Sect. 26.6, to reconstruct a superresolved image.

Nonlinear, two-photon excitation is used to improve resolution and overcome effects produced by out-of-plane background light caused by rings of the Bessel beam. By using an intense infrared wavelength to supply sufficient energy in the central lobe of the Bessel beam to cause fluorescence through two-photon, nonlinear excitation of the fluorophores. Fluorescence is produced by the central lobe but not by the surrounding rings, as shown in Fig. 26.21d. When used with structured illumination in fluorescence microscopy, excellent optical sectioning and high transverse resolution is achieved. A two-photon Bessel beam array is shown in Fig. 26.21e.

A comparison of attributes for the various advanced light-sheet microscopy techniques is shown in Table 26.1 [26.38].

26.8 Comparison of Techniques for Live-Cell Imaging

A comparison of stimulated-emission depletion (STED/RESLOFT), localization microscopy and structured illumination microscopy (SIM) is shown in Table 26.2 [26.41, 42]. Although reported resolutions of STED and localization microscopy are smaller than for SIM, rapid acquisition time and lower irradiance required for imaging live cells with SIM is particularly advantageous. The required photon flux rate increase of each technique is also compared to conventional microscopy.

Table 26.2 Comparison of superresolution techniques for live-cell imaging

Technique	Reported resolution (nm)	Photon flux increase required	Irradiance (W/cm^2)	Acquisition time (s)
STED/RESLOFT	*xy*: 20 nm *xyz*: 30 nm	100 1070	10^4–10^9	> 60 ≈ 1000
Localization microscopy	*xy*: 20 nm *xy*: 10 nm, *z*: 20 nm	100 14 400	10^3–10^4	> 20 1500
SIM	*xy*: 100 nm *xy*: 100 nm, *z*: 370 nm	4 8	10–10^2	0.1–1 ≈ 10

26.9 Summary

Superresolution optical microscopy has been under modern development since the early 1990s. The six major categories outlined in this chapter provide an overview of field, although there are a large number of variations not discussed that are also interesting. Each of the listed categories has demonstrated resolution well beyond the classical Rayleigh limit, and often object features are resolved to better than 100 nm using visible-light photons. In each case, no new information is created by the microscope optical system compared to when standard uniform illumination is used. Instead, clever ways to increase the transfer of information from the object to the reconstruction algorithm are implemented. Therefore, the imaging process necessarily involves longer acquisition time to achieve adequate signal-to-noise ratio. Trade-offs between how the object or illumination is modified to produce this superresolution have important implications for what type of objects can be imaged. For example, concern about photobleaching or phototoxicity can limit the use of some techniques with live cells. Development in this field continues with the popular goal to see deep within living tissue with optical photons at the resolution of a scanning electron microscope. Although industrial applications are also important, they are overshadowed by the momentum of the biological community. The interested reader is referred to recent reviews and certain microscopy websites for more general information [26.43–46].

References

26.1 Nobel Media AB: The Nobel Prize in Chemistry 2014. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2014/ (2014)

26.2 J.W. Goodman: *Introduction to Fourier optics* (Roberts, Englewood 2005)

26.3 C. Genet, T.W. Ebbesen: Light in tiny holes, Nature **445**, 39–46 (2007)

26.4 B. Hecht, B. Sick, U.P. Wild, V. Deckert, R. Zenobi, O.J. Martin, D.W. Pohl: Scanning near-field optical microscopy with aperture probes: Fundamentals and applications, J. Chem. Phys. **112**, 7761–7774 (2000)

26.5 L. Novotny, B. Hecht: *Principles of Nano-Optics* (Cambridge University Press, Cambridge 2012)

26.6 Z.H. Kim, S.R. Leone: High-resolution apertureless near-field optical imaging using gold nanosphere probes, J. Phys. Chem. B **110**, 19804–19809 (2006)

26.7 F.F. Froehlich, T.D. Milster: Detection of probe dither motion in near-field scanning optical microscopy, Appl. Opt. **34**, 7273–7279 (1995)

26.8 E. Betzig: Principles and applications of near-field scanning optical microscopy (NSOM). In: *Near Field Optics*, NATO ASI Series, Vol. 242, ed. by D.W. Pohl, D. Courjon (Springer, Dordrecht 1993) pp. 7–15

26.9 S. Hell, E. Stelzer: Properties of a 4Pi confocal fluorescence microscope, J. Opt. Soc. Am. A **9**, 2159–2166 (1992)

26.10 J. Bewersdorf, A. Egner, S.W. Hell: 4Pi microscopy. In: *Handbook of Biological Confocal Microscopy*, 3rd edn., ed. by J.B. Pawley (Springer, New York 2006) pp. 561–570

26.11 S. Hell, H. Stelzer: Fundamental improvement of resolution with a 4Pi-confocal fluorescence microscope using two-photon excitation, Optics Commun **93**, 277–282 (1992)

26.12 S.W. Hell, M. Schrader, H.T.M. Van Der Voort: Far-field fluorescence microscopy with three-dimensional resolution in the 100-nm range, J. Microsc. **187**, 1–7 (1997)

26.13 S.W. Hell, J. Wichmann: Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy, Opt. Lett. **19**, 780–782 (1994)

26.14 S.W. Hell: Increasing the resolution of far-field fluorescence light microscopy by point-spread-function engineering. In: *Topics in Fluorescence Spectroscopy*, ed. by J.R. Lakowicz (Springer, New York 2002) pp. 361–426

26.15 V. Westphal, L. Kastrup, S.W. Hell: Lateral resolution of 28 nm ($\lambda / 25$) in far-field fluorescence microscopy, Appl. Phys. B **77**, 377–380 (2003)

26.16 B. Harke, J. Keller, C.K. Ullal, V. Westphal, A. Sch, S.W. Hell: Resolution scaling in STED microscopy, Opt. Express **16**, 4154–4162 (2008)

26.17 S.W. Hell: Toward fluorescence nanoscopy, Nat. Biotechnol. **21**, 1347–1355 (2003)

26.18 M. Hofmann, C. Eggeling, S. Jakobs, S.W. Hell: Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins, Proc. Natl. Acad. Sci. USA **102**, 17565–17569 (2005)

26.19 S.W. Hell, S. Jakobs, L. Kastrup: Imaging and writing at the nanoscale with focused visible light through saturable optical transitions, Appl. Phys. A **77**, 859–860 (2003)

26.20 K.I. Willig, B. Harke, R. Medda, S.W. Hell: STED microscopy with continuous wave beams, Nat. Methods **4**, 915–918 (2007)

26.21 G. Vicidomini, G. Moneron, K.Y. Han, V. Westphal, H. Ta, M. Reuss, J. Engelhardt, C. Eggeling, S.W. Hell: Sharper low-power STED nanoscopy by time gating, Nat. Methods **8**, 571–573 (2011)

26.22 R.E. Thompson, D.R. Larson, W.W. Webb: Precise nanometer localization analysis for individual fluorescent probes, Biophys. J. **82**, 2775–2783 (2002)

26.23 H. Shroff, C.G. Galbraith, J.A. Galbraith, E. Betzig: Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics, Nat. Methods **5**, 417–423 (2008)

26.24 D.M. Owen, C. Rentero, J. Rossy, A. Magenau, D. Williamson, M. Rodriguez, K. Gaus: PALM imaging and cluster analysis of protein heterogeneity at the cell surface, J. Biophotonics **3**, 446–454 (2010)

26.25 E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, H.F. Hess: Imaging intracellular fluorescent proteins at nanometer resolution, Science **313**, 1642–1645 (2006)

26.26 S.T. Hess, T.P. Girirajan, M.D. Mason: Ultra-high resolution imaging by fluorescence photoactivation localization microscopy, Biophys. J. **91**, 4258–4272 (2006)

26.27 S.T. Hess, T.J. Gould, M. Gunewardene, J. Bewersdorf, M.D. Mason: Ultrahigh resolution imaging of biomolecules by fluorescence photoactivation localization microscopy. In: *Micro and Nano Technologies in Bioanalysis: Methods and Protocols*, ed. by R.S. By, J.W.L. Foote (Springer, New York 2009) pp. 483–522

26.28 S. Manley, J.M. Gillette, G.H. Patterson, H. Shroff, H.F. Hess, E. Betzig, J. Lippincott-Schwartz: High-density mapping of single-molecule trajectories with photoactivated localization microscopy, Nat. Methods **5**, 155–157 (2008)

26.29 S. Quirin, S.R.P. Pavani, R. Piestun: Optimal 3D single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions, Proc. Natl. Acad. Sci. **109**, 675–679 (2012)

26.30 M.J. Rust, M. Bates, X. Zhuang: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), Nat. Methods **3**, 793–796 (2006)

26.31 S. van de Linde, A. Löschberger, T. Klein, M. Heidbreder, S. Wolter, M. Heilemann, M. Sauer: Direct stochastic optical reconstruction microscopy with standard fluorescent probes, Nat. Protoc. **6**, 991–1009 (2011)

26.32 T. Dertinger, R. Colyer, G. Iyer, S. Weiss, J. Enderlein: Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI), Proc. Natl. Acad. Sci. **106**, 22287–22292 (2009)

26.33 M.G.L. Gustafsson: Surpassing the resolution limit by a factor of two using structured illumination microscopy, J. Microsc. **198**, 82–87 (2000)

26.34 M.G.L. Gustafsson: Nonlinear structured illumination microscopy: Wide field fluorescence imaging with theoretically unlimited resolution, Proc. Natl. Acad. Sci. **102**, 13081–13086 (2005)

26.35 Y.S. Hu, M. Zimmerley, Y. Li, R. Watters, H. Cang: Single-molecule super-resolution light-sheet microscopy, ChemPhysChem **15**, 577–586 (2014)

26.36 J. Huisken, D. Stainier: Selective plane illumination microscopy techniques in developmental biology, Development **136**, 1963–1975 (2009)

26.37 H. Siedentopf, R. Zsigmondy: Über Sichtbarmachung und Größenbestimmung ultramikoskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser, Ann. Phys. **1**, 1–39 (1903)

26.38 L. Gao, L. Shatio, B. Chen, E. Betzig: 3D live fluorescence imaging of cellular dynamics using Bessel beam plane illumination microscopy, Nat. Protoc. **9**, 1083–1101 (2014)

26.39 P. Santi: Light sheet fluorescence microscopy: A review, J. Histochem. Cytochem. **59**, 129–138 (2011)

26.40 P.J. Keller, H.U. Dodt: Light sheet microscopy of living or cleared specimens, Curr. Opin. Neurobiol. **22**, 138–143 (2012)

26.41 E. Betzig: Nobel Lecture: Single Molecules, Cells, and Super-Resolution Optics, http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2014/betzig-lecture.html (2016)

26.42 L. Schermelleh, R. Heintzmann, H. Leonhardt: A guide to super-resolution fluorescence microscopy, J. Cell Biol. **190**, 165–175 (2010)

26.43 B. Huang, M. Bates, X. Zhuang: Super resolution fluorescence microscopy, Annu. Rev. Biochem. **78**, 993–1016 (2009)

26.44 B.R. Long, D.C. Robinson, H. Zhong: Subdiffractive microscopy: Techniques, applications, and challenges, Wiley Interdiscip. Rev. Syst. Biol. Med. **6**, 151–168 (2014)

26.45 H. Shroff, S.T. Hess, E. Betzig, H.F. Hess, G.H. Patterson, J. Lippincott-Schwartz, M. W. Davidson: Practical Aspect of PALM Imaging, http://zeiss-campus.magnet.fsu.edu/articles/superresolution/palm/practicalaspects.html (2016)

26.46 J.S. Silfies, S.A. Schwartz, M.W. Davidson: Single-Molecule Super-Resolution Imaging, http://www.microscopyu.com/techniques/super-resolution/single-molecule-super-resolution-imaging (2016)

Bio/Nano Part F

Part F Bio/Nanotribology and Bio/Nanomechanics

27 **Nanotribology, Nanomechanics and Materials Characterization**
Bharat Bhushan, Columbus, USA

28 **Surface Forces and Nanorheology of Molecularly Thin Films**
Dong Woog Lee, Ulsan, Republic of Korea
Marina Ruths, Lowell, USA
Jacob N. Israelachvili, Santa Barbara, USA

29 **Atomic Scale Friction Phenomena**
Enrico Gnecco, Jena, Germany
Rémy Pawlak, Basel, Switzerland
Marcin Kisiel, Basel, Switzerland
Thilo Glatzel, Basel, Switzerland
Ernst Meyer, Basel, Switzerland

30 **Computer Simulations of Nanometer-Scale Indentation and Friction**
Susan B. Sinnott, University Park, USA
Seong-Jun Heo, Fremont, USA
Donald W. Brenner, Raleigh, USA
Judith A. Harrison, Annapolis, USA
Douglas L. Irving, Raleigh, USA

31 **Cellular Nanomechanics**
Roger D. Kamm, Cambridge, USA
Jan Lammerding, Ithaca, USA
Mohammad R. K. Mofrad, Berkeley, USA

32 **Nanomechanical Properties of Nanostructures and Scale Effects**
Bharat Bhushan, Columbus, USA

27. Nanotribology, Nanomechanics and Materials Characterization

Bharat Bhushan

Nanotribology and nanomechanics studies are needed to develop a fundamental understanding of interfacial phenomena on a small scale, and to study interfacial phenomena in micro/nanoelectromechanical systems (MEMS/NEMS), magnetic storage devices, and many other applications. Friction and wear of lightly loaded micro/nanocomponents are highly dependent on surface interactions (a few atomic layers). These structures are generally coated with molecularly thin films. Nanotribology and nanomechanics studies are also valuable in the fundamental understanding of interfacial phenomena in macrostructures, and provide a bridge between science and engineering. An atomic force microscope (AFM) tip is used to simulate a single-asperity contact with a solid or lubricated surface. AFMs are used to study the various tribological phenomena, which include surface roughness, adhesion, friction, scratching, wear, detection of material transfer, and boundary lubrication. In situ surface characterization of local deformation of materials and thin coatings can be carried out using a tensile stage inside an AFM. Mechanical properties such as hardness, Young's modulus of elasticity, and creep/relaxation behavior can be determined on micro- to picoscales using a depth-sensing indentation system in an AFM. Localized surface elasticity and viscoelastic mapping near surface regions can be obtained with nanoscale lateral resolution. Finally, an AFM can be used for nanofabrication/nanomachining.

27.1 **Description of AFM/FFM and Various Measurement Techniques** 871
27.1.1 Surface Roughness and Friction Force Measurements 872
27.1.2 Adhesion Measurements 875
27.1.3 Scratching, Wear, and Fabrication/Machining 875
27.1.4 Surface Potential Measurements 876
27.1.5 In situ Characterization of Local Deformation Studies 876
27.1.6 Nanoindentation Measurements 877
27.1.7 Localized Surface Elasticity and Viscoelasticity Mapping 878
27.1.8 Boundary Lubrication Measurements 881
27.1.9 AFM Tip Wear 881

27.2 **Surface Imaging, Friction, and Adhesion** 882
27.2.1 Atomic-Scale Imaging and Friction 882
27.2.2 Microscale Friction 884
27.2.3 Directionality Effect on Microfriction 887
27.2.4 Surface Roughness-Independent Microscale Friction 891
27.2.5 Velocity Dependence on Micro/Nanoscale Friction 893
27.2.6 Nanoscale Friction and Wear Mapping 898
27.2.7 Adhesion and Friction in Wet Environments 899
27.2.8 Separation Distance Dependence of Meniscus and van der Waals Forces 901
27.2.9 Scale Dependence in Friction 902

27.3 **Micro/Nanoscale Wear and Scratching, Local Deformation, and Nanofabrication/Nanomachining** 906
27.3.1 Nanoscale Wear 906
27.3.2 Microscale Scratching 906
27.3.3 Microscale Wear 907
27.3.4 In situ Characterization of Local Deformation 911
27.3.5 Nanofabrication/Nanomachining 914

27.4 **Indentation** 915
27.4.1 Picoindentation 915
27.4.2 Nanoindentation 915
27.4.3 Localized Surface Elasticity and Viscoelasticity Mapping 917

27.5 **Boundary Lubrication** 919
27.5.1 Perfluoropolyether Lubricants 919
27.5.2 Self-Assembled Monolayers 924
27.5.3 Liquid Film Thickness Measurements 926

27.6 **Conclusion** 928

References 929

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_27

The mechanisms and dynamics of the interactions of two contacting solids during relative motion, ranging from atomic scale to microscale, need to be understood in order to develop a fundamental understanding of adhesion, friction, wear, indentation, and lubrication processes. For most solid–solid interfaces of technological relevance, contact occurs at multiple asperities. Consequently, the importance of investigating single-asperity contacts in studies of the fundamental micro/nanomechanical and micro/nanotribological properties of surfaces and interfaces has long been recognized. The emergence and proliferation of proximal probes, in particular scanning probe microscopies (the scanning tunneling microscope and the atomic force microscope), the surface force apparatus, and computational techniques for simulating tip–surface interactions and interfacial properties, allowed systematic investigations of interfacial problems with high resolution as well as ways and means for modifying and manipulating nanoscale structures. These advances led to the appearance of the field of nanotribology. The word nanotribology was used for the first time in the title of a research paper by *Bhushan* et al. [27.1] and a first book on the field by Bhushan in 1995 (first edition of [27.2]). The first dedicated international meeting on the topic was organized in 1996 in Sesimbra, Portugal [27.3].

Nanotribology pertains to experimental and theoretical investigations of interfacial processes on scales ranging from the atomic- and molecular scale to the microscale, occurring during adhesion, friction, scratching, wear, indentation, and thin-film lubrication at sliding surfaces [27.1–14]. Proximal probes have also been used for mechanical and electrical characterization, in situ characterization of local deformation, and other nanomechanics studies [27.8, 11, 14, 15].

Nanotribological and nanomechanical studies are needed to develop a fundamental understanding of interfacial phenomena on a small scale, and to study interfacial phenomena in nanostructures used in magnetic storage devices, micro/nanoelectromechanical systems (MEMS/NEMS), and other applications [27.1–3, 7–20]. Friction and wear of lightly loaded micro/nanocomponents are highly dependent on surface interactions (a few atomic layers). These structures are generally coated with molecularly thin films. Nanotribological and nanomechanical studies also are valuable in the fundamental understanding of interfacial phenomena in macrostructures, and provide a bridge between science and engineering.

The surface force apparatus (SFA), the scanning tunneling microscopes (STMs), and atomic force and friction force microscopes (AFM and FFMs) are used widely in nanotribological and nanomechanical studies. Typical operating parameters are compared in Table 27.1. The SFA, developed in 1968, is commonly employed to study both static and dynamic properties of molecularly thin films sandwiched between two molecularly smooth surfaces. The STM, developed in 1981, allows the imaging of electrically conducting surfaces with atomic resolution, and has been used for the imaging of clean surfaces as well as of lubricant molecules. The introduction of the AFM in 1985 provided a method for measuring ultrasmall forces between a probe tip and an engineering (electrically conducting or insulating) surface. AFM has been used for morphological and surface roughness measurements of surfaces on the nanoscale, as well as for adhesion measurements. Subsequent modifications of the AFM led to the development of the FFM, designed for atomic- and microscale studies of friction. This instrument measures forces in the scanning direction. The AFM is being used

Table 27.1 Comparison of typical operating parameters in SFA, STM, and AFM/FFM used for micro/nanotribological studies

Operating parameter	SFA	STM[a]	AFM/FFM
Radius of mating surface/tip	≈ 10 mm[b]	5–100 nm	5–100 nm
Radius of contact area	10–40 μm	N/A	0.05–0.5 nm
Normal load	10–100 mN	N/A	< 0.1–500 nN
Sliding velocity	0.001–100 μm/s	0.02–200 μm/s (scan size ≈ 1 nm × 1 nm to 125 μm × 125 μm; scan rate < 1–122 Hz)	0.02–200 μm/s (scan size ≈ 1 nm × 1 nm to 125 μm × 125 μm; scan rate < 1–122 Hz)
Sample limitations	Typically atomically smooth, optically transparent mica; opaque ceramic, smooth surfaces can also be used	Electrically conducting samples	None of the above

[a] Can be used for atomic-scale imaging.
[b] Since stresses scale inverse of tip radius, SFA can provide very low stress measurement capabilities.

for various investigations including scratching, wear, indentation, detection of transfer of material, boundary lubrication, and fabrication and machining [27.1, 2, 14, 15]. Meanwhile, significant progress in understanding the fundamental nature of bonding and interactions in materials, combined with advances in computer-based modeling and simulation methods, has allowed theoretical studies of complex interfacial phenomena with high resolution in space and time. Such simulations provide insights into atomic-scale energetics, structure, dynamics, thermodynamics, transport, and rheological aspects of tribological processes.

The nature of interactions between two surfaces brought together closely, and those between two contacting surfaces as they are separated, have been studied experimentally with the SFA. This has led to a basic understanding of the normal forces between surfaces and the way in which these are modified by the presence of thin liquid or a polymer film. The frictional properties of such systems have been studied by moving the surfaces laterally. Such experiments have provided insights into the molecular-scale operation of lubricants such as thin liquid or polymer films. Complementary to these studies are those in which the AFM tip is used to simulate a single-asperity contact with a solid or lubricated surface (Fig. 27.1). These experiments have demonstrated that the relationship between friction and surface roughness is not always simple or obvious. AFM studies have also revealed much about the nanoscale nature of intimate contact during wear, indentation, and lubrication.

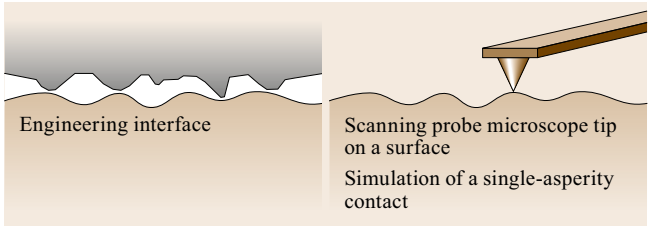


Fig. 27.1 Schematics of an engineering interface and scanning probe microscope tip in contact with an engineering interface

In this chapter, we present a review of significant aspects of nanotribological, nanomechanical, and materials characterization studies conducted using AFM/FFM.

27.1 Description of AFM/FFM and Various Measurement Techniques

The AFM was developed by Gerd Binnig et al. in 1985. It is capable of investigating surfaces of scientific and engineering interest on an atomic scale [27.21, 22]. The AFM relies on a scanning technique to produce very high-resolution, three-dimensional images of sample surfaces. It measures ultrasmall forces (less than 1 nN) present between the AFM tip surface mounted on a flexible cantilever beam and a sample surface. These small forces are obtained by measuring the motion of a very flexible cantilever beam having an ultrasmall mass, by a variety of measurement techniques including optical deflection, optical interference, capacitance, and tunneling current. The deflection can be measured to within 0.02 nm, so for a typical cantilever spring constant of 10 N/m, a force as low as 0.2 nN can be detected. To put these numbers in perspective, individual atoms and human hair are typically a fraction of a nanometer and about 75 μm in diameter, respectively, and a drop of water and an eyelash have a mass of about 10 μN and 100 nN, respectively. In the operation of high-resolution AFM, the sample is generally scanned rather than the tip because any cantilever movement would add vibrations. AFMs are available for the measurement of large samples, where the tip is scanned and the sample is stationary. To obtain an atomic resolution with the AFM, the spring constant of the cantilever should be weaker than the equivalent spring between atoms. A cantilever beam with a spring constant of about 1 N/m or lower is desirable. For high lateral resolution, tips should be as sharp as possible. Tips with a radius ranging from 5 to 50 nm are commonly available. Interfacial forces, adhesion, and surface roughness, including atomic-scale imaging, are routinely measured using the AFM.

A modification to the AFM providing a sensor to measure the lateral force led to the development of the friction force microscope (FFM) or the lateral force microscope (LFM), designed for atomic- and microscale studies of friction [27.1–3, 7, 9, 10, 14, 23–38] and lubrication [27.12, 20, 39, 39–54]. This instrument measures lateral or friction forces (in the plane of the sample surface and in the scanning direction). By using a standard or a sharp diamond tip mounted on a stiff cantilever beam, AFM is used in investigations of scratching and wear [27.1, 2, 7, 8, 11, 28, 55–58], nanoindentation [27.1, 11, 18, 28, 59–63], and nanofabrication/machining [27.2, 20, 28]. An oscillating cantilever is used for localized surface elasticity and viscoelastic mapping, referred to as dynamic AFM [27.36, 64–72]. In situ surface characterization of local deformation of materials and thin coatings has been carried out by imaging the sample surfaces using an AFM, during tensile deformation using a tensile stage [27.73–75].

27.1.1 Surface Roughness and Friction Force Measurements

Surface height imaging down to atomic resolution of electrically conducting surfaces is carried out using an STM. An AFM is also used for surface height imaging and roughness characterization down to nanoscale. Commercial AFM/FFM are routinely used for simultaneous measurements of surface roughness and friction force [27.2, 14]. These instruments are available for the measurement of small samples and large samples. In a small-sample AFM shown in Fig. 27.2a, the sample, generally no larger than 10 mm × 10 mm, is mounted on a piezoelectric crystal in the form of a cylindrical tube (referred to as a PZT tube scanner) which consists of separate electrodes to scan the sample precisely in the x–y plane in a raster pattern and to move the sample in the vertical (z) direction. A sharp tip at the free end of a flexible cantilever is brought into contact with the sample. The normal and frictional forces being applied at the tip–sample interface are measured using a laser beam deflection technique. A laser beam from a diode laser is directed by a prism onto the back of a cantilever near its free end, tilted downward at about 10° with respect to the horizontal plane. The reflected beam from the vertex of the cantilever is directed through a mirror onto a quad photodetector (a split photodetector with four quadrants). The differential signal from the top and bottom photodiodes provides the AFM signal which is a sensitive measure of the cantilever vertical deflection. Topographic features of the sample cause the tip to deflect in the vertical direction as the sample is scanned under the tip. This tip deflection will change the direction of the reflected laser beam, changing the intensity difference between the top and bottom sets of photodetectors (the AFM signal). In the AFM operating mode called the height mode, for topographic imaging or for any other operation in which the applied normal force is to be kept constant, a feedback circuit is used to modulate the voltage applied to the PZT scanner to adjust the height of the PZT, so that the cantilever vertical deflection (given by the intensity difference between the top and bottom detector) will remain constant during scanning. The PZT height variation is thus a direct measure of the surface roughness of the sample.

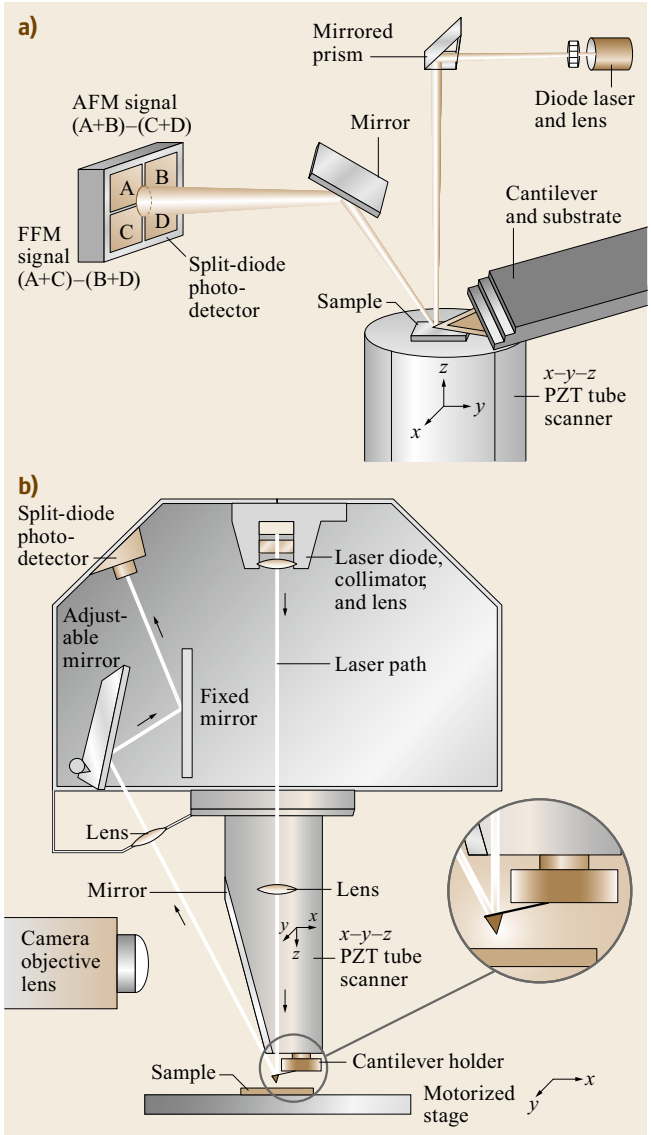


Fig. 27.2a,b Schematics **(a)** of a commercial small-sample atomic force microscope/friction force microscope (AFM/FFM), and **(b)** of a large-sample AFM/FFM

In a large-sample AFM, both force sensors using the optical deflection method and a scanning unit are mounted on the microscope head (Fig. 27.2b). Because of vibrations caused by the cantilever movement, lateral resolution of this design can be somewhat poorer than the design in Fig. 27.2a in which the sample is scanned instead of the cantilever beam. The advantage of the large-sample AFM is that large samples can be measured readily.

Most AFMs can be used for surface roughness measurements in the so-called tapping mode (TM) (intermittent contact mode), also referred to as dynamic (atomic) force microscopy. In the tapping mode, during scanning over the surface, the cantilever/tip assembly with a normal stiffness of 20–100 N/m (Bruker tapping mode etched Si probe or TESP) is sinusoidally vibrated at its resonance frequency (350–400 kHz) by a piezo mounted above it, and the oscillating tip slightly taps the surface. The piezo is adjusted using the feedback control in the z-direction to maintain a constant

(20–100 nm) oscillating amplitude (setpoint) and constant average normal force (Fig. 27.3) [27.2, 14]. The feedback signal to the z-direction sample piezo (to keep the setpoint constant) is a measure of surface roughness. The cantilever/tip assembly is vibrated at some amplitude, here referred to as the free amplitude, before the tip engages the sample. The tip engages the sample at some setpoint, which may be thought of as the amplitude of the cantilever as influenced by contact with the sample. The setpoint is defined as a ratio of the vibration amplitude after engagement to the vibration amplitude in free air before engagement. A lower setpoint gives a reduced amplitude and closer mean tip-to-sample distance. The amplitude should be kept large enough so that the tip does not get stuck to the sample because of adhesive attractions. Also the oscillating amplitude applies less average (normal) load as compared to the contact mode and reduces sample damage. The tapping mode is used in topography measurements to minimize the effects of friction and other lateral forces and to measure topography of soft surfaces.

To measure the friction force at the tip surface during sliding, the left hand and right hand sets of quadrants of the photodetector are used. In the so-called friction mode, the sample is scanned back and forth in a direction orthogonal to the long axis of the cantilever beam. A friction force between the sample and the tip will produce a twisting of the cantilever. As a result, the laser beam will be reflected out of the plane defined by the incident beam and the beam reflected vertically from an untwisted cantilever. This produces an intensity difference of the laser beam received in the left hand and right hand sets of quadrants of the photodetector. The intensity difference between the two sets of detectors (FFM signal) is directly related to the degree of twisting and hence to the magnitude of the friction force. One problem associated with this method is that any misalignment between the laser beam and the photodetector axis would introduce error in the measurement. However, by following the procedures developed by *Ruan* and *Bhushan* [27.25], in which the average FFM signal for the sample scanned in two opposite directions is subtracted from the friction profiles of each of the two scans, the misalignment effect is eliminated. This method provides three-dimensional maps of friction force. By following the friction force calibration procedures developed by *Ruan* and *Bhushan* [27.25], voltages corresponding to friction forces can be converted to force units [27.76, 77]. The coefficient of friction is obtained from the slope of friction force data measured as a function of normal loads typically ranging from 10 to 150 nN. This approach eliminates any contributions due to the adhesive forces [27.28]. To calculate the coefficient of friction based on a single point measurement, friction force should be divided by the sum of the applied normal load and intrinsic adhesive

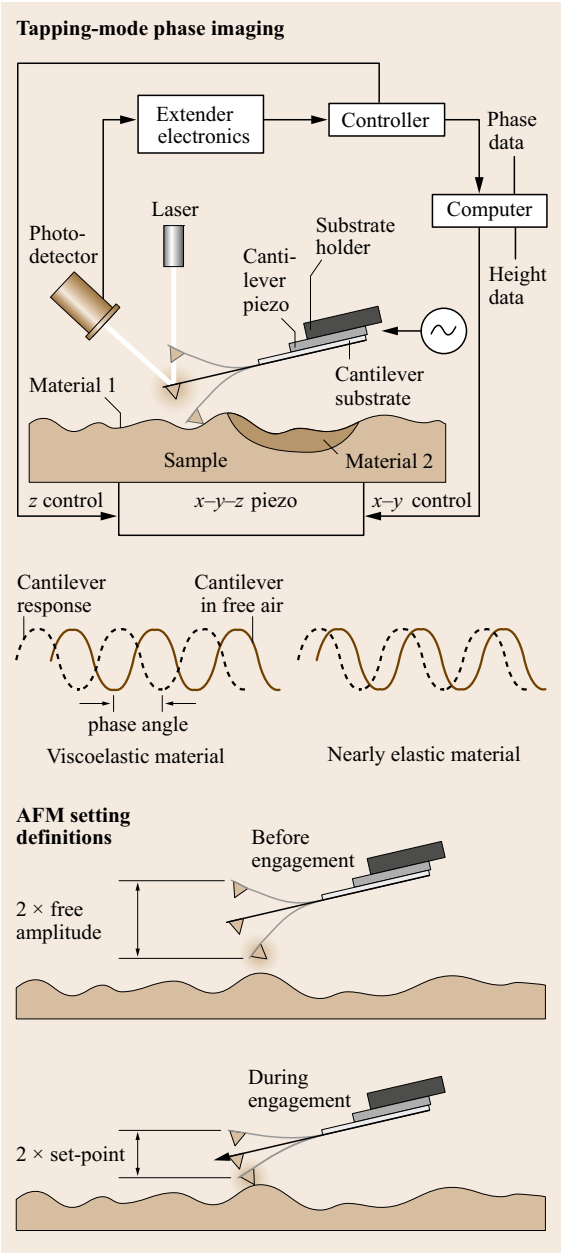


Fig. 27.3 Schematic of tapping mode used to obtain height and phase data and definitions of free amplitude and setpoint. During scanning, the cantilever is vibrated at its resonance frequency and the sample x–y–z piezo is adjusted by feedback control in the z-direction to maintain a constant setpoint. The computer records height (which is a measure of surface roughness) and phase angle (which is a function of the viscoelastic properties of the sample) data

force. Furthermore, it should be pointed out that for a single-asperity contact, the coefficient of friction is not independent of load (see discussion later).

Surface roughness measurements in the contact mode are typically made using a sharp, microfabricated square-pyramidal Si_3N_4 tip with a radius of 30–50 nm on a triangular cantilever beam (Fig. 27.4a) with normal stiffness on the order of 0.06–0.58 N/m with a normal natural frequency of 13–40 kHz (Bruker silicon nitride probe or NP) at a normal load of about 10 nN, and friction measurements are carried out in the load range of 1–100 nN. Surface roughness measurements in the tapping mode utilize a stiff cantilever with high resonance frequency; typically a square-pyramidal etched single-crystal silicon tip, with a tip radius of 5–10 nm, integrated with a stiff rectangular silicon cantilever beam (Fig. 27.4a) with a normal stiffness on the order of 17–60 N/m and a normal resonance frequency of 250–400 kHz (DI TESP), is used. Multiwalled carbon nanotube (MWNT) tips having a small diameter (few nm) and a length of about 1 μm (high aspect ratio) attached on the single-crystal silicon, square-pyramidal tips are used for the high-resolution imaging of surfaces and of deep trenches in the tapping mode (noncontact mode) (Fig. 27.4b) [27.78]. The MWNT tips are hydrophobic. To study the effect of radius of a single asperity (tip) on adhesion and friction, microspheres of silica with radii ranging from about 4 to 15 μm are attached at the end of cantilever beams. Optical micrographs of two of the microspheres at the ends of triangular cantilever beams are shown in Fig. 27.4c.

The tip is scanned in such a way that its trajectory on the sample forms a triangular pattern (Fig. 27.5). Scanning speeds in the fast and slow scan directions depend on the scan area and scan frequency. Scan sizes ranging from less than 1×1 nm to 125×125 μm and scan rates from less than 0.5 to 122 Hz typically can be used. Higher scan rates are used for smaller scan lengths. For

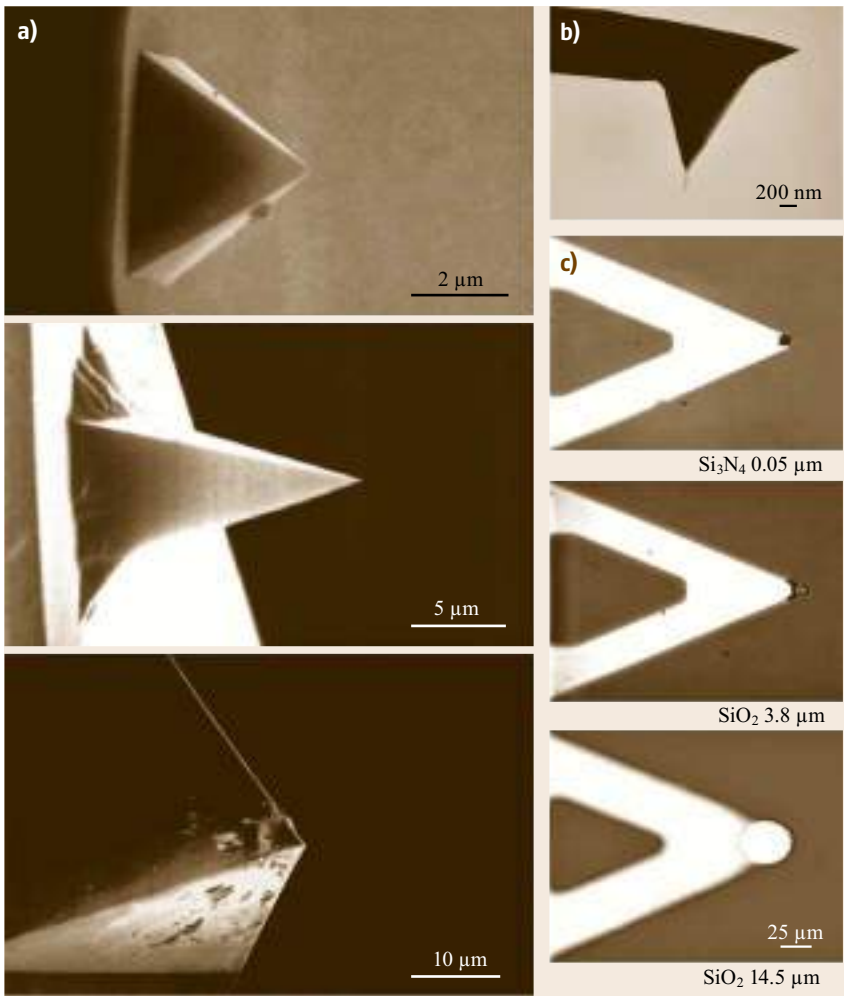


Fig. 27.4 **(a)** SEM micrographs of a square-pyramidal plasma-enhanced chemical vapor deposition (PECVD) Si_3N_4 tip with a triangular cantilever beam, a square-pyramidal etched single-crystal silicon tip with a rectangular silicon cantilever beam, and a three-sided pyramidal natural diamond tip with a square stainless steel cantilever beam. **(b)** SEM micrograph of a MWNT physically attached on the single-crystal silicon, square-pyramidal tip, and **(c)** optical micrographs of a commercial Si_3N_4 tip and two modified tips showing SiO_2 spheres mounted over the sharp tip, at the end of the triangular Si_3N_4 cantilever beams (radii of the tips are given in the figure)

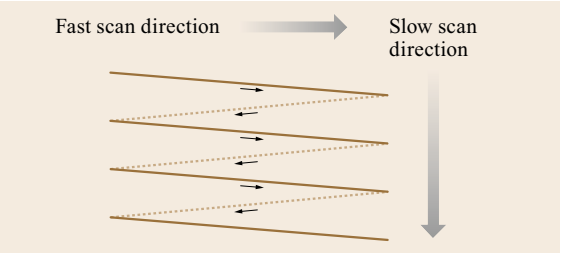


Fig. 27.5 Schematic of triangular pattern trajectory of the tip as the sample (or the tip) is scanned in two dimensions. During scanning, data are recorded only during scans along the *solid scan lines*

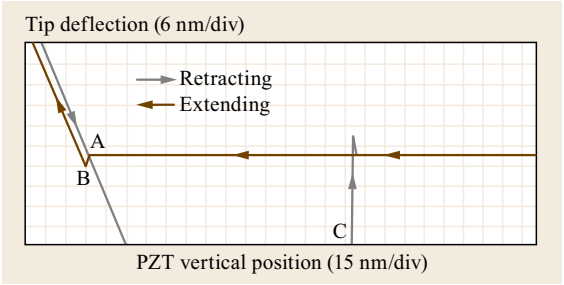


Fig. 27.6 Typical force–distance curve for contact between an Si_3N_4 tip and single-crystal silicon surface in measurements made in the ambient environment. Snap-in occurs at point A; contact between the tip and silicon occurs at point B; tip breaks free of adhesive forces at point C as the sample moves away from the tip

example, scan rates in the fast and slow scan directions for an area of $10\,\mu m \times 10\,\mu m$ scanned at 0.5 Hz are $10\,\mu m/s$ and 20 nm/s, respectively.

27.1.2 Adhesion Measurements

Adhesive force measurements are performed in the so-called force calibration mode. In this mode, force-distance curves are obtained, for an example see Fig. 27.6. The horizontal axis gives the distance that the piezo (and hence the sample) travels, and the vertical axis gives the tip deflection. As the piezo extends, it approaches the tip, which is at this point in free air and hence shows no deflection. This is indicated by the flat portion of the curve. As the tip approaches the sample within a few nanometers (point A), an attractive force exists between the atoms of the tip surface and the atoms of the sample surface. The tip is pulled towards the sample and contact occurs at point B on the graph. From this point on, the tip is in contact with the surface and as the piezo further extends, the tip gets further deflected. This is represented by the sloped portion of the curve. As the piezo retracts, the tip goes beyond the zero deflection (flat) line because of attractive forces (van der Waals forces and long-range meniscus forces), into the adhesive regime. At point C in the graph, the tip snaps free of the adhesive forces and is again in free air. The horizontal distance between points B and C along the retrace line gives the distance moved by the tip in the adhesive regime. This distance multiplied by the stiffness of the cantilever gives the adhesive force. Incidentally, the horizontal shift between the loading and unloading curves results from the hysteresis in the PZT tube [27.2, 14].

27.1.3 Scratching, Wear, and Fabrication/Machining

For microscale scratching, microscale wear, nanofabrication/nanomachining, and nanoindentation hardness measurements, an extremely hard tip is required. A three-sided pyramidal single-crystal natural diamond tip with an apex angle of 80° and a radius of about 100 nm mounted on a stainless steel cantilever beam with normal stiffness of about 25 N/m is used at relatively higher loads (1–150 μN; Fig. 27.4a). For scratching and wear studies, the sample is generally scanned in a direction orthogonal to the long axis of the cantilever beam (typically at a rate of 0.5 Hz) so that friction can be measured during scratching and wear. The tip is mounted on the cantilever such that one of its edges is orthogonal to the long axis of the beam. Therefore, wear during scanning along the beam axis is higher (about 2–3×) than that during scanning orthogonal to the beam axis. For wear studies, an area on the order of $2\,\mu m \times 2\,\mu m$ is scanned at various normal loads (ranging from 1 to 100 μN) for a selected number of cycles [27.2, 14, 28].

Scratching can also be performed at ramped loads and the coefficient of friction can be measured during scratching [27.58]. A linear increase in the normal load approximated by a large number of normal load increments of small magnitude is applied using a software interface (lithography module in Nanoscope III) that allows the user to generate controlled movement of the tip with respect to the sample. The friction signal is tapped out of the AFM and is recorded on a computer. A scratch length on the order of 25 μm and a velocity on the order of $0.5\,\mu m/s$ are used and the number of loading steps is usually taken to be 50.

Nanofabrication/nanomachining is conducted by scratching the sample surface with a diamond tip at specified locations and scratching angles. The normal load used for scratching (writing) is on the order of 1–100 μN with a writing speed on the order of $0.1–200\,\mu m/s$ [27.2, 8, 14, 20, 28, 79].

27.1.4 Surface Potential Measurements

Various electrical measurements on the nanoscale can be made using a conducting tip. The conducting tip acts as a voltage electrode to provide stimuli and monitor surface electrical properties [27.80]. Various electrical techniques are used and include scanning Kelvin probe microscopy (SKPM), scanning capacitance microscopy (SCM), and scanning spreading resistance microscopy (SSRM) [27.53, 81, 82].

Kelvin probe microscopy can be used to detect wear precursors and to study the early stages of localized wear [27.81, 83, 84]. In this technique (also called *nano-Kelvin probe* technique), the potential difference is measured between the tip and the sample by applying a DC bias potential and an oscillating (AC (alternating current)) potential to a conducting tip over a grounded substrate. Mapping of the surface potential is carried out in the so-called *lift mode* (Fig. 27.7). These measurements are made simultaneously with the topography scan in the tapping mode, using an electrically conducting (nickel-coated single-crystal silicon) tip. After each line of the topography scan is completed, the feedback loop controlling the vertical piezo is turned off, and the tip is lifted from the surface and traced over the same topography at a constant distance of 100 nm. During the lift mode, a DC bias potential and an oscillating potential (3–7 V) is applied to the tip. The frequency of oscillation is chosen to be equal to the resonance frequency of the cantilever (≈ 80 kHz). When a DC bias potential equal to the negative value of surface potential of the sample (on the order of +2 V) is applied to the tip, it does not vibrate. During scanning, a difference between the DC bias potential applied to the tip and the potential of the surface will create DC electric fields that interact with the oscillating charges (as a result of the AC potential), causing the cantilever to oscillate at its resonance frequency, as in tapping mode. However, a feedback loop is used to adjust the DC bias on the tip to exactly nullify the electric field, and thus the vibrations of the cantilever. The required bias voltage follows the localized potential of the surface. The surface potential was obtained by reversing the sign of the bias potential provided by the electronics [27.81, 84].

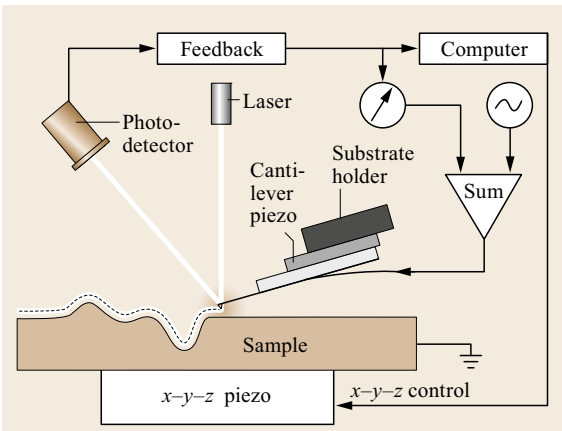


Fig. 27.7 Schematic of lift mode used to make surface potential measurement. The topography is collected in tapping mode in the primary scan. The cantilever piezo is de-activated. Using topography information of the primary scan, the cantilever is scanned across the surface at a constant height above the sample. An oscillating voltage at the resonant frequency is applied to the tip and a feedback loop adjusts the DC (direct current) bias of the tip to maintain the cantilever amplitude at zero. The output of the feedback loop is recorded by the computer and becomes the surface potential map

Surface and subsurface changes of structure and/or chemistry can cause changes in the measured potential of a surface. Thus, mapping of the surface potential after sliding can be used for detecting wear precursors and studying the early stages of localized wear [27.81, 83, 84].

27.1.5 In situ Characterization of Local Deformation Studies

In situ characterization of local deformation of materials can be carried out by performing tensile, bending, or compression experiments inside an AFM and by observing nanoscale changes during the deformation experiment [27.18]. In these experiments, small deformation stages are used to deform the samples inside an AFM. For tensile testing of the polymeric films carried out by *Bobji* and *Bhushan* [27.73, 74] and *Tambe* and *Bhushan* [27.75] and of the hair by *Seshadri* and *Bhushan* [27.85, 86], a tensile stage was used (Fig. 27.8). The stage with a left-right combination lead screw (which helps to move the slider in the opposite direction) was used to stretch the sample to minimize the movement of the scanning area, which was kept close to the center of the tensile specimen. One end of the sample was mounted on the slider via a force sensor to monitor the tensile load. The samples were stretched for various strains using a stepper motor and the same control area at different strains was imaged. In order to better locate the control area for imaging, a set of four markers was created at the corners of a 30 μm × 30 μm square at the center of the sample by scratching the sample with a sharp silicon tip. The scratching depth was controlled such that it did not affect the cracking behavior of the coating. A minimum displacement of 1.6 μm could be obtained. This corresponded to a strain

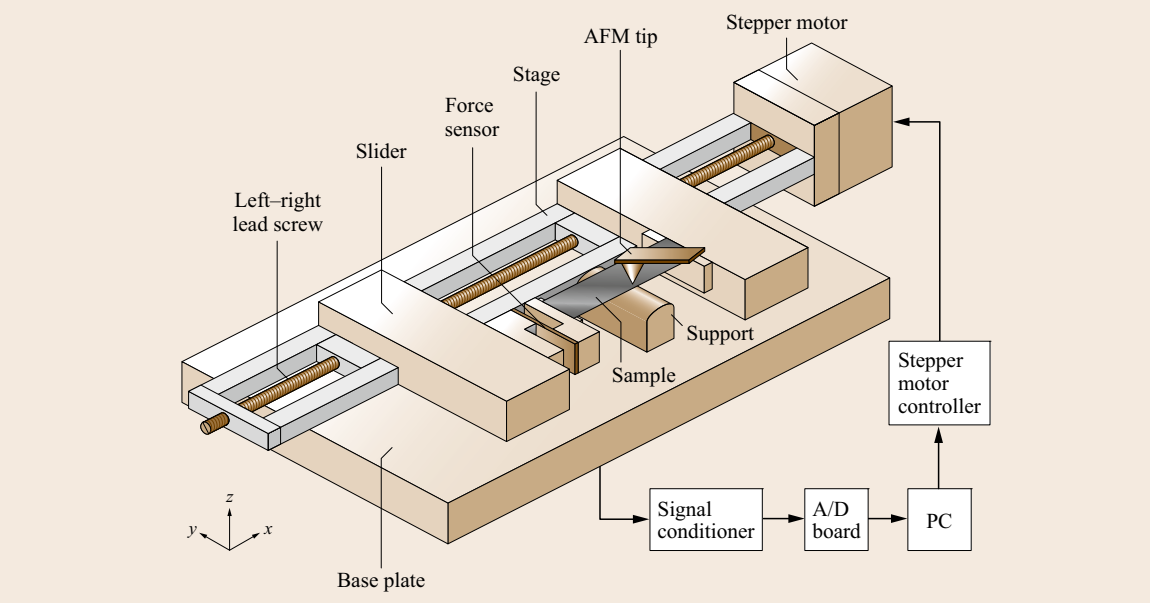


Fig. 27.8 Schematic of the tensile stage to conduct in situ tensile testing of the polymeric films in AFM

increment of 8×10^{-3}% for a sample length of 38 mm. The maximum travel distance was about 100 mm. The resolution of the force sensor was 10 mN with a capacity of 45 N. While stretching, a stress–strain curve was obtained during the experiment to study any correlation between the degree of plastic strain and the propensity of cracks.

27.1.6 Nanoindentation Measurements

For nanoindentation hardness measurements the scan size is set to zero, and then a normal load is applied to make the indents using the diamond tip. During this procedure, the tip is continuously pressed against the sample surface for about two seconds at various indentation loads. The sample surface is scanned before and after the scratching, wear, or indentation to obtain the initial and the final surface topography, at a low normal load of about 0.3 μN using the same diamond tip. An area larger than the indentation region is scanned to observe the indentation marks. Nanohardness is calculated by dividing the indentation load by the projected residual area of the indents [27.61].

Direct imaging of the indent allows one to quantify piling up of ductile material around the indenter. However, it becomes difficult to identify the boundary of the indentation mark with great accuracy. This makes the direct measurement of contact area somewhat inaccurate. A technique with the dual capability of depth-sensing as well as in situ imaging, which is most appropriate in nanomechanical property studies, is used for accurate measurement of hardness with shallow depths [27.2, 14, 60]. This nano/picoindentation system is used for the load–displacement measurement process and for subsequent in situ imaging of the indent, if required. The indentation system, shown in Fig. 27.9, consists of a three-plate transducer with electrostatic actuation hardware used for the direct application of a normal load and a capacitive sensor used for the measurement of vertical displacement. The AFM head is replaced with this transducer assembly while the specimen is mounted on the PZT scanner, which remains stationary during indentation experiments. The transducer consists of a three (Be-Cu)-plate capacitive structure, and the tip is mounted on the center plate. The upper and lower plates serve as drive electrodes, and the load is applied by applying appropriate voltage to the drive

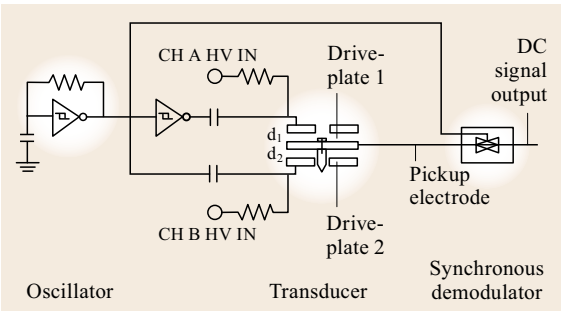


Fig. 27.9 Schematic of a nano/picoindentation system with three-plate transducer with electrostatic actuation hardware and capacitance sensor (after [27.60])

electrodes. Vertical displacement of the tip (indentation depth) is measured by measuring the displacement of the center plate relative to the two outer electrodes using the capacitance technique. Indent area and consequently hardness value can be obtained from the load–displacement data. The Young's modulus of elasticity is obtained from the slope of the unloading curve.

27.1.7 Localized Surface Elasticity and Viscoelasticity Mapping

Localized Surface Elasticity

Indentation experiments provide a single-point measurement of the Young's modulus of elasticity calculated from the slope of the indentation curve during unloading. Localized surface elasticity maps can be obtained using dynamic force microscopy, in which an oscillating tip is scanned over the sample surface in contact under steady and oscillating load. Lower frequency operation modes in the kHz range, such as force modulation mode [27.64, 66] or pulsed force mode [27.87], are well suited for soft samples such as polymers. However, if the tip-sample contact stiffness becomes significantly higher than the cantilever stiffness, the sensitivity of these techniques strongly decreases. In this case, sensitivity of the measurement of stiff materials can be improved by using high-frequency operation modes in the MHz range with a lateral motion, such as acoustic (ultrasonic) force microscopy, referred to as atomic force acoustic microscopy (AFAM), or contact resonance spectroscopy [27.68, 88, 89]. Inclusion of vibration frequencies other than only the first cantilever flexural or torsional resonance frequency also allows additional information to be obtained.

In the negative lift mode force modulation technique, during primary scanning height data is recorded in tapping mode as described earlier. During interleave scanning, the entire cantilever/tip assembly is moved up and down at the force modulation holder's bimorph resonance frequency (about 24 kHz) at some amplitude, here referred to as the force modulation amplitude, and the z-direction feedback control for the sample x–y–z piezo is deactivated (Fig. 27.10a) [27.64, 66, 69]. During this scanning process, height information from the primary scan is used to maintain a constant lift scan height. This eliminates the influence of height on the measured signals during the interleave scan. Lift scan height is the mean tip-to-sample distance between the tip and sample during the interleave scan. The lift scan height is set such that the tip is in constant contact with the sample, i. e., a constant static load is applied. (A higher lift scan height gives a closer mean tip-to-sample distance.) In addition, the tip motion caused by the bimorph vibration results in a modulating periodic force. The sample surface resists the oscillations of the tip to a greater or lesser extent depending upon the sample's stiffness. The computer records amplitude (which is a function of elastic stiffness of the material). Contact analyses can be used to obtain a quantitative measure of localized elasticity of soft surfaces [27.66]. Etched single-crystal silicon cantilevers with integrated tips (Bruker force modulation etched Si probe or FESP) with a radius of 25−50 nm, a stiffness of 1−5 N/m, and a natural frequency of 60−100 kHz, are commonly used for the measurements. Scanning is normally set to a rate of 0.5 Hz along the fast axis.

In the AFAM technique [27.68, 88, 89], the cantilever/tip assembly is moved either in the normal or lateral mode, and the contact stiffness is evaluated by comparing the resonance frequency of the cantilever in contact with the sample surface to those of the free vibrations of the cantilever. Several free resonance frequencies are measured. On the basis of the shift of the measured frequencies, the contact stiffness is determined by solving the characteristic equation for the tip vibrating in contact with the sample surface. The elastic modulus is calculated from contact stiffness using Hertz analysis for a spherical tip indenting a plane. Contact stiffness is equal to 8 × contact radius × reduced shear modulus in shear mode.

In the lateral mode using the AFAM technique, the sample is glued on cylindrical pieces of aluminum which serves as ultrasonic delay lines coupled to an ultrasonic shear wave transducer (Fig. 27.10b) [27.34, 68, 89]. The transducer is driven with frequency sweeps to generate in-plane lateral sample surface vibrations. These couple to the cantilever via the tip–sample contact. To measure torsional vibrations of the cantilever at frequencies up to 3 MHz, the original electronic circuit of the lateral channel of the AFM (using a lowpass filter with limited bandwidth to a few hundred kHz) was replaced by a high-speed scheme which bypasses the lowpass filter. The high-frequency signal was fed to a lock-in amplifier, digitized using a fast analog/digital (A/D) card, and fed into a broadband amplifier, followed by an root mean square (RMS)-to-DC converter and read by a computer. Etched single-crystal silicon cantilevers (normal stiffness of 3.8−40 N/m) with integrated tips are used.

Viscoelastic Mapping

Another form of dynamic force microscopy, phase contrast microscopy, is used to detect the contrast in viscoelastic (viscous energy dissipation) properties of different materials across the surface [27.48, 65, 69, 70, 72, 90, 91]. In these techniques, both deflection amplitude and phase angle contrasts are measured, which are measures of the relative stiffness and viscoelastic

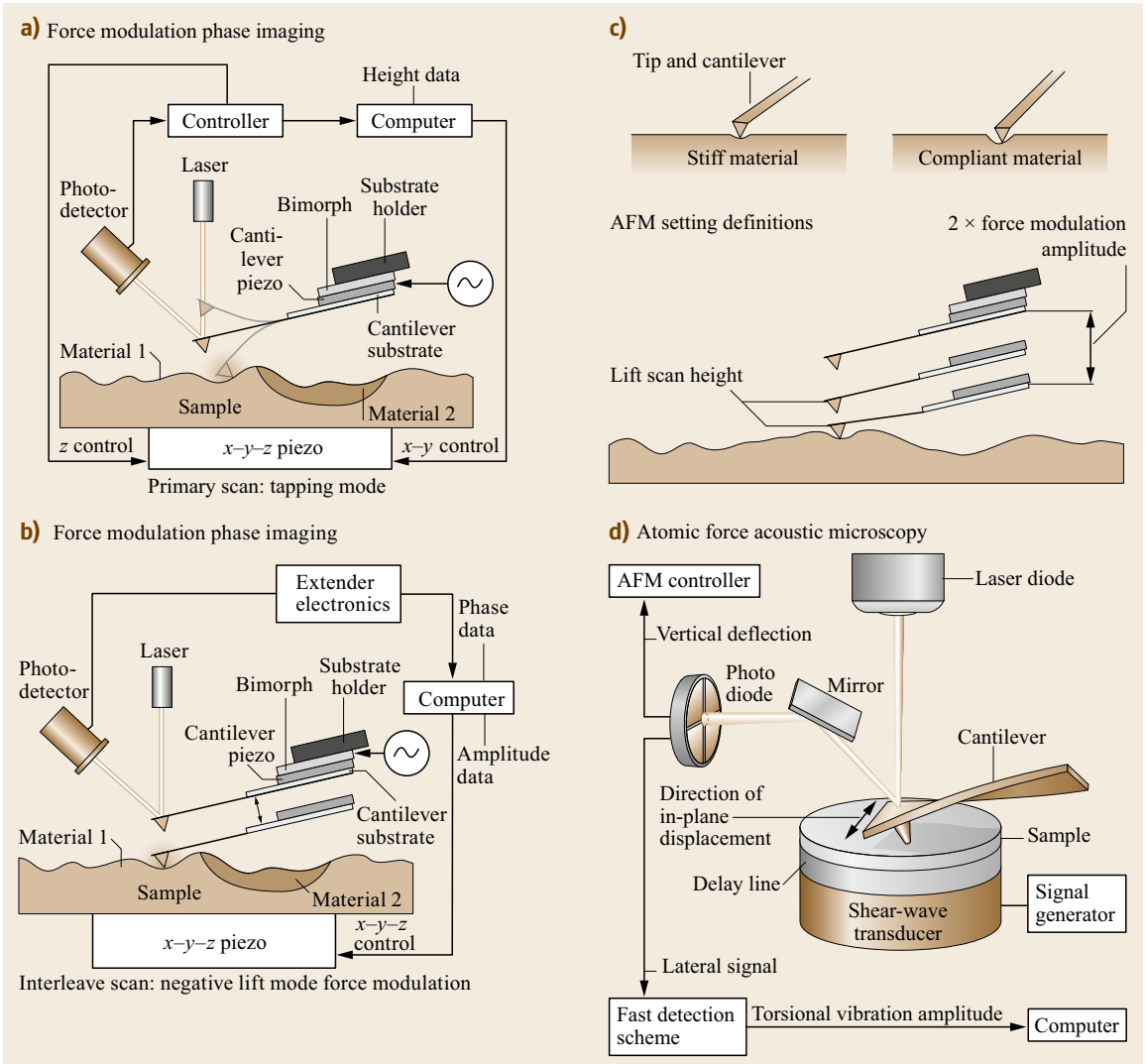


Fig. 27.10 **(a–c)** Schematic of force modulation mode used to obtain amplitude (stiffness) and definitions of force modulation amplitude and lift scan height. During primary scanning, height data is recorded in tapping mode. During interleave scanning, the entire cantilever/tip assembly is vibrated at the bimorph's resonance frequency and the z-direction feedback control for the sample x–y–z piezo is deactivated. During this scanning, height information from the primary scan is used to maintain a constant lift scan height. The computer records amplitude (which is a function of material stiffness) during the interleave scan. **(d)** Schematic of an AFM incorporating shear wave transducer which generates in-plane lateral sample surface vibrations. Because of the forces between the tip and the surface, torsional vibrations of the cantilever are excited (after [27.34]). The shift in contact resonance frequency is a measure of contact stiffness

properties, respectively. Two phase measurement techniques – tapping mode and torsional resonance (TR) mode – have been developed. We describe them next.

In the tapping mode (TM) technique, as described earlier, the cantilever/tip assembly is sinusoidally vibrated at its resonant frequency, and the sample x–y–z piezo is adjusted using feedback control in the z-direction to maintain a constant setpoint (Fig. 27.3) [27.69, 70]. The feedback signal to the z-direction sample piezo (to keep the setpoint constant) is a measure of surface roughness. The extender electronics is used to measure the phase angle lag between the cantilever piezo drive signal and the cantilever response during sample engagement. As illustrated in Fig. 27.3, the phase angle lag (at least partially) is a function of the viscoelastic properties of the sample material.

A range of tapping amplitudes and setpoints can be used for measurements. A commercially etched single-crystal silicon tip (DI TESP), with a radius of 5–10 nm, a stiffness of 20–100 N/m, and a natural frequency of 350–400 kHz, is normally used for tapping mode. Scanning is normally set to a rate of 1 Hz along the fast axis.

In the torsional mode (TR mode), a tip is vibrated at high frequency at the resonance frequency of the cantilever beam. An etched single-crystal silicon cantilever with integrated tip (DI FESP) with a radius of about 5–10 nm, normal stiffness of 1–5 N/m, torsional stiffness of about 30 times normal stiffness, and torsional natural frequency of 800 kHz is normally used. A major difference between the TM and TR modes is the directionality of the applied oscillation – a normal (compressive) amplitude exerted for the TM and a torsional amplitude for the TR mode. The TR mode is expected to provide good contrast in the tribological and mechanical properties of the near-surface region as compared to the TM. Two of the reasons are as follows. (1) In the TM, the interaction is dominated by the vertical properties of the sample, so the tip spends a small fraction of its time in near-field interaction with the sample. Furthermore, the distance between the tip and the sample changes during the measurements, which changes interaction time and forces, and affects measured data. In the TR mode, the distance remains nearly constant. (2) The lateral stiffness of a cantilever is typically about two orders of magnitude larger than the normal (flexural) stiffness. Therefore, in the TM, if the sample is relatively rigid, much of the deformation occurs in the cantilever beam, whereas in the TR mode, much of the deformation occurs in the sample. A few comments on the special applications of the TR mode are made next. Since most of the deformation occurs in the sample, the TR mode can be used to measure stiff and hard samples. Furthermore, properties of thin films can be measured more readily with the TR mode. For both the TM and TR modes, if the cantilever is driven to vibrate at frequencies above resonance, it would have less motion (high apparent stiffness), leading to higher sample deformation and better contrast. It should be further noted that the TM exerts a compressive force, whereas the TR mode exerts torsional force, therefore normal and shear properties are measured in the TM and TR modes, respectively.

In the TR mode, the torsional vibration of the cantilever beam is achieved using a specially designed cantilever holder. It is equipped with a piezo system mounted in a cantilever holder, in which two piezos vibrate out-of-phase with respect to each other. A tuning process prior to scanning is used to select the torsional vibration frequency. The piezo system excites torsional vibration at the cantilever's resonance frequency. The torsional vibration amplitude of the tip (TR amplitude) is detected by the lateral segments of the split-diode photodetector (Fig. 27.11) [27.71]. The TR mode measures surface roughness and phase angle as follows. During the measurement, the cantilever/tip assembly is

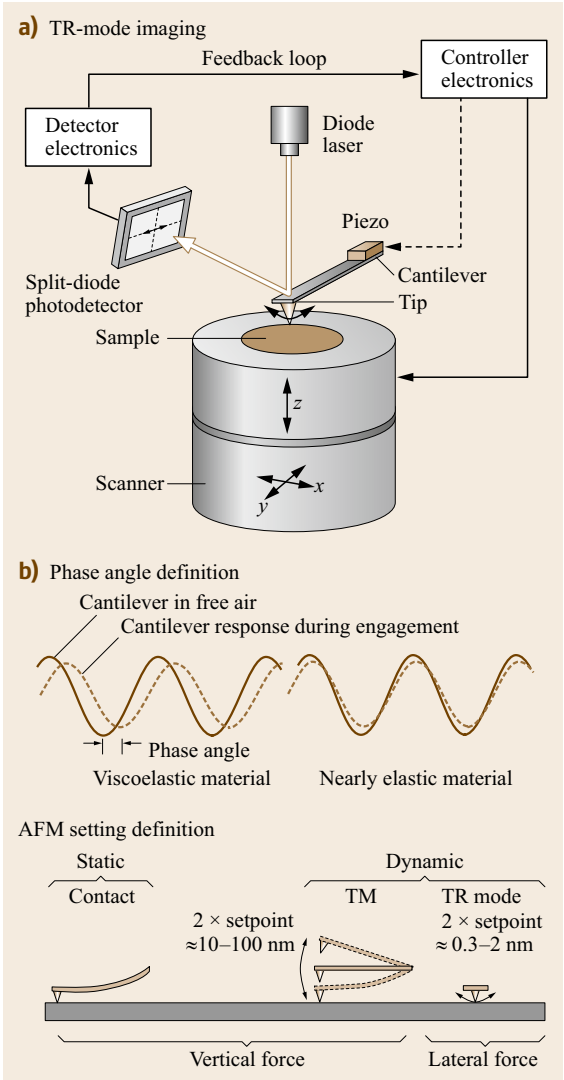


Fig. 27.11a,b Schematic of torsional resonance mode shown *at the top* (**a**). Two examples of the phase angle response are shown *in the middle* (**b**). One is for materials exhibiting viscoelastic and the other nearly elastic properties. Three AFM settings are compared *at the bottom*: contact, tapping mode (TM), and TR modes. The TR mode is a dynamic approach with a laterally vibrating cantilever tip that can interact with the surface more intensively than other modes. Therefore, more detailed near-surface information is available

first vibrated at its resonance at some amplitude dependent upon the excitation voltage, before the tip engages the sample. Next, the tip engages the sample at some setpoint. A feedback system coupled to a piezo stage is used to keep a constant TR amplitude during scanning. This is done by controlling the vertical position of the sample using a piezo moving in the z-direction, which changes the degree of tip interaction. The displacement of the sample z piezo gives a roughness image of the sample. A phase angle image can be obtained by measuring the phase lag of the cantilever vibration response in the torsional mode during engagement with respect to the cantilever vibration response in free air before engagement. The control feedback of the TR mode is similar to that of tapping, except that the torsional resonance amplitude replaces flexural resonance amplitude [27.71].

Chen and *Bhushan* [27.72] used a variation to the approach just described (referred to as mode I here). They performed measurements at constant normal cantilever deflection (constant load; mode II) instead of using the constant setpoint in the *Kasai* et al. [27.71] approach. Their approach overcomes the meniscus adhesion problem present in mode I and reveals true surface properties.

Song and *Bhushan* [27.92] presented a forced torsional vibration model for a tip-cantilever assembly under viscoelastic tip–sample interaction. This model provides a relationship of torsional amplitude and phase shift with lateral contact stiffness and viscosity which can be used to extract in-plane interfacial mechanical properties.

Various operating modes of AFM used for surface roughness, localized surface elasticity, and viscoelastic mapping and friction force measurements (to be discussed later) are summarized in Table 27.2.

27.1.8 Boundary Lubrication Measurements

To study nanoscale boundary lubrication properties, adhesive forces are measured in the force calibration mode, as previously described. The adhesive forces are also calculated from the horizontal intercept of friction versus normal load curves at a zero value of friction force. For friction measurements, the samples are typically scanned using an Si tip over an area of $2 \times 2\,\mu m^2$ at the normal load ranging from 5 to 130 nN. The samples are generally scanned with a scan rate of 0.5 Hz resulting in a scanning speed of $2\,\mu m/s$. Velocity effects on friction are studied by changing the scan frequency from 0.1 to 60 Hz, while the scan size is maintained at $2 \times 2\,\mu m^2$, which allows velocity to vary from 0.4 to $240\,\mu m/s$. To study the durability properties, the friction force and coefficient of friction are monitored during scanning at a normal load of 70 nN and a scanning speed of $0.8\,\mu m/s$, for a desired number of cycles [27.42, 43, 47].

27.1.9 AFM Tip Wear

AFM tips wear during use. Wear can result in an increase in tip radius and adhesion between the tip and the sample, thus reducing the image resolution and introducing artifacts. The tip geometry can be measured by AFM imaging using an ultrasharp geometry (sharper than the AFM tip), known as a tip characterizer, and analyzing the data for tip reconstruction using a tip de-

Table 27.2 Summary of various operating modes of AFM for surface roughness, stiffness, phase angle, and friction

Operating mode	Direction of cantilever vibration	Vibration frequency of cantilever (kHz)	Vibration amplitude (nm)	Feedback control	Data obtained
Contact	N/A			Constant normal load	Surface height, friction
Tapping	Vertical	350–400	10–100	Setpoint (constant tip amplitude)	Surface height, phase angle (normal viscoelasticity)
Force modulation	Vertical	10–20 (bimorph)	10–100	Constant normal load	Surface height, amplitude (normal stiffness)
Lateral	Lateral (AFAM)	100–3000 (sample)	≈ 5 (sample)	Constant normal load	Shift in contact resonance (normal stiffness, friction)
TR mode I	Torsional	≈ 800	0.3–2	Setpoint (constant tip amplitude)	Surface height, phase angle (lateral viscoelasticity)
TR mode II	Torsional	≈ 800	0.3–2	Constant normal load	Surface height, amplitude and phase angle (lateral stiffness and lateral viscoelasticity)
TR mode III	Torsional	> 800 in contact	0.3–2	Constant normal load	Shift in contact resonance (friction)

convolution algorithm [27.93]. A grating sample with an array of sharp tips with a radius less than 10 nm is commonly used.

Tao and *Bhushan* [27.93] developed fluoropolymer coatings deposited on the silicon probe to reduce their tip wear.

27.2 Surface Imaging, Friction, and Adhesion

27.2.1 Atomic-Scale Imaging and Friction

Surface height imaging down to atomic resolution of electrically conducing surfaces can be carried out using an STM. An AFM can also be used for surface height imaging and roughness characterization down to the nanoscale. Figure 27.12 shows a sequence of STM images at various scan sizes of solvent-deposited C_{60} film on 200 nm thick gold-coated freshly cleaved mica [27.94]. The film consists of clusters of C_{60} molecules 8 nm in diameter. The C_{60} molecules within a cluster appear to pack into a hexagonal array with a spacing of about 1 nm, however, they do not follow any long-range order. The measured cage diameter of the C_{60} molecule is about 0.7 nm, very close to the projected diameter of 0.71 nm.

In an AFM measurement during surface imaging, the tip comes into intimate contact with the sample surface and leads to surface deformations with finite tip–sample contact area (typically few atoms). The finite size of the contact area prevents the imaging of individual point defects, and only the periodicity of the atomic lattice can be imaged. Figure 27.13a shows the topography image of a freshly cleaved surface of highly oriented pyrolytic graphite (HOPG) [27.26]. The periodicity of the graphite is clearly observed.

To study the friction mechanisms on an atomic scale, a freshly cleaved HOPG was studied by *Mate* et al. [27.23] and *Ruan* and *Bhushan* [27.26]. Figure 27.14a shows the atomic-scale friction force map (raw data) and Fig. 27.13a shows the friction force maps (after two-dimensional (2-D) spectrum filtering with high-frequency noise truncated) [27.26]. Figure 27.14a also shows a line plot of friction force profiles along some crystallographic direction. The actual shape of the friction profile depends upon the spatial location of the axis of tip motion. Note that a portion of atomic-scale lateral force is conservative. *Mate* et al. [27.23] and *Ruan* and *Bhushan* [27.26] reported that the average friction force linearly increased with normal load and was reversible with load. Friction profiles were similar during the sliding of the tip in either direction.

During scanning, the tip moves discontinuously over the sample surface and jumps with discrete steps from one potential minimum (well) to the next. This leads to a saw-tooth-like pattern for the lateral motion (force) with a periodicity of the lattice constant. This motion is called the stick-slip movement of the tip [27.2, 23, 26]. The observed friction force includes two components – conservative and periodic, and non-

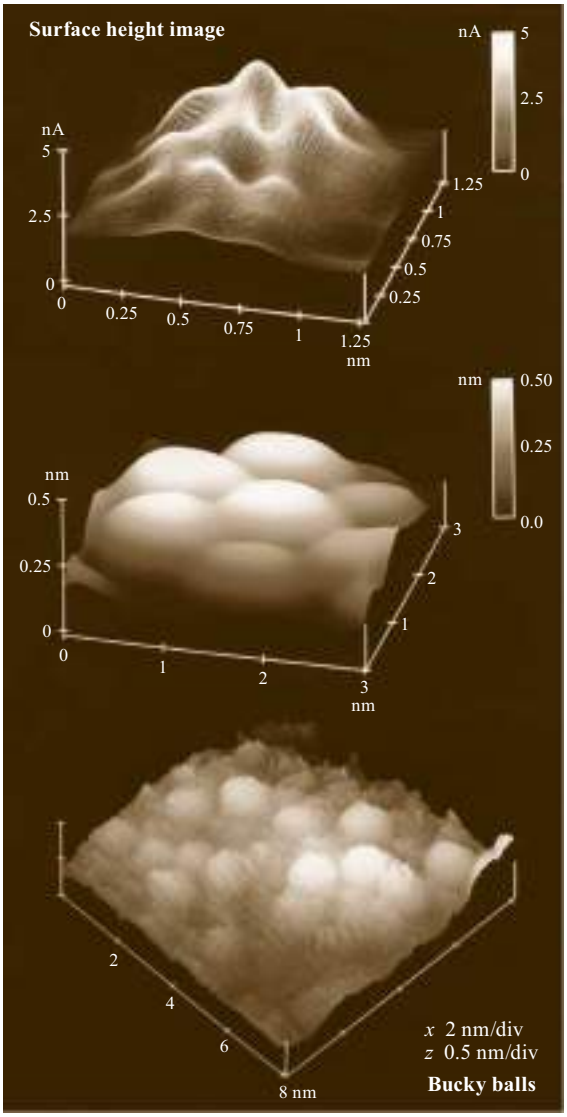


Fig. 27.12 STM images of solvent-deposited C_{60} film on a gold-coated freshly cleaved mica at various scan sizes (after [27.95])

conservative and constant. If the relative motion of the sample and tip were simply that of two rigid collections of atoms, the effective force would be a conservative force oscillating about zero. Slow reversible elastic deformation would also contribute to conservative force. The origin of the nonconservative direction-dependent force component would be phonon generation, viscous dissipation, or plastic deformation.

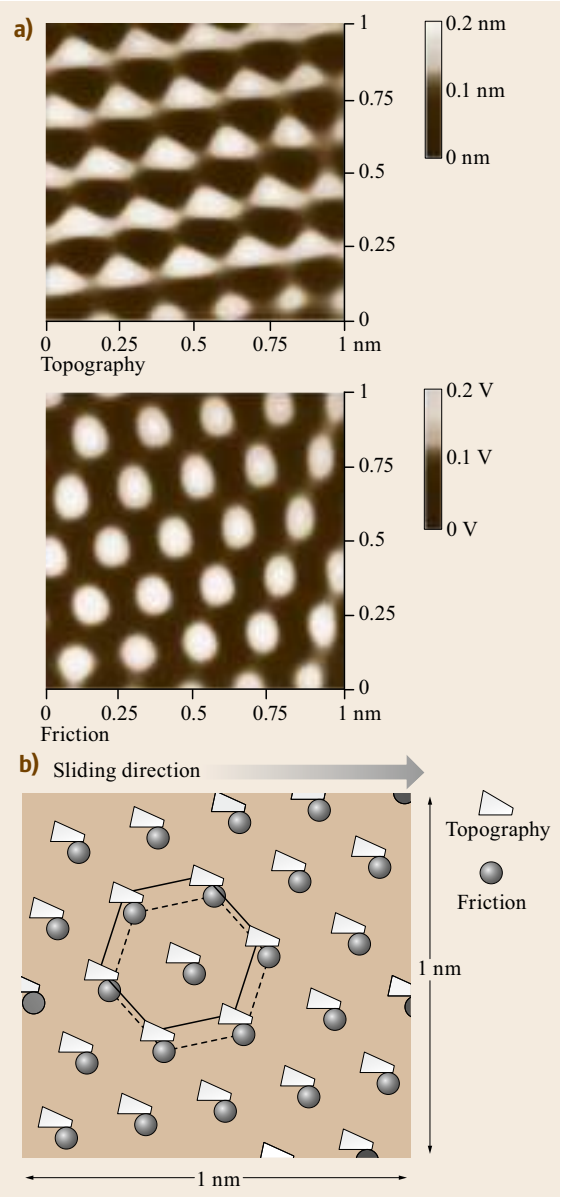


Fig. 27.13 **(a)** Gray-scale plots of surface topography and friction force maps (2-D spectrum filtered), measured simultaneously, of a $1\,\text{nm} \times 1\,\text{nm}$ area of freshly cleaved HOPG, showing the atomic-scale variation of topography and friction, and **(b)** schematic of superimposed topography and friction maps from **(a)**; the *symbols* correspond to maxima. Note the spatial shift between the two plots (after [27.26])

Stick-slip on the atomic scale, discussed above, is the result of the energy barrier required to be overcome to jump over the atomic corrugations on the sample surface. It corresponds to the energy required for the jump of the tip from a stable equilibrium position on the surface to a neighboring position. The perfect atomic regularity of the surface guarantees the periodicity of the lateral force signal, independent of the actual atomic structure of the tip apex. A few atoms (based on the magnitude of the friction force, less than ten) on a tip sliding over an array of atoms on the sample are expected to undergo the stick-slip. For simplicity, Fig. 27.14b shows a simplified model for one atom on a tip with a one-dimensional spring mass system. As the sample surface slides against the AFM tip, the tip remains *stuck* initially until it can overcome the energy (potential) barrier, which is illustrated by a sinusoidal interaction potential as experienced by the tip. After some motion, there is enough energy stored in the spring which leads to *slip* into the neighboring stable equilibrium position. During the slip and before attaining stable equilibrium, stored energy is converted into vibrational energy of the surface atoms in the range of 10^{13} Hz (phonon generation) and decays within the range of 10^{-11} s into heat. (A wave of atoms vibrating in concert are termed a phonon). The stick-slip phenomenon, resulting from irreversible atomic jumps, can be theoretically modeled with classical mechanical models [27.96, 97]. The Tomanek-Zhong-Thomas model [27.97] is the starting point for determining friction force during atomic-scale stick-slip. The AFM model describes the total potential as the sum of the potential acting on the tip due to interaction with the sample and the elastic energy stored in the cantilever. Thermally activated stick-slip behavior can explain the velocity effects on friction, to be presented later.

Finally, based on Fig. 27.13a, the atomic-scale friction force of HOPG exhibited the same periodicity as that of the corresponding topography, but the peaks in friction and those in the topography are displaced relative to each other (Fig. 27.13b). A Fourier expansion of the interatomic potential was used by *Ruan* and *Bhushan* [27.26] to calculate the conservative interatomic forces between the atoms of the FFM tip and those of the graphite surface. Maxima in the interatomic forces in the normal and lateral directions do not occur at the same location, which explains the observed shift between the peaks in the lateral force and those in the corresponding topography.

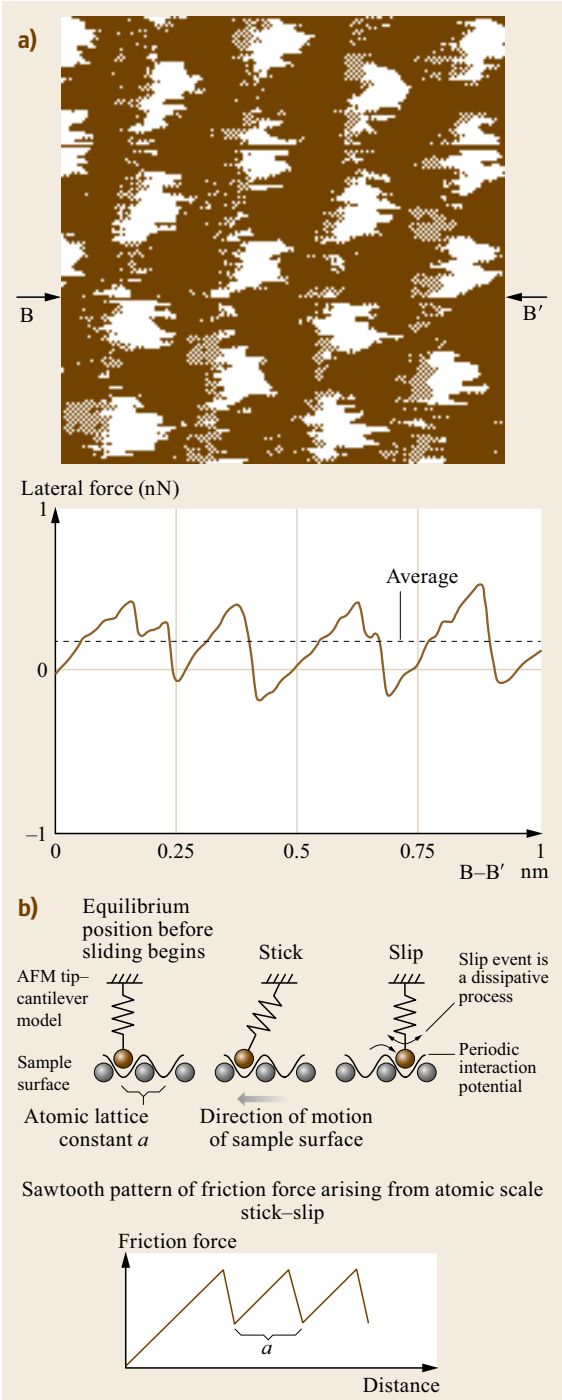


Fig. 27.14 **(a)** Gray-scale plot of a friction force map (raw data) of a $1 \times 1\,\mathrm{nm}^2$ area of freshly cleaved HOPG, showing atomic-scale variation of friction force. High points are shown by a *lighter color*. Also shown is a line plot of the friction force profile along the line indicated by the *arrows*. The normal load was 25 nN and the cantilever normal stiffness was 0.4 N/m (after [27.26]). **(b)** Schematic of a model for a tip atom sliding on an atomically flat periodic surface. The schematic shows the tip jumping from one potential minimum to another, resulting in stick-slip behavior ◄

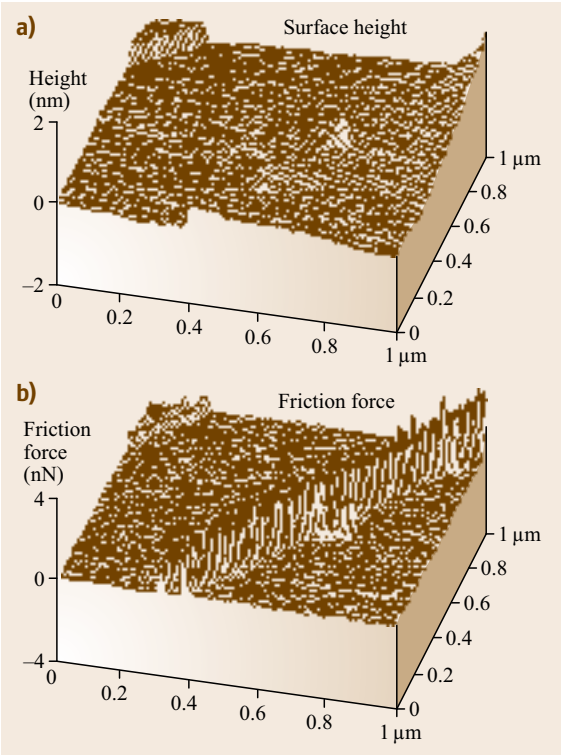


Fig. 27.15 **(a)** Surface roughness and **(b)** friction force maps at a normal load of 42 nN of a freshly cleaved HOPG surface against an Si_3N_4 FFM tip. Friction in the *line-shaped region* is over an order of magnitude larger than the smooth areas (after [27.26])

27.2.2 Microscale Friction

Local variations in the microscale friction of cleaved graphite are observed (Fig. 27.15). Microscale friction is defined as the friction measured with a scan size equal to or larger than $1\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$. These arise from structural changes that occur during the cleaving process [27.27]. The cleaved HOPG surface is largely atomically smooth but exhibits line-shaped regions in which the coefficient of friction is more than an order of magnitude larger. Transmission electron microscopy indicates that the line-shaped regions consist of graphite planes of different orientation, as well as of amorphous carbon. Differences in friction have also been observed for multiphase ceramic materials [27.56]. Figure 27.16 shows the surface roughness and friction

force maps of Al_2O_3-TiC (70–30 wt%). TiC grains have a Knoop hardness of about 2800 kg/mm^2 and Al_2O_3 has 2100 kg/mm^2, therefore, TiC grains do not polish as much and result in a slightly higher elevation (about 2–3 nm higher than that of Al_2O_3 grains). TiC grains exhibit higher friction force than Al_2O_3 grains. The coefficients of friction of TiC and Al_2O_3 grains are 0.034 and 0.026, respectively, and the coefficient of friction of Al_2O_3-TiC composite is 0.03. Local variation in friction force also arises from the scratches present on the Al_2O_3-TiC surface. *Meyer* et al. [27.98] also used FFM to measure structural variations of organic mono- and multilayer films. All of these measurements suggest that the FFM can be used for structural mapping of the surfaces. FFM measurements can also be used to map chemical variations, as indicated by the use of the FFM with a modified probe tip to map the spatial arrangement of chemical functional groups in mixed organic monolayer films [27.99]. Here, sample regions that had stronger interactions with the functionalized probe tip exhibited greater friction.

Local variations in the microscale friction of nominally rough, homogeneous material surfaces can be significant, and are seen to depend on the local surface slope rather than the surface height distribution (Fig. 27.17). This dependence was first reported by

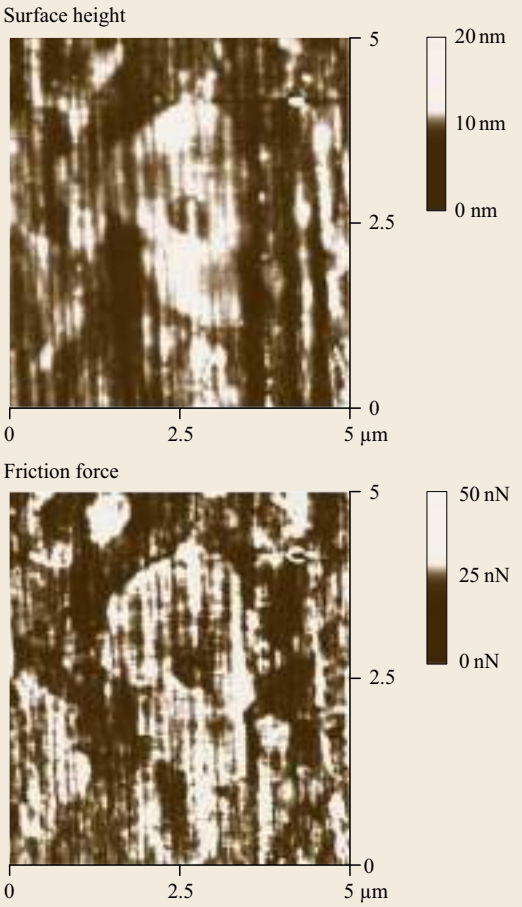


Fig. 27.16 Gray-scale surface roughness ($\sigma = 0.80$ nm) and friction force maps (mean = 7.0 nN, $\sigma = 0.90$ nN) for Al_2O_3-TiC (70–30 wt%) at a normal load of 138 nN (after [27.56])

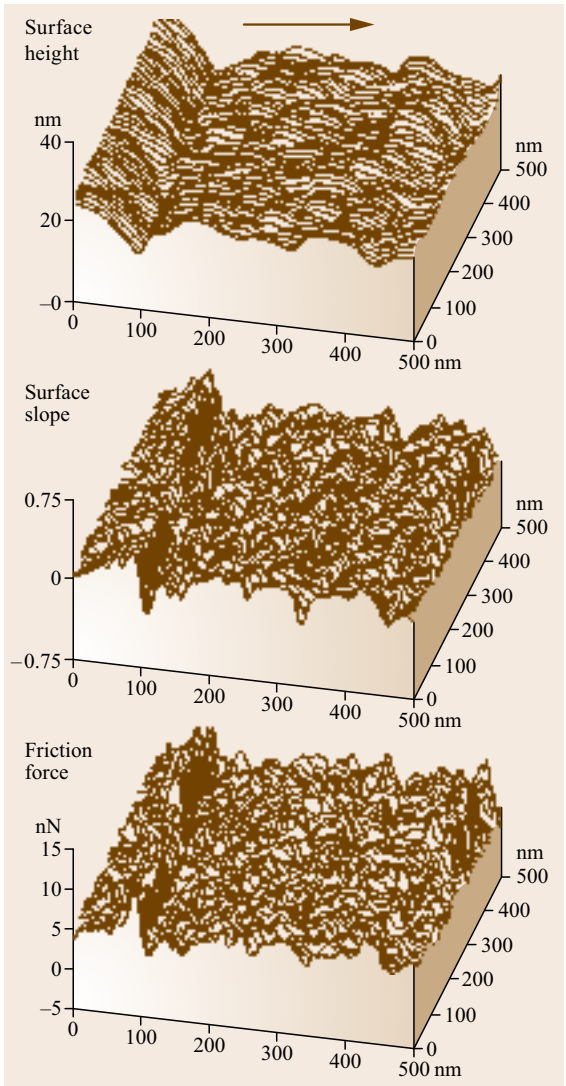


Fig. 27.17 Surface roughness map ($\sigma = 4.4$ nm), surface slope map taken in the sample-sliding direction (the *horizontal axis*; mean = 0.023, $\sigma = 0.197$), and friction force map (mean = 6.2 nN, $\sigma = 2.1$ nN) for a lubricated thin-film magnetic rigid disk for a normal load of 160 nN (after [27.28])

Bhushan and *Ruan* [27.24], *Bhushan* et al. [27.28] and *Bhushan* [27.79] and later discussed in more detail by *Koinkar* and *Bhushan* [27.100] and *Sundararajan* and *Bhushan* [27.101]. In order to elegantly show any correlation between local values of friction and surface roughness, surface roughness and friction force maps of a gold-coated ruler with somewhat rectangular grids and a silicon grid with square pits were obtained (Fig. 27.18). Figures 27.17 and 27.18 show the surface roughness map, the slopes of the roughness map taken along the sliding direction (surface slope map), and the friction force map for various samples. There is a strong correlation between the surface slopes and friction forces. For example, in Fig. 27.18, the friction force is high locally at the edge of the grids and pits with a positive slope and is low at the edges with a negative slope.

We now examine the mechanism of microscale friction, which may explain the resemblance between the slope of surface roughness maps and the corresponding friction force maps [27.2, 7, 13, 14, 24, 26–28, 101, 102]. There are three dominant mechanisms of friction; adhesive, ratchet, and plowing [27.103, 104]. As a first order, we may assume these to be additive. The adhesive mechanism cannot explain the local variation in friction. Next, we consider the ratchet mechanism. We consider a small tip sliding over an asperity making an angle θ with the horizontal plane (Fig. 27.19). The normal force W (normal to the general surface) applied by the tip to the sample surface is constant. The friction force F on the sample would be a constant for a smooth surface if the friction mechanism does not change. For the rough surface shown in Fig. 27.19, if the adhesive mechanism does not change during sliding, the local value of the coefficient of friction remains constant,

$$\mu_0 = \frac{S}{N} , \tag{27.1}$$

where S is the local friction force and N is the local normal force. However, the friction and normal forces are measured with respect to global horizontal and normal axes, respectively. The measured local coefficient of friction μ_1 in the ascending part is

$$\mu_1 = \frac{F}{W} = \frac{\mu_0 + \tan\theta}{1 - \mu_0 \tan\theta} \approx \mu_0 + \tan\theta, \quad \text{for small } \mu_0 \tan\theta \tag{27.2}$$

indicating that in the ascending part of the asperity one may simply add the friction force and the asperity slope to one another. Similarly, on the right-hand side (descending part) of the asperity,

$$\mu_2 = \frac{\mu_0 - \tan\theta}{1 + \mu_0 \tan\theta} \approx \mu_0 - \tan\theta, \quad \text{for small } \mu_0 \tan\theta . \tag{27.3}$$

For a symmetrical asperity, the average coefficient of friction experienced by the FFM tip traveling across the whole asperity is

$$\begin{aligned} \mu_{\text{ave}} &= \frac{\mu_1 + \mu_2}{2} \\ &= \mu_0 \frac{1 + \tan^2\theta}{1 - \mu_0^2 \tan^2\theta} \\ &\approx \mu_0 (1 + \tan^2\theta), \quad \text{for small } \mu_0 \tan\theta . \end{aligned} \tag{27.4}$$

Finally, we consider the plowing component of friction with the tip sliding in either direction, which is [27.14, 79]

$$\mu_{\text{p}} \approx \tan\theta . \tag{27.5}$$

Because in FFM measurements we notice little damage of the sample surface, the contribution from plowing is expected to be small, and the ratchet mechanism is believed to be the dominant mechanism for the local variations in the friction force map. With the tip sliding over the leading (ascending) edge of an asperity, the surface slope is positive; it is negative during sliding over the trailing (descending) edge of an asperity. Thus, measured friction is high at the leading edge of asperities and low at the trailing edge. In addition to the slope effect, the collision of the tip when encountering an asperity with a positive slope produces additional torsion of the cantilever beam leading to higher measured friction force. When encountering an asperity with the same negative slope, however, there is no collision effect and hence no effect on torsion. This effect also contributes to the difference in friction forces when the tip scans up and down on the same topography feature. The ratchet mechanism and the collision effects thus semiquantitatively explain the correlation between the slopes of the roughness maps and friction force maps observed in Figs. 27.17 and 27.18. We note that in the ratchet mechanism, the FFM tip is assumed to be small compared to the size of asperities. This is valid since the typical radius of curvature of the tips is about 10–50 nm. The radii of curvature of the asperities of the samples measured here (the asperities that produce most of the friction variation) are found to typically be about 100–200 nm, which is larger than that of the FFM tip [27.105]. It is important to note that the measured local values

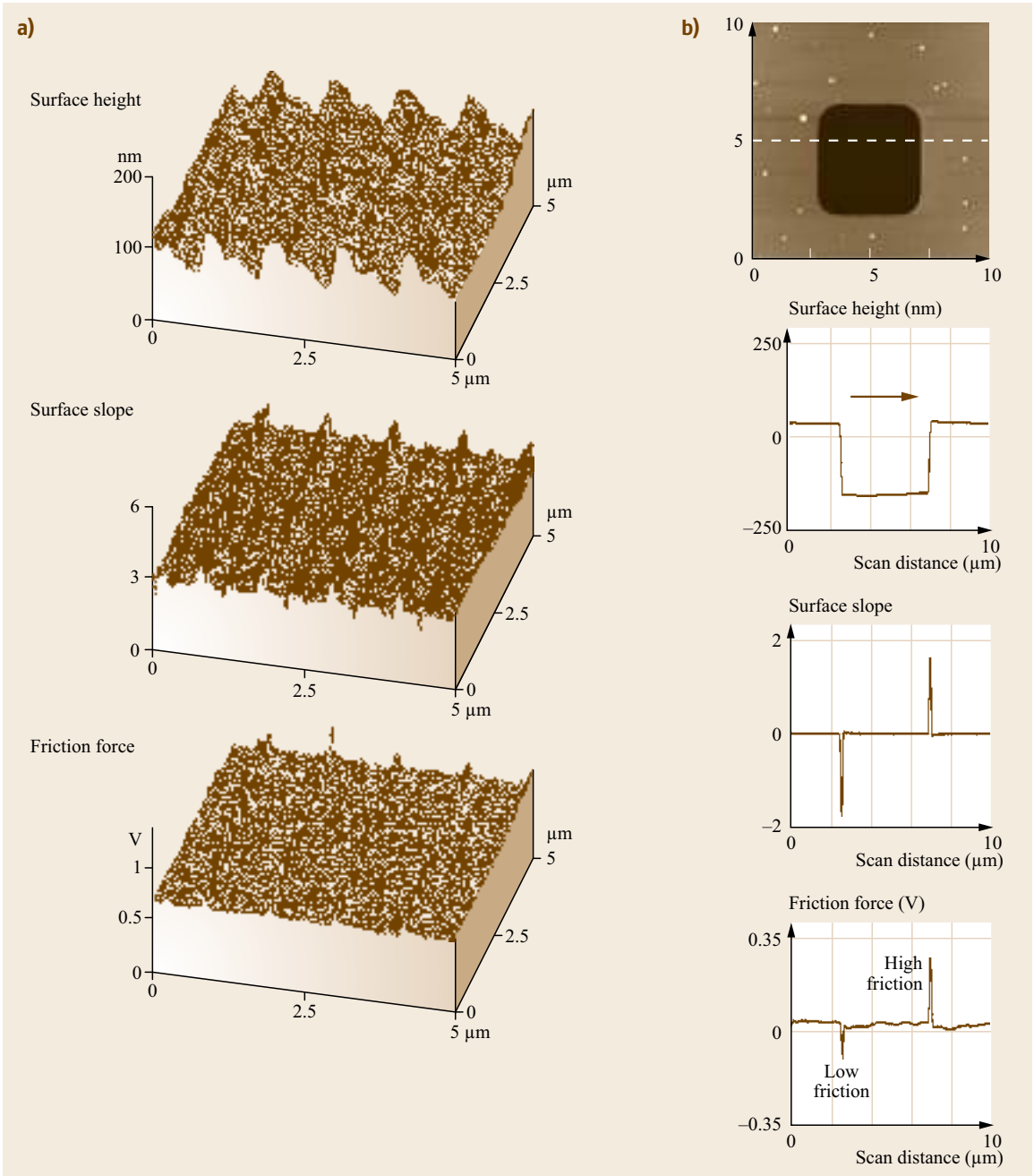


Fig. 27.18a,b Surface roughness map, surface slope map taken in the sample-sliding direction (*the horizontal axis*), and friction force map for **(a)** a gold-coated ruler (with somewhat rectangular grids with a pitch of 1 µm and a ruling step height of about 70 nm) at a normal load of 25 nN and **(b)** a silicon grid (with 5 µm square pits of depth 180 nm and a pitch of 10 µm) (after [27.101])

of friction and normal forces are measured with respect to global (and not local) horizontal and vertical axes, which are believed to be relevant in applications.

27.2.3 Directionality Effect on Microfriction

During friction measurements, the friction force data from both the forward (trace) and backward (re-

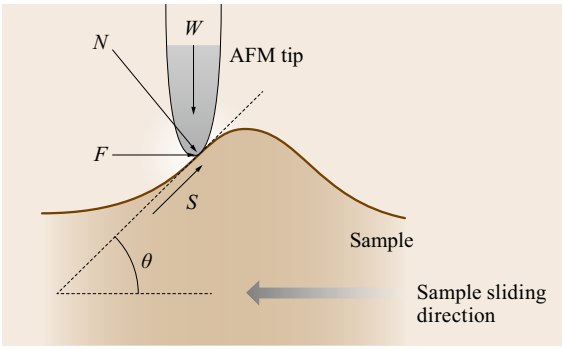


Fig. 27.19 Schematic illustration showing the effect of an asperity (making an angle θ with the horizontal plane) on the surface in contact with the tip on local friction in the presence of an adhesive friction mechanism. W and F are the normal and friction forces, respectively, and S and N are the force components along and perpendicular to the local surface of the sample at the contact point, respectively

trace) scans are useful in understanding the origins of the observed friction forces. Magnitudes of material-induced effects are independent of the scanning direction whereas topography-induced effects are different between forward and backward scanning directions. Since the sign of the friction force changes as the scanning direction is reversed (because of the reversal of torque applied to the end of the tip), addition of the friction force data of the forward and backward scan eliminates the material-induced effects while topography-induced effects still remain. Subtraction of the data between forward and backward scans does not eliminate either effect (Fig. 27.20) [27.101].

Owing to the reversal of the sign of the retrace (R) friction force with respect to the trace (T) data, the friction force variations due to topography are in the same direction (peaks in trace correspond to peaks in retrace). However, the magnitudes of the peaks in trace and retrace at a given location are different. An increase in the friction force experienced by the tip when scanning up a sharp change in topography is more than the decrease in the friction force experienced when scanning down the same topography change, partly because of collision effects discussed earlier. Asperities on engineering surfaces are asymmetrical, which also affect the magnitude of friction force in the two directions. Asymmetry in tip shape may also have an effect on the directionality effect of friction. We will note later that the magnitude of surface slopes are virtually identical, therefore, the tip shape asymmetry should not have much effect.

Figure 27.21 shows the surface height and friction force data for a gold ruler and a silicon grid in the trace and retrace directions. Subtraction of two friction data

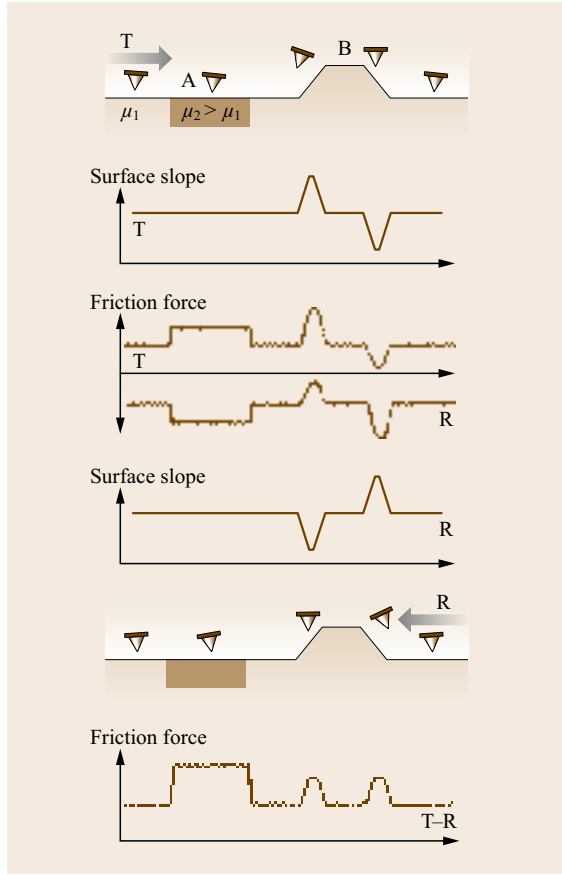


Fig. 27.20 Schematic of friction forces expected when a tip traverses a sample that is composed of different materials and sharp changes in topography. A schematic of surface slope is also shown

yields a residual peak because of the differences in the magnitudes of friction forces in the two directions. This effect is observed at all locations of significant changes in topography.

In order to facilitate comparison of the directionality effect on friction, it is important to take into account the sign change of the surface slope and friction force in the trace and retrace directions. Figure 27.22 shows surface height, surface slope, and friction force data for the two samples in the trace and retrace directions. The correlations between surface slope and friction forces are clear. The third column in the figures shows retrace slope and friction data with an inverted sign (-retrace). Now we can compare trace data with -retrace data. It is clear that the friction experienced by the tip is dependent upon the scanning direction because of surface topography. In addition to the effect of topographical changes discussed earlier, during surface-finishing processes, material can be transferred preferentially onto

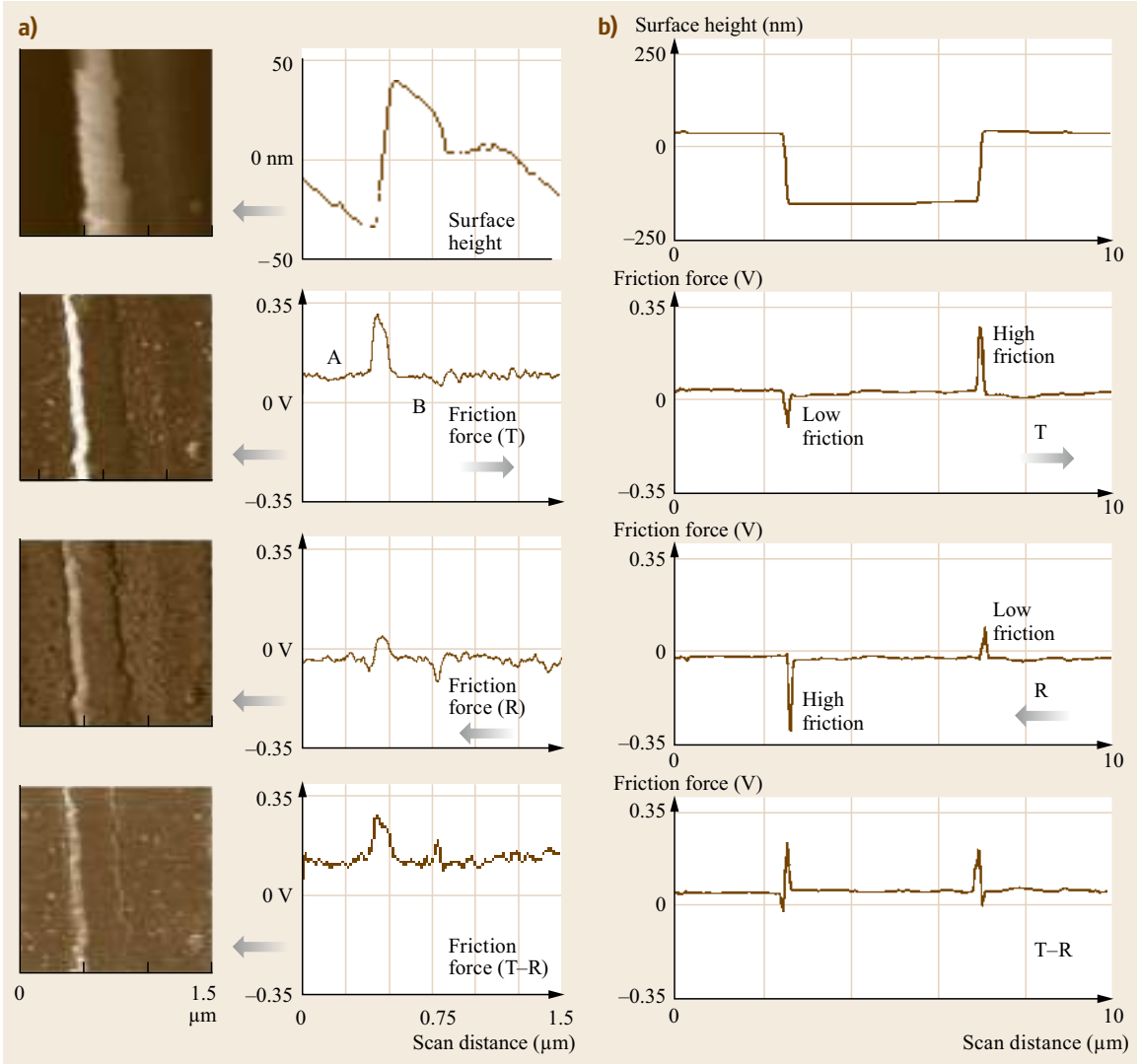


Fig. 27.21 **(a)** Gray-scale images and two-dimensional profiles of surface height and friction forces across a single ruling of the gold-coated ruling, and **(b)** two-dimensional profiles of surface height and friction forces across a silicon grid pit. Friction force data in trace and retrace directions, and subtracted force data are presented (after [27.101])

one side of the asperities, which also causes asymmetry and direction dependence. Reduction in local variations and in the directionality of friction properties requires careful optimization of surface roughness distributions and of surface-finishing processes.

The directionality as a result of a surface asperities effect will be also manifested in macroscopic friction data, i. e., the coefficient of friction may be different in one sliding direction than that in the other direction. The asymmetrical shape of the asperities accentuates this effect. The frictional directionality can also exist in materials with particles having a preferred orientation. The directionality effect in friction on a macroscale is observed in some magnetic tapes. In a macroscale test, a 12.7 mm wide polymeric magnetic tape was wrapped over an aluminum drum and slid in a reciprocating motion with a normal load of 0.5 N and a sliding speed of about 60 mm/s [27.79]. The coefficient of friction as a function of sliding distance in either direction is shown in Fig. 27.23. We note that the coefficient of friction on a macroscale for this tape is different in different directions. Directionality in friction is sometimes observed on the macroscale; on the microscale this is the norm [27.2, 16]. On the macroscale, the effect of surface asperities normally is averaged out over a large number of contacting asperities.

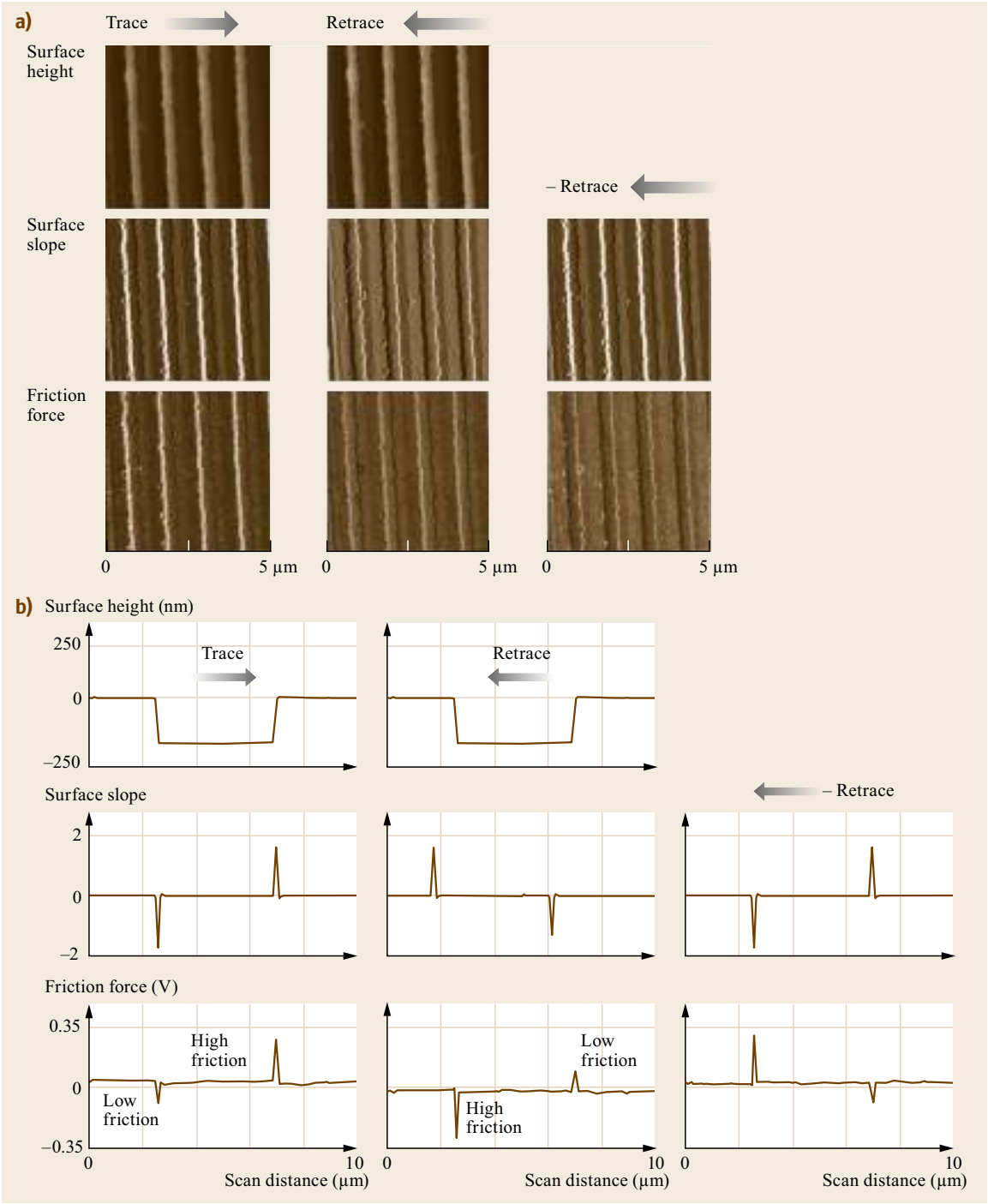


Fig. 27.22 (**a**) Gray-scale images of surface heights, surface slopes, and friction forces for scans across a gold-coated ruling, and (**b**) two-dimensional profiles of surface heights, surface slopes, and friction forces for scans across the silicon grid pit. *Arrows* indicate the tip sliding direction (after [27.101])

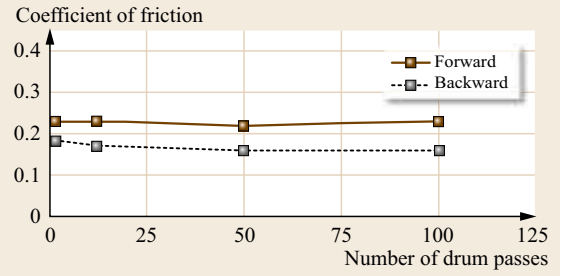


Fig. 27.23 Coefficient of macroscale friction as a function of drum passes for a polymeric magnetic tape sliding over an aluminum drum in a reciprocating mode in both directions. Normal load = 0.5 N over 12.7 mm wide tape, sliding speed = 60 mm/s (after [27.79])

27.2.4 Surface Roughness–Independent Microscale Friction

As just reported, the friction contrast in conventional friction measurements is based on interactions dependent upon interfacial material properties superimposed by roughness-induced lateral forces, and the cantilever twist is dependent on the sliding direction because of the local surface slope. Hence it is difficult to separate the friction-induced from the roughness-induced cantilever twist in the image. To obtain the roughness-independent friction, lateral or torsional modulation techniques are used in which the tip is oscillated in-plane with a small amplitude at a constant normal load, and change in shape and magnitude of cantilever resonance is used as a measure of friction force [27.4, 34–37, 106, 107]. These techniques also allow measurements over a very small area (few nm to few μm).

Scherer et al. [27.33] and *Reinstaedtler* et al. [27.34, 35] used lateral mode for friction measurements (Fig. 27.10b). *Bhushan* and *Kasai* [27.37] used the TR mode for these measurements (Fig. 27.11). Before engagement, the cantilever is driven into torsional motion of the cantilever/tip assembly with a given normal vibration amplitude (vibration amplitude in free air). After engagement, the vibration amplitude decreases due to the interaction between the tip and the sample, the vibration frequency increases, and phase shift occurs. During scanning, the normal load is kept constant, and the vibration amplitude of the cantilever is measured at the contact frequency.

As mentioned earlier, the shift in contact resonance frequency both in lateral and in TR modes is a measure of contact stiffness as shown schematically in Fig. 27.24. At an excitation voltage above a certain value, as a result of microslip at the interface, a flattening of the resonance frequency spectra occurs

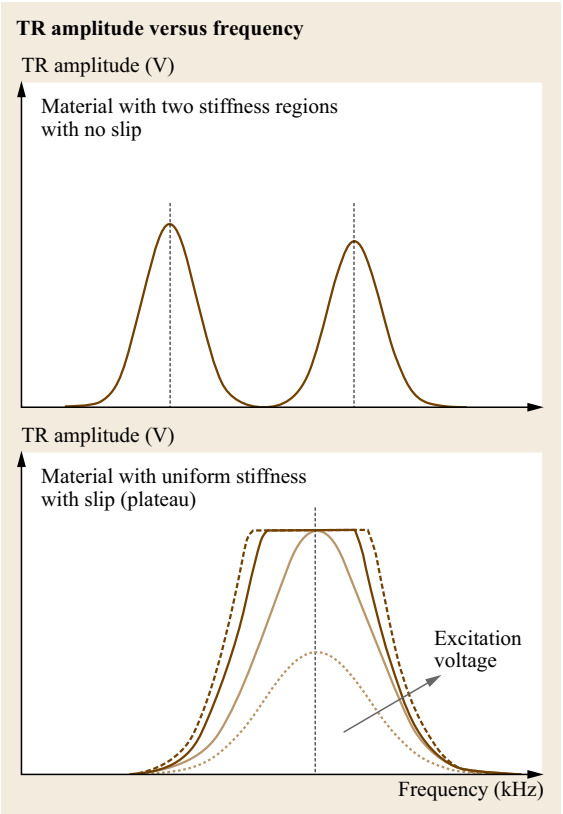


Fig. 27.24 Schematic showing frequency profiles of the TR amplitude for materials with two phases and a single phase. The maximum TR amplitude at the contact resonance frequency of the resonance curve with a flattened top, resulting from slip, can be used for friction force measurement

(Fig. 27.22). At lowest excitation voltage, the AFM tip sticks to the sample surface and follows the motion like an elastic contact with viscous damping, and the resonance curve is Lorentzian with a well-defined maximum. The excitation voltage should be high enough to initiate microslip. The maximum torsional amplitude at a given resonance frequency is a function of friction force and sample stiffness, so the technique is not valid for inhomogeneous samples. If the torsional stiffness of the cantilever is very high compared to the sample stiffness, the technique should work.

Reinstaedtler et al. [27.34] performed lateral mode experiments on bare Si and Si lubricated with 5 nm thick chemically bonded perfluoropolyether (Z-DOL) lubricant film. Figure 27.25a shows the amplitude of the cantilever torsional vibration as a function of frequency on a bare silicon sample. The frequency sweep was adjusted such that a contact resonance frequency was covered. The different curves correspond to dif-

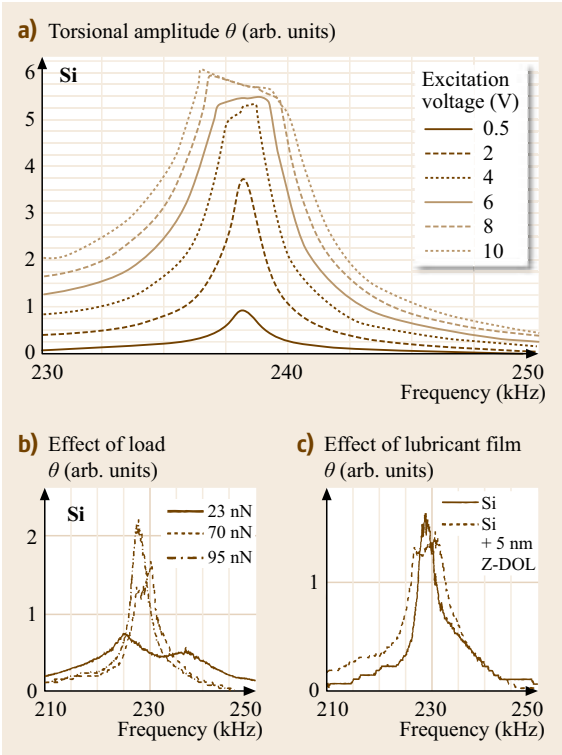


Fig. 27.25a–c Torsional vibration amplitude of the cantilever as a function of excitation frequency. **(a)** Measurement on bare silicon. The *different curves* correspond to increasing excitation voltages applied to the transducer and, hence, increasing surface amplitudes. **(b)** Measurement on silicon lubricated with a 5 nm thick Z-DOL layer. *Curves* for three different static loads are shown. The transducer was excited with 5 V of amplitude. **(c)** Measurement with a static load of 70 nN and 7 V excitation amplitude. The *two curves* correspond to bare silicon and lubricated silicon, respectively (after [27.34])

ferent excitation voltages applied to the shear wave transducer. At low amplitudes, the shape of the resonance curve is Lorentzian. Above a critical excitation amplitude of the transducer (excitation voltage = 4 V corresponding to ≈ 0.2 nm lateral surface amplitude as measured by interferometry), the resonance curve flattens out, and the frequency span of the flattened part increases further with the excitation amplitude. Here, the static force applied was 47 nN and the adhesion force was 15 nN. The resonance behavior of the tip-cantilever system in contact with the lubricated silicon sample (Fig. 27.25b) was similar to that of the bare silicon sample. By increasing the static load, the critical amplitude for the appearance of the flattening increases. The deviations from the Lorentzian resonance curve became visible at static loads lower than 95 nN. As shown in Fig. 27.25c, the resonance curve obtained at the same normal load of 70 nN and at the same excitation voltage (7 V) is more flattened on the lubricated sample than on the bare silicon which led us to conclude that the critical amplitude is lower on the lubricated sample than on the bare sample. These experiments clearly demonstrate that torsional vibration of an AFM cantilever at ultrasonic frequencies leads to stick-slip phenomena and sliding friction. Above a critical vibration amplitude, sliding friction sets in.

Bhushan and *Kasai* [27.37] performed friction measurements on a silicon ruler and demonstrated that friction data in TR mode is essentially independent of surface-roughness and sliding direction. Figure 27.26 shows surface height and friction force maps on a silicon ruler obtained using the TR mode and contact mode techniques. A comparison is made between the TR mode and contact mode friction force maps. For easy comparison, the line scan profiles near the central area are shown on top of the gray-scale maps. The vertical scales of the friction force profiles in the two graphs are selected to cover the same range of friction force so that direct comparison can be made, i. e., 0.25 V in full scale for the TR mode corresponds to 0.5 V for the contact mode in these measurements. As expected, for the trace scan, small downward peaks in the TR mode map and large upward and downward peaks in the contact mode map are observed. The positions of these peaks coincide with those of the surface slope. Therefore, the peaks in the friction signals are attributed to a topography-induced effect. For the retrace scan, the peak pattern for the TR mode remains similar, but for the contact mode, the pattern becomes reversed.

The subtraction image for the TR mode shows almost flat contrast, since the trace and retrace friction data profiles are almost identical. For the contact mode, the subtraction image shows that the topography-induced contribution still exists. As stated earlier, the addition image of the TR mode and the addition image of the contact mode enhance the topography-induced effect, which is observed in the figure.

A closer look at the silicon ruler images at one pitch was taken, and the associated images are shown in Fig. 27.27. The surface height profiles in the TR mode and contact mode are somewhat different. The TR mode shows sharper edges than those in contact mode. The ratios of the change in amplitude at the steps to the change in the mean amplitude in the TR mode and in the contact mode are a measure of topography effects. The ratio in the contact mode (≈ 85%) is about seven times larger than that of the TR mode (≈ 12%).

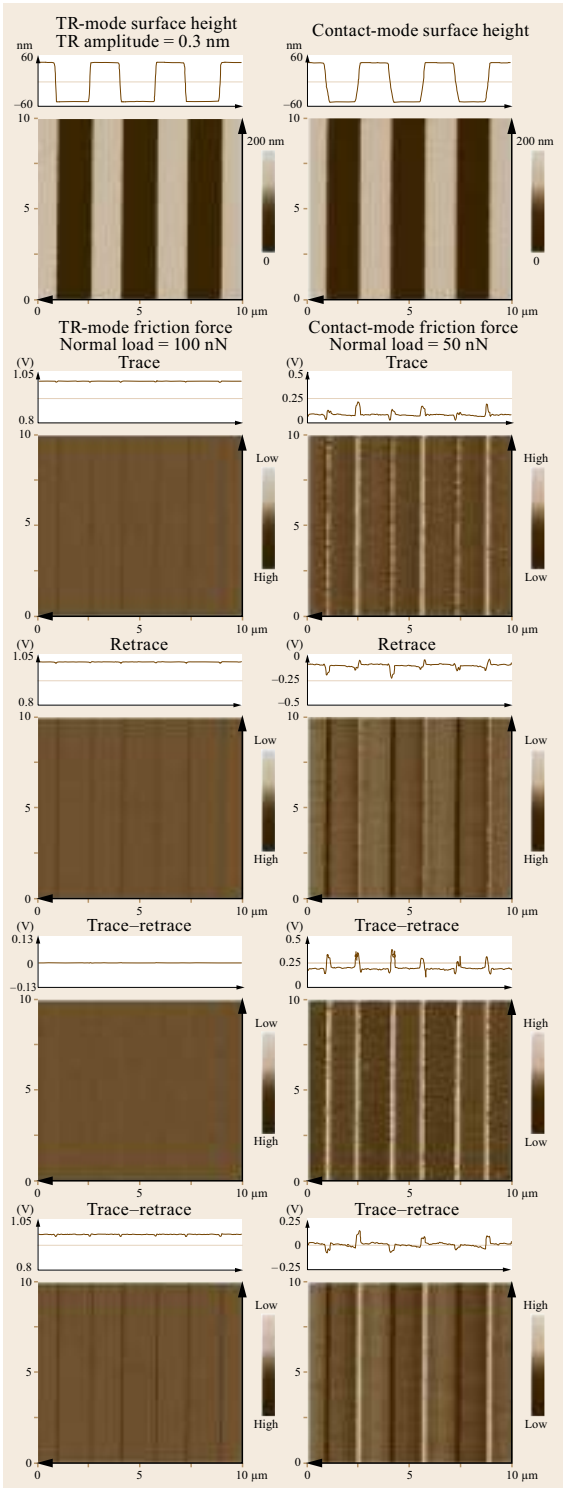


Fig. 27.26 A comparison between TR mode friction and contact mode friction maps together with line scans, on the silicon ruler. The TR mode surface height and contact mode surface height images are also shown (after [27.37])

27.2.5 Velocity Dependence on Micro/Nanoscale Friction

AFM/FFM experiments can be generally conducted at relative velocities as high as about 100–250 µm/s. To simulate applications, it is of interest to conduct friction experiments at higher velocities (up to 1 m/s). Furthermore, high-velocity experiments would be useful to study velocity dependence on friction and wear. One approach has been to mount samples on a shear wave transducer (ultrasonic transducer) and then drive it at very high frequencies (in the MHz range) as reported earlier (Fig. 27.10) [27.32–36, 106, 108]. The coefficient of friction on the nanoscale is estimated based on the contact resonance frequency and requires the solution of the characteristic equations for the tip vibrating in contact with the sample surface. The approach is complex and is dependent upon various assumptions.

An alternative approach is to utilize piezo stages with large amplitude ($\approx 10-100\,\mu m$) and relatively low resonance frequency (few kHz) and directly measure the friction force on microscale using the FFM signal without any analysis with assumptions used in the previous approaches using shear wave transducers. The commercial AFM setup modified with this approach can yield sliding velocities up to 200 mm/s [27.38, 109]. In the high-velocity piezo stage shown in Fig. 27.28b, the single axis piezo stage is oriented such that the scanning axis is perpendicular to the long axis of the AFM cantilever (which corresponds to the 90 °C scan angle mode of the commercial AFM). The displacement is monitored using an integrated capacitive feedback sensor, located diametrically opposite to the piezo crystal. The capacitive change, corresponding to the stage displacement, gives a measure of the amount of displacement and can be used as feedback to the piezo controller for better guiding and tracking accuracy during scanning. The closed loop position control of the piezoelectric driven stages using capacitive feedback sensors provides linearity of motion better than 0.01% with nanometer resolution and a stable drift-free motion [27.38]. In the ultrahigh-velocity piezo stage shown in Fig. 27.28a, a rectangular monolithic piezoceramic plate (the stator) with two excitation electrodes is resonated using a 12 V power supply. Depending on the desired direction of the motion, the left or right electrode is excited to produce high-frequency eigenmode oscillations up to 200 kHz. Simultaneous eigenmodes result in quasielliptical motion. An alumina friction tip (pusher) attached to the plate pushes a slider with a glued friction bar which rests on a set of bearings. Through its contact with the friction bar, the piezoceramic plate provides microimpulses and drives the slider forward or backward.

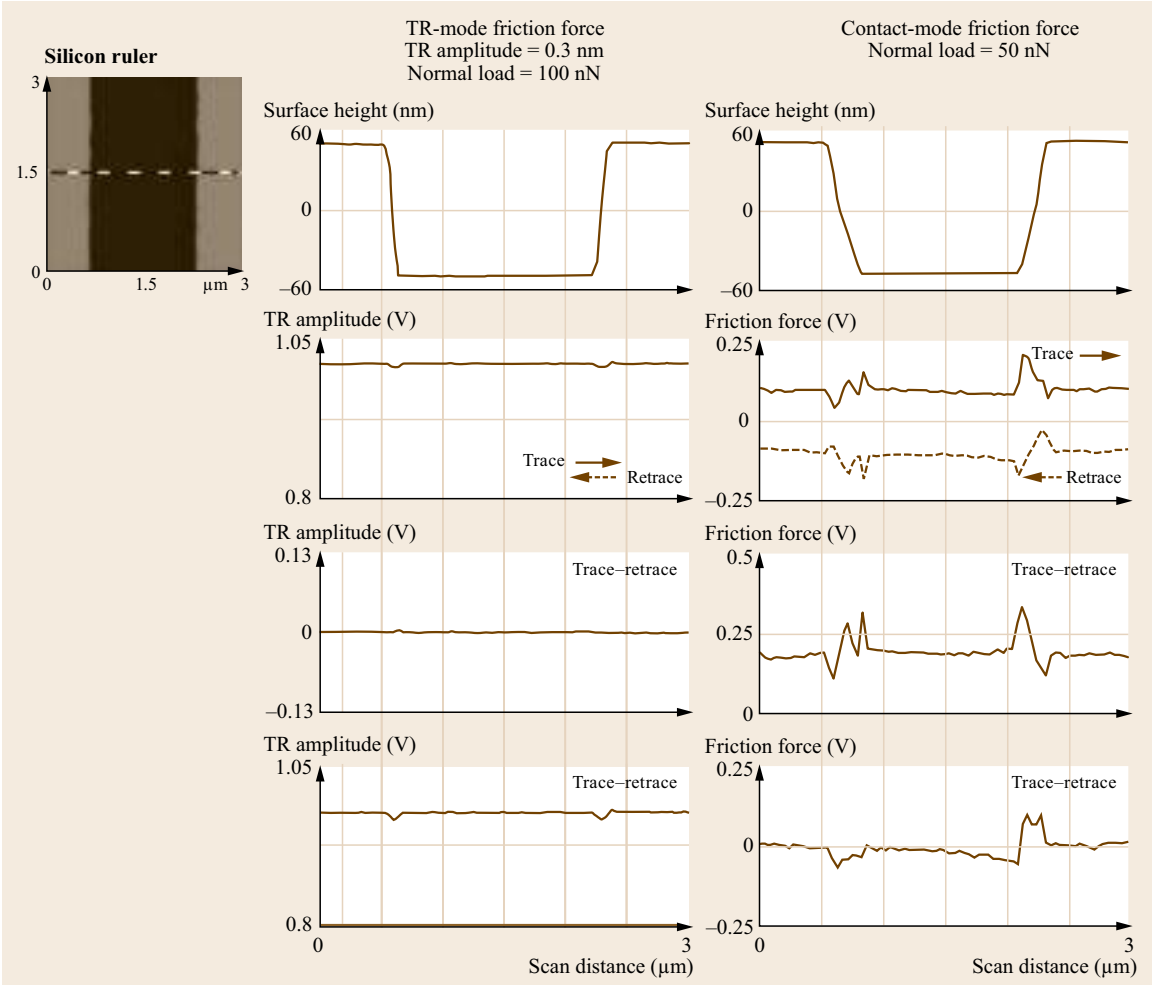


Fig. 27.27 A comparison of the line scans of TR mode friction and contact mode friction on a selected pitch of the silicon ruler (after [27.37])

While the longitudinal oscillation component provides energy as the driving force, the transverse component serves to change the pressure of the friction tip against the friction bar. The transverse oscillation energy determines the maximum frictional force and hence the holding and driving force of the stage. An optical position reference photosensor is located approximately in the middle of the travel range and is used to reference the absolute position of the stage within 1 μm repeatability. During motion, the increments of the linear scale from a home (reference) position point are converted to determine the position using a linear optical encoder.

The block diagram of the high-speed data collection and processing system used for the friction force measurement is shown in Fig. 27.28c. During the experiments, the AFM cantilever is held stationary by maintaining a scan size of zero. The mounted sample is scanned below the AFM tip by moving stages, and the normal and torsional deflections of the tip are recorded by a photodiode detector. The raw deflection signals from the optical detection system are directly routed to a high speed data acquisition A/D board. Raw friction data is acquired at a high sampling rate up to 80 kilosample/s.

Velocity dependence on friction for Si(100), diamond-like carbon (DLC), self-assembled monolayer, and perfluoropolyether lubricant films were studied by *Tambe* and *Bhushan* [27.38, 110–113] and *Tao* and *Bhushan* [27.109, 114]. The friction force as a function of velocity for Si(100) and DLC (deposited by filtered cathodic arc) is shown in Fig. 27.29 on a logarithm velocity scale (middle column). The solid lines in the figure represent the results for a scan length of 1000 μm

with velocity ranging from 1000 to $2\times10^5\,\mu m/s$ using the ultrahigh-velocity stage. The dotted lines represent results for a 25 μm scan length with velocity ranging from 5 to 500 μm/s using the high-velocity stage. To clearly show the friction force dependence on velocity in the lower range, the test results with velocity varying from 5 to 500 μm/s for 25 μm are shown on a magnified scale in the left column of Fig. 27.29.

For the Si(100) sample, the friction force decreased with velocity at low velocities ($v < 10\,\mu m/s$) and then increased linearly with $\log(v)$ for a 25 μm scan length. For the 1000 μm scan length, the friction force increased linearly with $\log(v)$ when velocity was lower than $2\times10^4\,\mu m/s$. When velocities were higher than $2\times10^4\,\mu m/s$, the friction force increased linearly with velocity. For DLC, the friction force increased linearly with $\log(v)$ from 5 to 500 μm/s for the 25 μm scan length. For the 1000 μm scan length, the friction force increases with velocity until about $2\times10^4\,\mu m/s$ where the friction force reaches a maximum, then the friction force decreased with velocity.

For different samples, the change in the friction force with velocity involves different mechanisms in relation to sample surface conditions. The silicon surface is hydrophilic, while the DLC surface is nearly hydrophobic. Under the ambient condition, a thin layer of water film is condensed on a hydrophilic sample surface. On a hydrophobic surface with high contact angle, the water film would be difficult to form on the sample surface, and the effect of the water film on the adhesive force and friction force could be neglected.

On the silicon surface, when the velocity is lower than 10 μm/s, the friction force decreased with velocity. This can be explained as follows. The water meniscus bridges develop as a function of time around the tip until reaching the equilibrium condition and are the dominant contributor to the friction force [27.2, 14, 103, 104]. The motion of the tip results in continuous breaking and reforming of the meniscus bridges. As the tip sliding velocity exceeds a critical velocity (10 μm/s), the menisci cannot reform due to insufficient time, and the meniscus force no longer plays a dominant role. Between 10 and $2\times10^4\,\mu m/s$, the friction increases linearly with $\log(v)$ for both 25 and 1000 μm scan lengths. This logarithmic dependence can be explained by the atomic-scale stick-slip phenomenon [27.111, 114]. At a velocity larger than $2\times10^4\,\mu m/s$, the friction increases linearly with velocity, and this trend can be explained by viscous shear; see the friction force plotted as a function of velocity on a linear scale at a magnified scale in the right column of Fig. 27.29.

To explain the atomic-scale stick-slip mechanism of friction, the motion for the tip is expressed by a spring

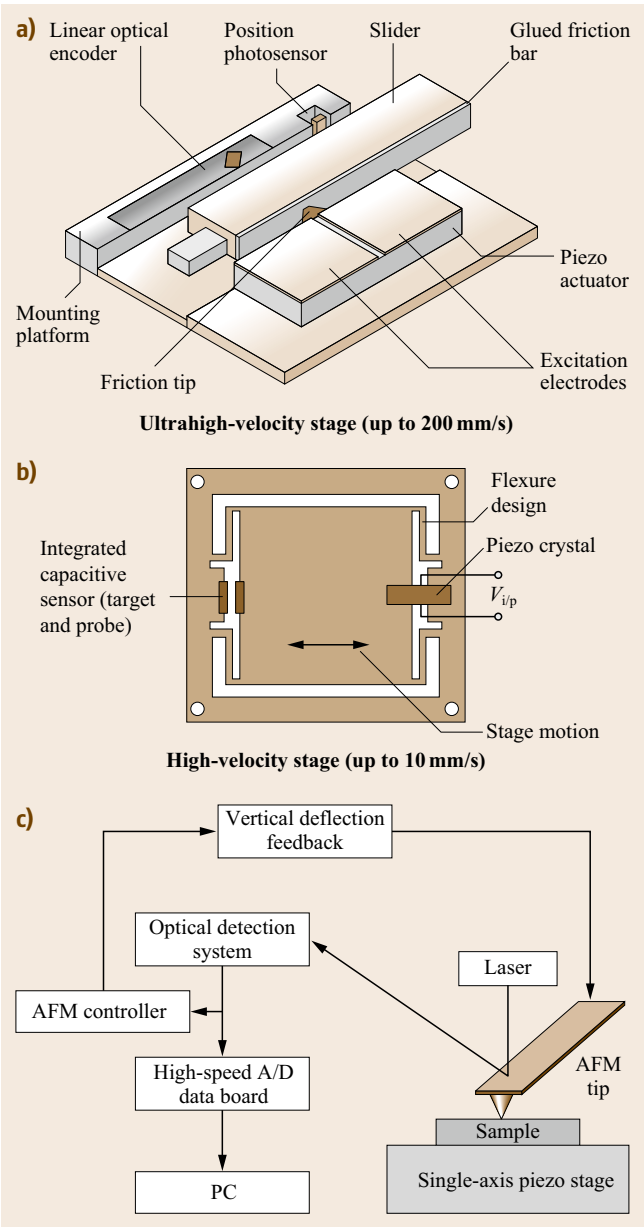


Fig. 27.28a–c Schematics of (a) an ultrahigh-velocity piezo stage, (b) high-velocity piezo stage, and (c) a block diagram of the high-speed data collection and processing system used for friction force measurement (after [27.38, 109])

mass model [27.115] as follows

$$m\ddot{x}_t = -\eta\dot{x}_t - k(x_M - x_t) - F\,, \qquad (27.6)$$

where m is the effective mass of the system, η the viscous damping coefficient, k the spring constant of the cantilever, $x_M = v_M t$ the equilibrium position of the

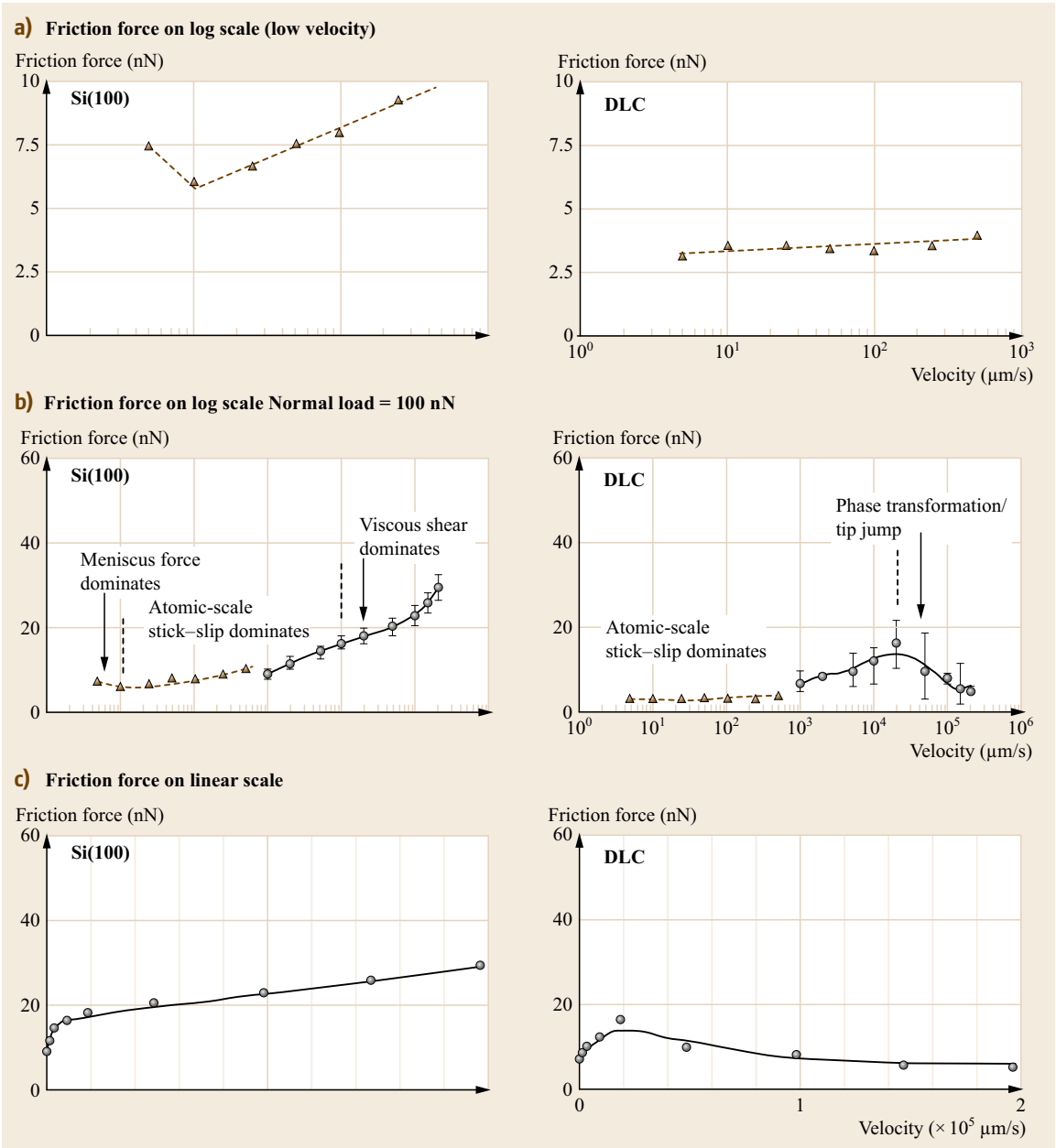


Fig. 27.29a–c Friction force as a function of sliding velocity obtained for a 25 μm scan length using a high-velocity stage (*dotted line*) and a 1000 μm scan length using an ultrahigh-velocity stage (*solid line*). In the *top two rows* (**a**,**b**), velocity is plotted on a log scale. (**a**) shows a lower velocity range – between 1 and 500 μm/s. (**c**) shows the data at a higher range of velocity on the linear scale (after [27.114])

cantilever, x_t the position of the tip, and F the external force. The lateral force is expressed as $F_l = k(x_M - x_t)$, and the friction force F_{fric} is the lateral force averaged over time.

For velocities lower than 2×10^4 μm/s, the damping part ($\eta\dot{x}_t$) in (27.6) is comparatively low, and atomic-scale stick-slip dominates. To investigate the stick-slip, *Tomlinson* [27.96] assumed a periodic surface with potential

$$V(x) = V_0\left[1-\cos\left(\frac{2\pi x}{a}\right)\right], \qquad (27.7)$$

where V_0 is surface barrier potential height, and a is the lattice constant of the surface. Then the force (F) in (27.6) is expressed as

$$F = V'(x) = \frac{2\pi V_0}{a} \sin\left(\frac{2\pi x}{a}\right) . \tag{27.8}$$

On the basis of the Tomlinson model, and taking into account the effect of thermal activation, or the elastic energy stored in the cantilever during sliding, *Gnecco* et al. [27.116] derived the relationship between the friction force and velocity, which is expressed as

$$F_{\text{stick-slip}} = F_0 + c \ln v , \tag{27.9}$$

where F_0 and c are constants.

When the tip slides on a solid surface covered by a viscous film such as a water film at high velocities, the friction force (F_{fric}) is related to the velocity and viscosity of the film given by [27.103, 104]

$$F_{\text{fric}} = \mu N + \eta \dot{\gamma} A \approx \mu N + \frac{\eta v A}{d} , \tag{27.10}$$

where μ is the coefficient of friction between the dry sliding bodies, N is the applied load, τ is the shear stress, A is the real contact area, η is the viscosity of the film, $\dot{\gamma}$ is the velocity gradient, v is the sliding velocity, and d is the thickness of the film. From (27.10), the relationship between the friction force and the sliding velocity is linear when sliding on a viscous coating. The relationship is consistent with the conclusion by *Helman* et al. [27.117] about the linear relationship between the friction force and the sliding velocity ($F_{\text{fric}} \approx \eta v_{\text{M}}$) at high sliding velocities for a spring mass model in (27.6) which simulates the AFM tip sliding on a viscous liquid.

The sliding of the tip on a hydrophilic surface with water films at low velocities, middle velocities, and high velocities is schematically illustrated in Fig. 27.30a [27.114]. It should be noted that the stick-slip mechanism considered by *Gnecco* et al. [27.116] was based on investigations on a dry surface. In this study, although the water was condensed on the Si(100) surface, the water film on the surface would not have significant effect on the energy dissipation due to the surface variation at relatively low velocities. Thus, the linear relationship between friction and $\log(v)$ could be maintained. When the velocity increases above a certain value, the tip would lose direct contact with the sample surface and shear the water film. At velocities higher than $2\times10^4\,\mu\text{m/s}$, the asperity deformation from the high velocity impact could be another mechanism as proposed by *Tambe* and *Bhushan* [27.111].

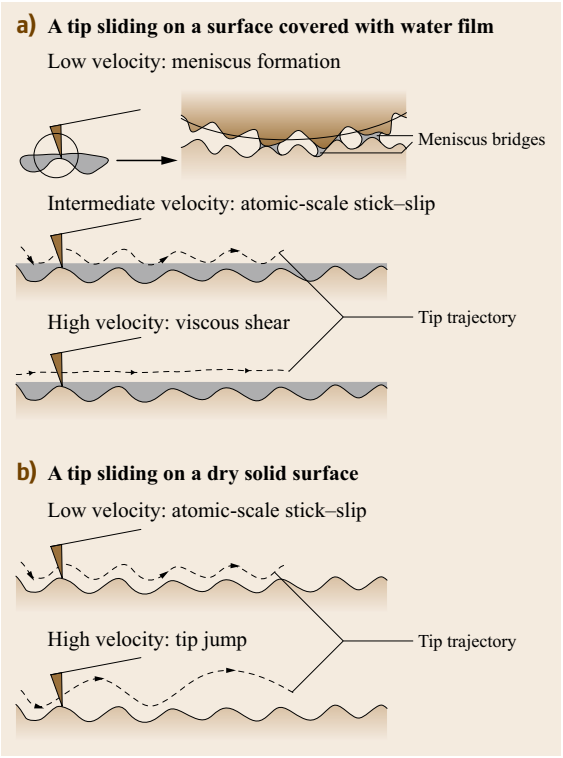


Fig. 27.30a,b Schematics of a tip sliding at different velocities on **(a)** a water-covered surface, and **(b)** a dry surface (after [27.114])

For the DLC film, since the surface is nearly hydrophobic, a uniform water film would not form on the surface. When sliding at a velocity lower than $1000\,\mu\text{m/s}$, the friction force increased linearly with $\log(v)$, which could also be explained by atomic-scale stick-slip. At velocities higher than $1000\,\mu\text{m/s}$, the friction force increased with velocity until the local maximum at the velocity of $2\times10^4\,\mu\text{m/s}$, then decreased with velocity. The decreasing trend in friction at higher velocities could be due to tip jump during sliding. The tip jump is illustrated in Fig. 27.30b. The tip jump results in the reduction of lateral force during sliding. Variation of the friction force with distance, indicative of the tip jump, was observed from the lateral force signal (not shown). When damping is low and velocity is high, the tip could jump several periodical cycles or several peaks [27.118]. At a given low damping coefficient, the slip results in a low transient lateral force, as discussed by *Fusco* and *Fasolino* [27.118]. Thus, the average lateral force (friction force) over the scan length is low. The tip jump could also cause high velocity impact of asperities on the DLC surface, resulting in the phase transformation of DLC from sp^3 to sp^2, as explained by *Tambe* and *Bhushan* [27.111]. The

sp^2-phase layer can act as a lubricant and reduce the interfacial friction.

27.2.6 Nanoscale Friction and Wear Mapping

Contrary to the classical friction laws postulated by Amontons and Coulomb centuries ago, a nanoscale friction force is found to be strongly dependent on the normal load and sliding velocity. Many materials, coatings, and lubricants that have wide applications show reversals in friction behavior corresponding to transitions between different friction mechanisms [27.14, 38, 110, 111, 119]. Most of the analytical models developed to explain nanoscale friction behavior have remained limited in their focus and have left investigators at a loss when trying to explain friction behavior scaling multiple regimes. Nanoscale friction maps provide fundamental insights into friction behavior. They help identify and classify the dominant friction mechanisms, as well as determine the critical operating parameters that influence transitions between different mechanisms [27.111, 112]. Figure 27.31 shows a nanoscale friction map for DLC with the friction mapped as a function of the normal load and the sliding velocity [27.120]. The contours represent constant friction force lines. The friction force is seen to increase with the normal load as well as velocity. The increase in friction force with velocity is the result of atomic-scale stick-slip. This is a result of thermal activation of the irreversible jumps of the AFM tip that arise from overcoming of the energy barrier between the adjacent atomic positions, as described earlier. The concentric contour lines corresponding to the constant friction force predicts a peak point, a point where the friction force reaches maxima and beyond which point any further increase in the normal load or the sliding velocity results in a decrease in friction force. This characteristic behavior for DLC is the result of phase transformation of DLC into a graphite-like phase by sp^3 to sp^2 phase transition, as described earlier. During the AFM experiments, the Si_3N_4 tip gives rise to contact pressures in the range of 1.8–4.4 GPa for DLC for normal loads of 10–150 nN [27.121]. A combination of the high contact pressures that are encountered on the nanoscale and the high frictional energy dissipation arising from the asperity impacts at the tip–sample interface due to the high sliding velocities accelerates a phase transition process whereby a low shear strength graphite-like layer is formed at the sliding interface.

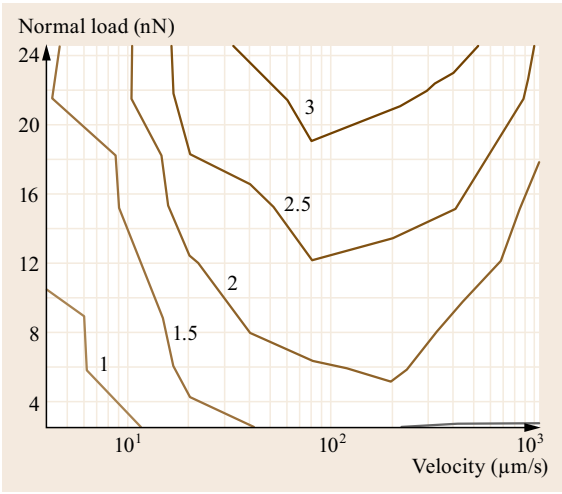


Fig. 27.31 Contour map showing friction force dependence on normal load and sliding velocity for DLC (after [27.97])

Similar to friction mapping, one way of exploring the broader wear patterns is to construct wear mechanism maps that summarize data and models for wear, thereby showing mechanisms for any given set of conditions to be identified [27.119, 122–124]. Wear of sliding surfaces can occur through one or more wear mechanisms, including adhesive, abrasive, fatigue, impact, corrosive, and fretting [27.103, 104]. *Tambe* and *Bhushan* [27.121, 124] performed AFM experiments to develop nanoscale wear maps. Figure 27.32 shows a nanowear map generated for a DLC sample by simultaneously varying the normal load and the sliding velocity over the entire scan area. The wear map was generated for a normal load range of 0–1000 nN and sliding velocity range of 0–2.5 mm/s. Wear debris, believed to result from phase transformation of DLC by sp^3 to sp^2 phase transition, was seen to form only for a high value of sliding velocities times normal loads, i. e., only beyond a certain threshold of friction energy dissipation [27.121, 124]. Hence the wear region exhibits a transition line indicating that for low velocities and low normal loads there is no phase transformation. For clarity, the wear mark corners are indicated by white dots in the AFM image (top) and the two zones of interest over the entire wear mark are schematically illustrated in Fig. 27.32 (top).

Nanoscale friction and wear mapping are novel techniques for investigating friction and wear behavior on the nanoscale over a range of operating parameters. By simultaneously varying the sliding velocity and normal load over a large range of values, nanoscale friction and wear behavior can be mapped, and the transitions between different wear mechanisms can be investigated. These maps help identify and demarcate critical operating parameters for different wear mechanisms and are very important tools in the process of design and selection of materials/coatings.

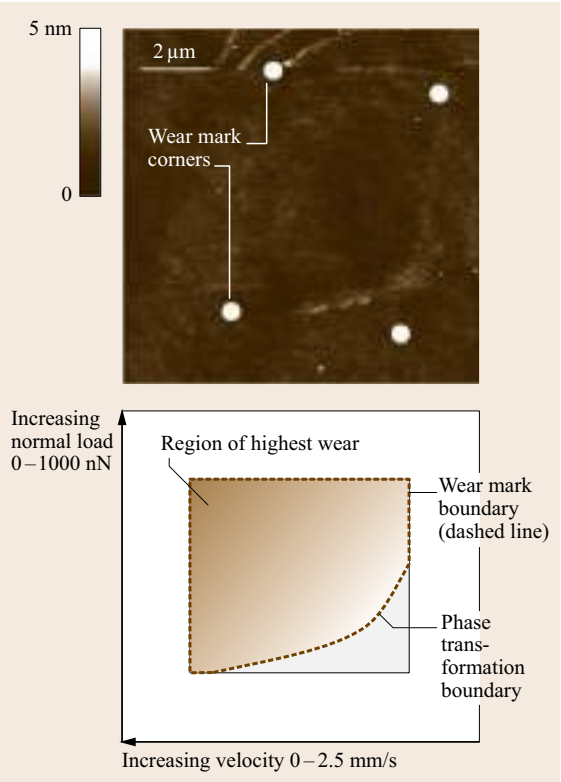


Fig. 27.32 Nanowear map (AFM image and schematic) illustrating the effect of sliding velocity and normal load on the wear of DLC resulting from phase transformation. The *curved area* shows debris lining and is indicative of the minimum frictional energy needed for phase transformation. For clarity, the wear mark corners are indicated by *white dots* in the AFM image and the various zones of interest over the entire wear mark are schematically illustrated (after [27.124])

27.2.7 Adhesion and Friction in Wet Environments

Experimental Observations

Relative humidity affects adhesion and friction for dry and lubricated surfaces [27.19, 31, 125]. Figure 27.33 shows the variation of single point adhesive force measurements as a function of tip radius on an Si(100) sample for several humidities. The adhesive force data are also plotted as a function of relative humidity for several tip radii. The general trend at humidities up to the ambient is that a 50 nm radius Si_3N_4 tip exhibits a lower adhesive force as compared to the other microtips of larger radii. However, in the latter case, values are similar. Thus, for the microtips there is no appreciable variation in adhesive force with tip radius at a given humidity up to the ambient. The adhesive force increases as relative humidity increases for all tips.

The sources of adhesive force between a tip and a sample surface are van der Waals attraction and meniscus formation [27.19, 103, 104]. The relative magnitudes of the forces from the two sources are dependent upon various factors including the distance between the tip and the sample surface, their surface roughness, their hydrophobicity, and the relative humidity [27.126]. For most rough surfaces, the meniscus contribution dominates at moderate to high humidities which arise from the capillary condensation of water vapor from the environment. If enough liquid is present to form a meniscus bridge, the meniscus force should increase with an increase in tip radius (proportional to tip radius for a spherical tip). In addition, an increase in the tip radius results in an increased contact area leading to higher values of van der Waals forces. However, if nanoasperities on the tip and the sample are considered then the number of contacting and near-contacting asperities forming meniscus bridges increases with an increase of humidity leading to an increase in meniscus forces. These explain the trends observed in Fig. 27.33. From the data, the tip radius has little effect on the adhesive forces at low humidities but increases with tip radius at high humidity. Adhesive force also increases with an increase in humidity for all tips. This observation suggests that thickness of the liquid film at low humidities is insufficient to form continuous meniscus bridges to affect adhesive forces in the case of all tips.

Figure 27.33 also shows the variation in the coefficient of friction as a function of the tip radius at a given humidity, and as a function of the relative humidity for a given tip radius for Si(100). It can be observed that for 0% RH (relative humidity), the coefficient of friction is about the same for the tip radii except for the largest tip, which shows a higher value. At all other humidities, the trend consistently shows that the coefficient of friction increases with the tip radius. An increase in friction with tip radius at low to moderate humidities arises from the increased contact area (due to higher van der Waals forces) and higher values of the shear forces required for a larger contact area. At high humidities, similar to adhesive force data, an increase with tip radius occurs because of both contact area and meniscus effects. Although the AFM/FFM measurements are able to measure the combined effect of the contribution of van der Waals and meniscus forces towards the friction force or adhesive force, it is difficult to measure their individual contributions separately. It can be seen that for all tips, the coefficient of friction increases with humidity to about ambient, beyond which it starts to decrease. The initial increase in the coefficient of friction with humidity arises from the fact that the thickness of

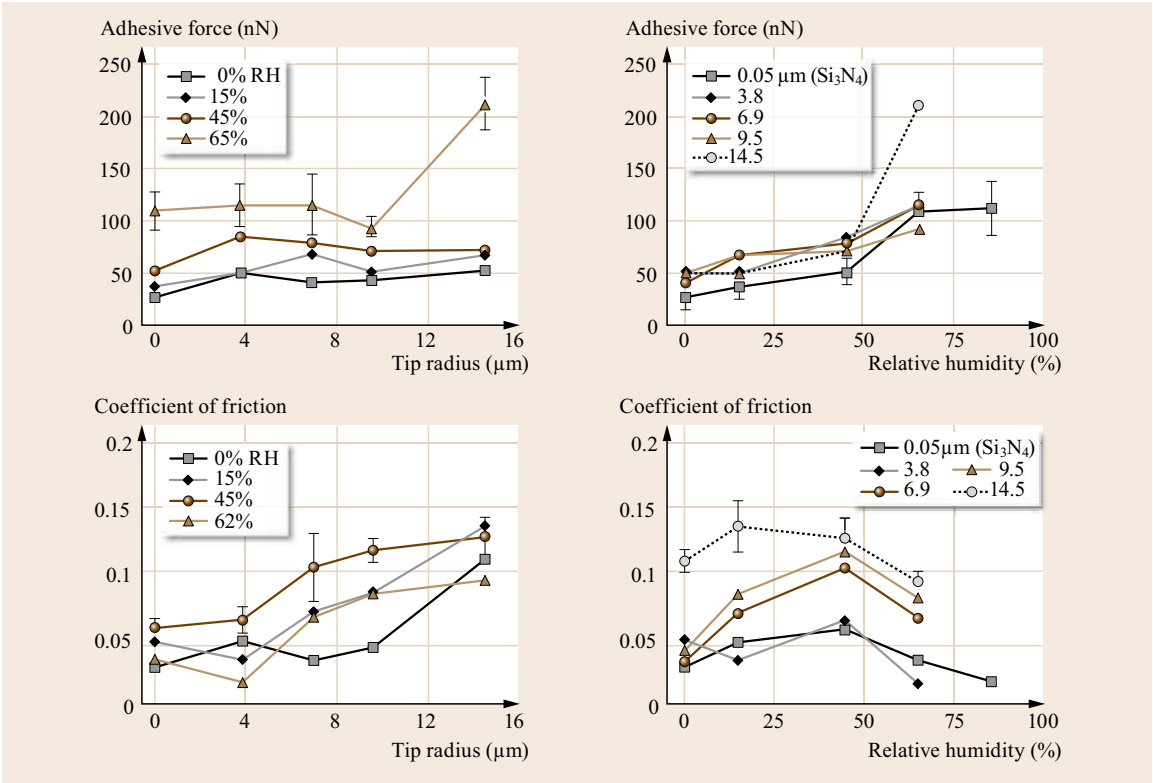


Fig. 27.33 Adhesive force and coefficient of friction as a function of tip radius at several humidities and as a function of relative humidity at several tip radii on Si(100) (after [27.31])

the water film increases with an increase in the humidity, which results in a larger number of nanoasperities forming meniscus bridges and leads to higher friction (larger shear force). The same trend is expected with the microtips beyond 65% RH. This is attributed to the fact that at higher humidities, the adsorbed water film on the surface acts as a lubricant between the two surfaces. Thus, the interface is changed at higher humidities, resulting in lower shear strength and hence lower friction force and coefficient of friction.

Adhesion and Friction Force Expressions for a Single-Asperity Contact

We now obtain the expressions for the adhesive force and coefficient of friction for a single-asperity contact with a meniscus formed at the interface (Fig. 27.34). For a spherical asperity of radius R in contact with a flat and smooth surface with the composite modulus of elasticity E^* and in the presence of liquid with a concave meniscus, the attractive meniscus force (adhesive force), designated as F_m or W_{ad}, is given as [27.19, 103, 104]

$$W_{ad} = 2\pi R\gamma(\cos\theta_1 + \cos\theta_2)\,, \tag{27.11}$$

where γ is the surface tension of the liquid, and θ_1 and θ_2 are the contact angles of the liquid with surfaces 1 and 2, respectively. For an elastic contact for both extrinsic (W) and intrinsic (W_{ad}) normal load, the friction force is given as,

$$F_e = \pi\tau\left[\frac{3(W + W_{ad})R}{4E^*}\right]^{2/3}, \tag{27.12}$$

where W is the external load, and τ is the average shear strength of the contacts. (The surface energy effects are not considered here.) Note that adhesive force increases linearly with an increase in the tip radius, and the friction force increases with an increase in the tip radius as $R^{2/3}$ and with normal load as $(W + W_{ad})^{2/3}$. The experimental data in support of $W^{2/3}$ dependence on the friction force can be found in various references [27.67]. The coefficient of friction μ_e is obtained from (27.12) as

$$\mu_e = \frac{F_e}{(W + W_{ad})} = \pi\tau\left(\frac{3R}{4E^*}\right)^{2/3}\frac{1}{(W + W_{ad})^{1/3}}\,. \tag{27.13}$$

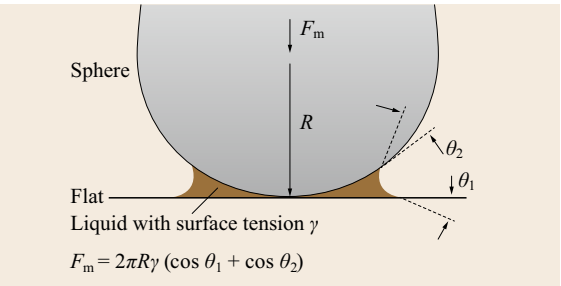


Fig. 27.34 Meniscus formation from a liquid condensate at the interface for a sphere in contact with a plane surface

In the plastic contact regime [27.103, 104], the coefficient of friction μ_p is obtained as

$$\mu_p = \frac{F_p}{W + W_{ad}} = \frac{\tau}{H_s}\,, \tag{27.14}$$

where H_s is the hardness of the softer material. Note that in the plastic contact regime, the coefficient of friction is independent of external load, adhesive contributions, and surface geometry.

For comparisons, for multiple-asperity contacts in the elastic contact regime the total adhesive force W_{ad} is the summation of adhesive forces at n individual contacts,

$$W_{ad} = \sum_{i=1}^{n} (W_{ad})_i \tag{27.15}$$

and

$$\mu_e \approx \frac{3.2\tau}{E^*(\sigma_p/R_p)^{1/2} + (W_{ad}/W)}\,,$$

where σ_p and R_p are the standard deviation of summit heights and average summit radius, respectively. Note that the coefficient of friction depends upon the surface roughness. In the plastic contact regime, the expression for μ_p in (27.14) does not change.

The source of the adhesive force in a wet contact in the AFM experiments being performed in an ambient environment includes mainly attractive meniscus force due to capillary condensation of water vapor from the environment. The meniscus force for a single contact increases with an increase in the tip radius. A sharp AFM tip in contact with a smooth surface at low loads (on the order of a few nN) for most materials can be simulated as a single-asperity contact. At higher loads, for rough and soft surfaces, multiple contacts would occur. Furthermore, at low loads (nN range) for most materials, the local deformation would be primarily elastic. Assuming that the shear strength of contacts does not change, the adhesive force for smooth and hard surfaces at low normal load (on the order of a few nN) (for a single-asperity contact in the elastic contact regime) would increase with an increase in tip radius, and the coefficient of friction would decrease with an increase in total normal load as $(W + W_{ad})^{-1/3}$ and would increase with an increase of tip radius as $R^{2/3}$. In this case, Amontons' law of friction which states that the coefficient of friction is independent of the normal load and is independent of the apparent area of contact does not hold. For a single-asperity plastic contact and multiple-asperity plastic contacts, neither the normal load nor tip radius comes into play in the calculation of the coefficient of friction. In the case of multiple-asperity contacts, the number of contacts increase with an increase of normal load, therefore the adhesive force increases with an increase in load.

In the data presented earlier in this section, the effect of tip radius and humidity on the adhesive forces and the coefficient of friction is investigated for experiments with the Si(100) surface at loads in the range of 10–100 nN. The multiple-asperity elastic-contact regime is relevant for this study involving large tip radii. An increase in humidity generally results in an increase in the number of meniscus bridges, which would increase the adhesive force. As was suggested earlier, that increase in humidity also may decrease the shear strength of contacts. A combination of an increase in adhesive force and a decrease in shear strength would affect the coefficient of friction. An increase in the tip radius would increase the meniscus force (adhesive force). A substantial increase in the tip radius may also increase interatomic forces. These effects influence the coefficient of friction with an increase in the tip radius.

27.2.8 Separation Distance Dependence of Meniscus and van der Waals Forces

When two surfaces are in close proximity, sources of adhesive forces are weak van der Waals attraction and meniscus formation. Relative magnitudes of the forces from the two sources are dependent upon various factors including the interplanar separation, their surface roughness, their hydrophobicity, and relative humidity (liquid volume) [27.126]. Meniscus contribution dominates at moderate to high humidities, and van der Waals forces dominate at asperities a few nm apart. In some micro/nanocomponents, it is important to know the relative contribution of two sources as a function of given interplanar separation so as to design an interface for low adhesion. For example, if two ultrasmooth surfaces come in close proximity with interplanar separation on

the order of a nm, van der Waals forces may dominate, and their magnitude may be reduced by creating bumps on one of the interfaces. This analysis is also of interest in AFM studies to understand distance dependence of adhesive forces as the tip goes in and out of contact.

Stifter et al. [27.126] modeled the contact of a parabolic-shaped tip and a flat, smooth sample surface. The tip may represent a surface asperity on an interface or an AFM tip in an AFM experiment. They calculated van der Waals and meniscus forces as a function of different parameters, namely, tip geometry, tip–sample starting distance, relative humidity, surface tension, and contact angles. They compared the meniscus forces with van der Waals forces to understand their relative importance in various operating conditions.

The interacting force between tip and sample in dry conditions is the Lennard-Jones force derived from the Lennard-Jones potential. The Lennard-Jones potential is composed of two interactions – the van der Waals attraction and the Pauli repulsion. Van der Waals forces are significant because they are always present. For a parabolic tip above a half-plane with a distance D between tip and plane, the Lennard-Jones potential is obtained by integrating the atomic potential over the volume of the tip and sample. It is given as [27.126]

$$V(D) = \frac{c}{12}\left(-\frac{A}{D} + \frac{B}{210D^7}\right), \qquad (27.16)$$

where c is the width of the parabolic tip (= diameter in the case of a spherical tip), and A and B are two potential parameters, where A is the Hamaker constant. This equation provides expressions for attractive and repulsive parts. The calculations were made for Lennard-Jones force (total) and van der Waals force (attractive part) for two Hamaker constants: 0.04×10^{-19} J (representative of polymers) and 3×10^{-19} J (representative of ceramics), and meniscus force for a water film ($\gamma_\ell = 72.5\,\mathrm{N/m}$). Figure 27.35 shows various forces as a function of separation distance. The effect of two relative humidities and three tip radii was also studied as to their effect on meniscus forces. The two dashed curves indicate the spread of possible van der Waals forces for two Hamaker constants. The figure shows that meniscus forces exhibit weaker distance dependence. The meniscus forces can be stronger or weaker than van der Waals forces for distances smaller than about 0.5 nm. For longer distances, the meniscus forces are stronger than the van der Waals forces. Van der Waals forces must be considered for a tip–sample distance up to a few nm ($D < 5$ nm). The meniscus forces operate up to the break of the meniscus in the range from 5 to 20 nm [27.126].

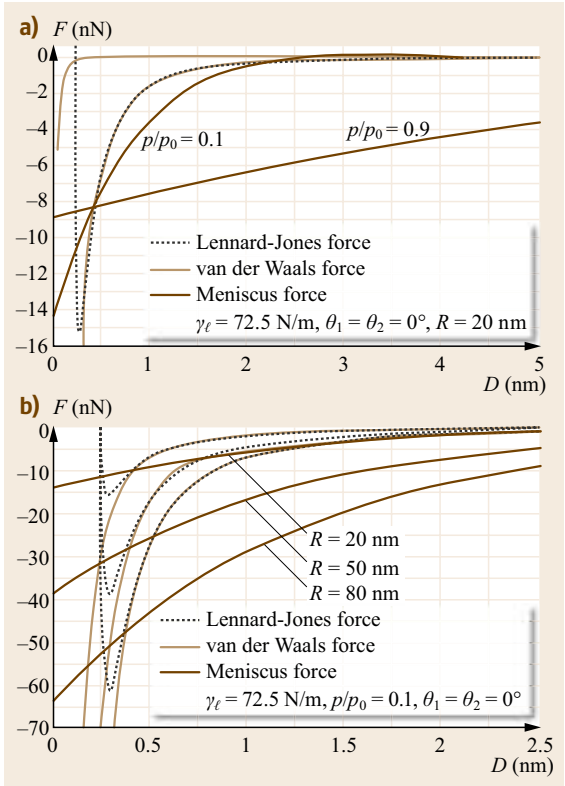


Fig. 27.35a,b Relative contribution of meniscus, van der Waals and Lennard-Jones forces (F) as a function of separation distance (D) and at **(a)** two values of relative humidity (p/p_o) for tip radius of 20 nm and Hamaker constants of 0.04×10^{-19} and 3.0×10^{-19} J, and **(b)** three tip radii (R) and a Hamaker constant of 3.0×10^{-19} J (after [27.126])

27.2.9 Scale Dependence in Friction

Table 27.3 presents the adhesive force and the coefficient of friction data obtained on the nanoscale and microscale [27.25, 110, 127, 128]. Adhesive force and coefficient of friction values on the nanoscale are about half to one order of magnitude lower than that on the microscale. Scale dependence is clearly observed in this data. As a further evidence of scale dependence, Table 27.4 shows the coefficient of friction measured for Si(100), HOPG, natural diamond, and DLC on the nanoscale and microscales. It is clearly observed that friction values are scale-dependent.

To estimate the scale length, the apparent contact radius at test loads is calculated and presented in Table 27.3. Mean apparent pressures are also calculated and presented. For nanoscale AFM experiments, it is assumed that an AFM tip coming in contact with a flat surface represents a single-asperity and elastic contact,

Table 27.3 Micro- and nanoscale values of adhesive force and coefficient of friction in micro- and nanoscale measurements (after [27.127])

Sample	Adhesive force		Coefficient of friction	
	Microscale[a] (μN)	Nanoscale[b] (nN)	Microscale[a]	Nanoscale[b]
Si(100)	685	52	0.47	0.06
DLC	325	44	0.19	0.03
Z-DOL	315	35	0.23	0.04
HDT	180	14	0.15	0.006

[a] Versus 500 μm radius Si(100) ball.
[b] Versus 50 nm radius Si_3N_4 tip.

Table 27.4 Micro- and nanoscale values of the coefficient of friction, typical physical properties of specimen, and calculated apparent contact radii and apparent contact pressures at loads used in micro- and nanoscale measurements. For calculation purposes it is assumed that contacts on micro- and nanoscale are single-asperity elastic contacts (after [27.134])

Sample	Coefficient of friction		Elastic modulus (GPa)	Poisson's ratio	Hardness (GPa)	Apparent contact radius at test load for		Mean apparent pressure at test load for	
	microscale	nanoscale				microscale (μm) (upper limit)	nanoscale (nm)	microscale (GPa) (lower limit)	nanoscale (GPa)
Si(100) wafer	0.47[a]	0.06[c]	130[e,f]	0.28[f]	9–10[e,f]	0.8–2.2[a]	1.6–3.4[c]	0.05–0.13[a]	1.3–2.8[c]
Graphite (HOPG)	0.1[b]	0.006[c]	9–15[g] (9)	– (0.25)	0.01[j]	62[b]	3.4–7.4[c]	0.082[b]	0.27–0.58[c]
Natural diamond	0.2[b]	0.05[c]	1140[h]	0.07[h]	80–104[g,h]	21[b]	1.1–2.5[c]	0.74[b]	2.5–5.3[c]
DLC film	0.19[a]	0.03[d]	280[i]	0.25[i]	20–30[i]	0.7–2.0[a]	1.3–2.9[d]	0.06–0.16[a]	1.8–3.8[d]

[a] 500 μm radius Si(100) ball at 100–2000 μN and 720 μm/s in dry air, *Bhushan* et al. [27.127]
[b] 3 mm radius Si_3N_4 ball (elastic modulus 310 GPa, Poisson's ratio 0.22 (*Bhushan* and *Gupta* [27.129])), at 1 N and 800 μm/s, *Ruan* and *Bhushan* [27.25]
[c] 50 nm radius Si_3N_4 tip at load range from 10–100 nN and 0.5 nm/s in dry air, *Ruan* and *Bhushan* [27.25]
[d] 50 nm radius Si_3N_4 tip at load range from 10–100 nN in dry air, *Bhushan* et al. [27.127]
[e] *Bhushan* and *Venkatesan* [27.130]
[f] *Properties of Silicon* [27.131]
[g] *Bhushan* and *Gupta* [27.129]
[h] *Field* [27.132]
[i] *Bhushan* [27.18]
[j] *National Carbon Company* [27.133]

and Hertz analysis was used for the calculations. In the microscale experiments, a ball coming in contact with a flat surface represents multiple-asperity contacts due to the roughness, and the contact pressure of the asperity contacts is higher than the apparent pressure. For calculation of a characteristic scale length for multiple asperity contacts, which is equal to the apparent length of contact, Hertz analysis was also used. This analysis provides an upper limit on apparent radius and lower limit on the mean contact pressure.

There are several factors responsible for the differences in the coefficients of friction at the micro- and nanoscale. Among them are the contributions from wear and contaminant particles, transition from elasticity to plasticity, and the meniscus effect [27.134–137]. The contribution of wear and contaminant particles is more significant at the macro/microscale because of the larger number of trapped particles, referred to as third-body contribution. It can be argued that for the nanoscale AFM experiments the asperity contacts are predominantly elastic (with average real pressure being less than the hardness of the softer material), and adhesion is the main contribution to the friction, whereas for the microscale experiments the asperity contacts are predominantly plastic, and deformation is an important factor. It will be shown later that hardness has a scale effect; it increases with decreasing scale and is responsible for less deformation on a smaller scale. The meniscus effect results in an increase of friction with increasing tip radius (Fig. 27.33). Therefore, third-

body contribution, scale-dependent hardness, and other properties transition from elastic contacts in nanoscale contacts to plastic deformation in microscale contacts, and meniscus contribution plays an important role.

Friction is a complex phenomenon, which involves asperity interactions involving adhesion and deformation (plowing). Adhesion and plastic deformation imply energy dissipation, which is responsible for friction (Fig. 27.36) [27.103, 104]. A contact between two bodies takes place on high asperities, and the real area of contact (A_r) is a small fraction of the apparent area of contact. During contact of two asperities, a lateral force may be required for asperities of a given slope to climb against each other. This mechanism is known as the ratchet mechanism, and it also contributes to the friction. Wear and contaminant particles present at the interface, referred as the *third body*, also contribute to the friction (Fig. 27.36). In addition, during contact even at low humidity, a meniscus is formed (Fig. 27.34). Generally, any liquid that wets or has a small contact angle on surfaces will condense from vapor into cracks and pores on surfaces as bulk liquid and in the form of annular-shaped capillary condensate in the contact zone. A quantitative theory of scale effects in friction should consider the scale effect on physical properties relevant to various contributions.

According to the adhesion and deformation model of friction, the coefficient of dry friction μ is a sum of the adhesion component μ_a and the deformation (plowing) component μ_d. The latter, in the presence of particles, is a sum of the asperity summits deformation component μ_{ds} and particles deformation component μ_{dp}, so that the total coefficient of friction is [27.136]

$$\begin{aligned}\mu &= \mu_a + \mu_{ds} + \mu_{dp} \\ &= \frac{F_a + F_{ds} + F_{dp}}{W} \\ &= \frac{A_{ra}\tau_a + A_{ds}\tau_{ds} + A_{dp}\tau_{dp}}{W} ,\end{aligned} \tag{27.17}$$

where W is the normal load, F is the friction force, A_{ra}, A_{ds}, A_{dp} are the real areas of contact during adhesion, two-body deformation and with particles, respectively, and τ is the shear strength. The subscripts a, ds, and dp correspond to adhesion, summit deformation, and particle deformation, respectively.

The adhesional component of friction depends on the real area of contact and adhesion shear strength. The real area of contact is scale-dependent due to the scale dependence of the surface roughness (for elastic and plastic contact) and due to the scale dependence of hardness (for plastic contact) [27.136]. We limit the analysis here to the multiple-asperity contact. For this case, scale L is defined as apparent size of contact between two bodies. (For completeness, for a single-asperity contact, scale is defined as the contact diameter.) It is suggested by *Bhushan* and *Nosonovsky* [27.135] that for many materials, dislocation-assisted sliding (microslip) is the main mechanism responsible for shear strength. They considered dislocation-assisted sliding based on the assumption that contributing dislocations are located in a subsurface volume. The thickness of this volume is limited by the distance which dislocations can climb ℓ_s (material parameter) and by the radius of contact a. They showed that τ_a is scale-dependent. Based on this, the adhesional components of the coefficient of friction in the case of elastic contact μ_{ae} and in the case of plastic contact μ_{ap} are given as [27.135]

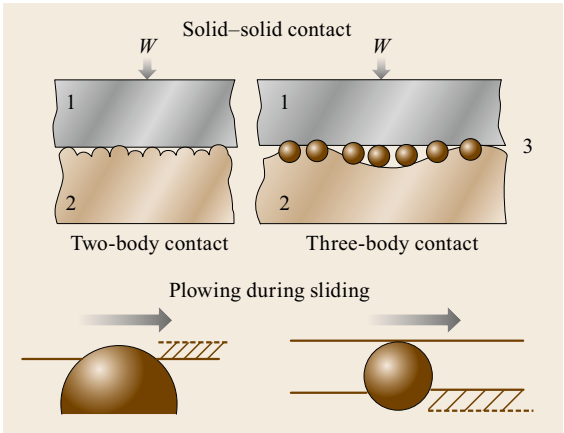


Fig. 27.36 Schematic of two- and three-body dry rough surface contacts

$$\mu_{ae} = \frac{\mu_{ae0}}{\sqrt{\ell + (\ell_s\sqrt{a_0})}} \left(\frac{L}{L_{\ell c}}\right)^{m-n} \sqrt{1 + \left(\frac{L_s}{L}\right)^m}, \quad L < L_{\ell c} \tag{27.18}$$

$$\mu_{ap} = \mu_{ap0} \sqrt{\frac{1 + (\ell_d\sqrt{a_0})}{1 + (\ell_s\sqrt{a_0})}} \sqrt{\frac{1 + (L_s/L)^m}{1 + (L_d/L)^m}}, \quad L < L_{\ell c} , \tag{27.19}$$

where μ_{ae0} and μ_{ap0} are values of the coefficient of friction at the macroscale ($L \geq L_{\ell c}$), m and n are indices which characterize the scale dependence of surface parameters, $\bar{a}_0$ is the macroscale value of mean contact radius, $L_{\ell c}$ is the long wavelength limit for scale dependence of the contact parameters, ℓ_s and ℓ_d are material-specific characteristic length parameters, and L_s and L_d are length parameters related to ℓ_s and ℓ_d. Scale dependence of the adhesional component of the

coefficient of friction is presented in Fig. 27.37, based on (27.18) and (27.19).

On the basis of the assumption that multiple asperities of two rough surfaces in contact have a conical shape, the two-body deformation component of friction can be determined as [27.103, 104]

$$\mu_{ds} = \frac{2\tan\theta_r}{\pi}\,, \tag{27.20}$$

where θ_r is the roughness angle (or attack angle) of a conical asperity. Mechanical properties affect the real area of contact and shear strength, and these cancel out in (27.16) [27.136]. On the basis of a statistical analysis of a random Gaussian surface [27.136]

$$\mu_{ds} = \frac{2\sigma_0}{\pi\beta_0 *}\left(\frac{L}{L_{\ell c}}\right)^{n-m} = \mu_{ds0}\left(\frac{L}{L_{\ell c}}\right)^{n-m}, \quad L < L_{\ell c}\,, \tag{27.21}$$

where μ_{ds0} is the value of the coefficient of summits deformation component of the coefficient of friction at macroscale ($L \geq L_{\ell c}$), and σ_0 and β_0^* are macroscale values of standard deviation of surface heights and correlation length, respectively, for a Gaussian surface. The scale dependence for the two-body deformation component of the coefficient of friction is presented in Fig. 27.38 (top curve) for $m = 0.5$, $n = 0.2$, based on (27.21). The coefficient of friction increases with decreasing scale, according to (27.21). This effect is a consequence of increasing average slope or roughness angle.

For the three-body deformation, it is assumed that wear and contaminant particles at the borders of the contact region are likely to leave the contact region, while the particles in the center are likely to stay (Fig. 27.39). The plowing three-body deformation is plastic and, assuming that particles are harder than the bodies, the shear strength τ_{dp} is equal to the shear yield strength of the softer body τ_Y, the three-body deformation component of the coefficient of friction is given by [27.137]

$$\mu_{dp} = \mu_{dp0} n_{tr} \frac{\overline{d}^2}{\overline{d}_0^2} \frac{\sqrt{1+2\ell_d\sqrt{\overline{d}}}}{\sqrt{1+2\ell_d\sqrt{\overline{d}_0}}}\,, \tag{27.22}$$

where $\overline{d}$ is the mean particle diameter, $\overline{d}_0$ is the macroscale value of mean particle diameter, n_{tr} is the number of trapped particles divided by total number of particles, and μ_{dp0} is the macroscale ($L \to \infty$, $n_{tr} \to 1$)

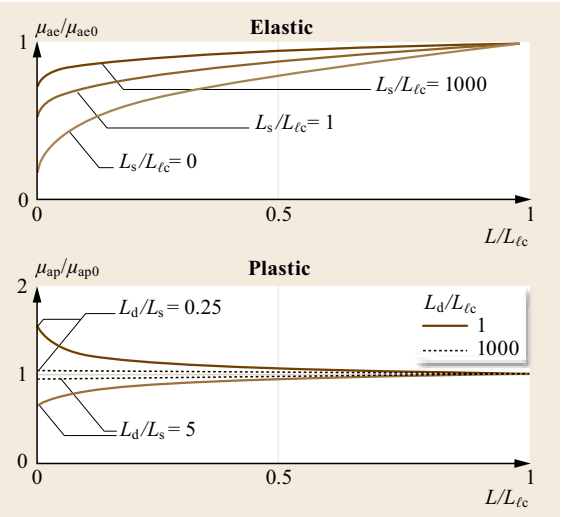


Fig. 27.37 Normalized results for the adhesional component of coefficient of friction, as a function of $L/L_{\ell c}$ for the multiple-asperity contact. Data are presented for $m = 0.5$, $n = 0.2$. For the multiple-asperity plastic contact, data are presented for two values of $L_d/L_{\ell c}$ (after [27.136])

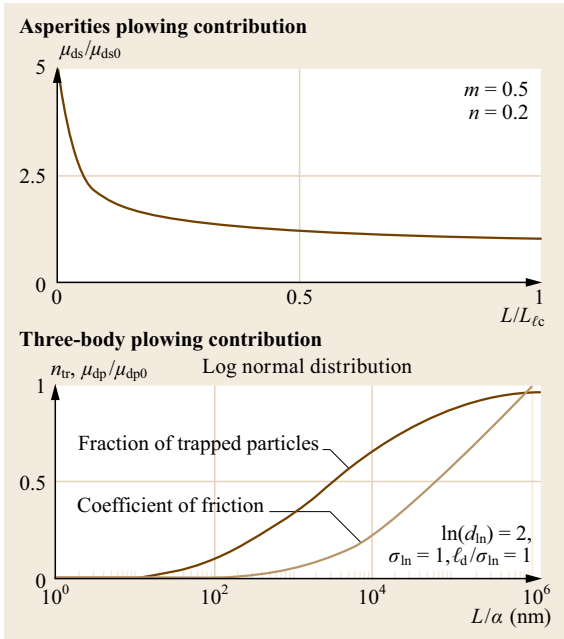


Fig. 27.38 Normalized results for the two-body deformation component of the coefficient of friction, and the number of trapped particles divided by the total number of particles and three-body deformation component of the coefficient of friction, normalized by the macroscale value for log-normal distribution of debris size, where α is the probability of a particle in the border zone to leave the contact region. Various constants given in the figure correspond to log-normal distribution (after [27.136])

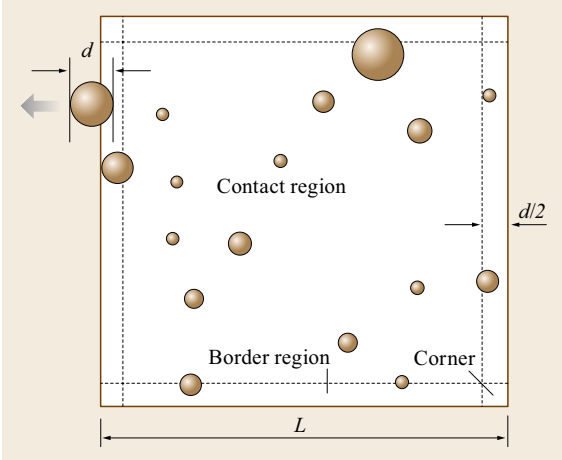


Fig. 27.39 Schematics of debris at the contact zone and at its border region. A particle of diameter d in the border region of $d/2$ is likely to leave the contact zone (after [27.136])

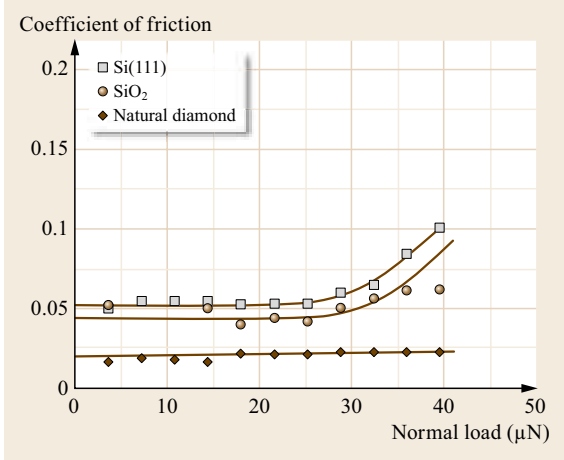


Fig. 27.40 Coefficient of friction as a function of normal load and for Si(111), SiO_2 coating and natural diamond. Inflections in the *curves* for Si and SiO_2 correspond to the contact stresses equal to the hardnesses of these materials (after [27.60])

value of the third-body deformation component of the coefficient of friction. Scale dependence of μ_{dp} is shown in Fig. 27.38 (bottom curve) based on (27.22). Based on the scale effect predictions presented in Figs. 27.37 and 27.38, trends in the experimental results in Table 27.3 can be explained.

Scale dependence of meniscus effects in friction, wear, and interface temperature can be analyzed in a similar way [27.137].

To demonstrate the load dependence of friction at the nano/microscale, the coefficient of friction as a function of normal load is presented in Fig. 27.40. The coefficient of friction was measured by *Bhushan* and *Kulkarni* [27.60] for a Si_3N_4 tip versus Si, SiO_2, and natural diamond using an AFM. They reported that for low loads, the coefficient of friction is independent of load and increases with increasing load after a certain load. It is noted that the critical value of loads for Si and SiO_2 corresponds to stresses equal to their hardness values, which suggests that the transition to plasticity plays a role in this effect. The friction values at higher loads for Si and SiO_2 approach that of macroscale values.

27.3 Micro/Nanoscale Wear and Scratching, Local Deformation, and Nanofabrication/Nanomachining

27.3.1 Nanoscale Wear

Bhushan and *Ruan* [27.24] conducted nanoscale wear tests on polymeric magnetic tapes using conventional silicon nitride tips at two different loads of 10 and 100 nN (Fig. 27.41). For a low normal load of 10 nN, measurements were made twice. There was no discernible difference between consecutive measurements for this load. However, as the load was increased from 10 to 100 nN, topographical changes were observed during subsequent scanning at a normal load of 10 nN; material was pushed in the sliding direction of the AFM tip relative to the sample. The material movement is believed to occur as a result of plastic deformation of the tape surface. Thus, deformation and movement of the soft materials on a nanoscale can be observed.

27.3.2 Microscale Scratching

The AFM can be used to investigate how surface materials can be moved or removed on micro- to nanoscales, for example, in scratching and wear [27.2, 7] (where these things are undesirable) and nanofabrication/nanomachining (where they are desirable). Figure 27.42a shows microscratches made on Si(111) at various loads and a scanning velocity of 2 μm/s after ten cycles [27.28]. As expected, the scratch depth increases linearly with load. Such microscratching measurements can be used to study failure mechanisms on

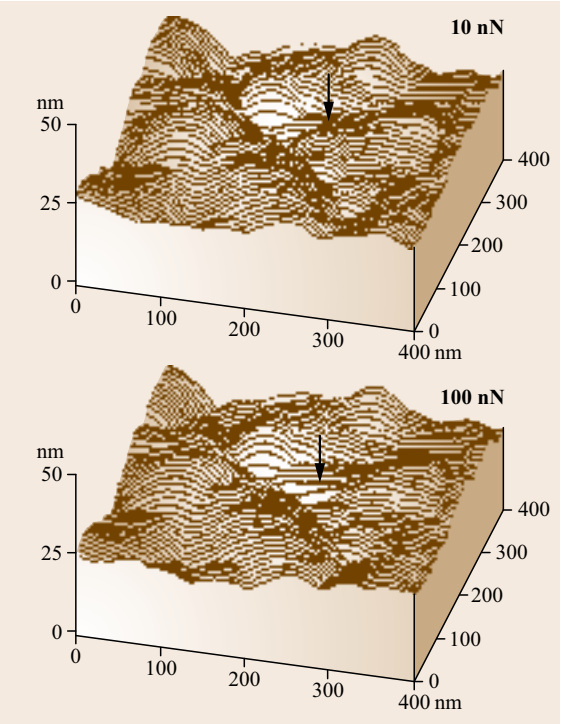


Fig. 27.41 Surface roughness maps of a polymeric magnetic tape at the applied normal load of 10 and 100 nN. Location of the change in surface topography as a result of nanowear is indicated by *arrows* (after [27.24])

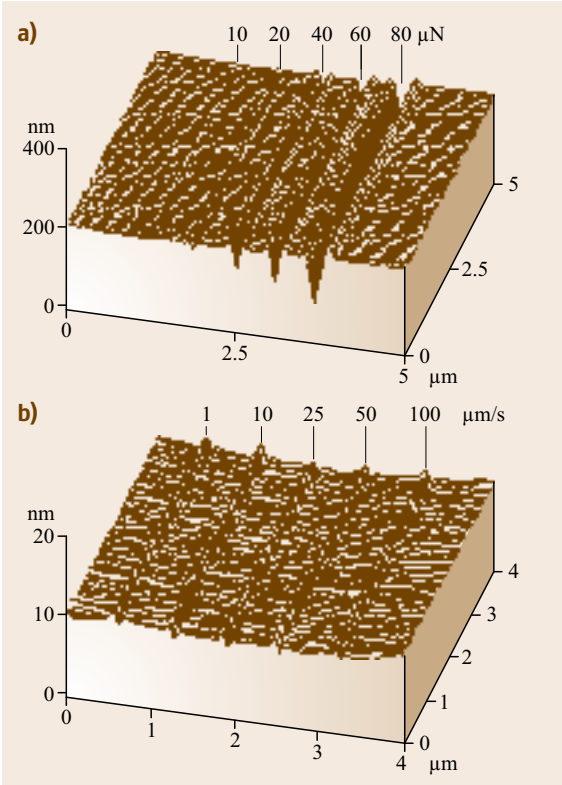


Fig. 27.42a,b Surface plots of **(a)** Si(111) scratched for ten cycles at various loads and a scanning velocity of 2 μm/s (after [27.28]). Note that *x*- and *y*-axes are in μm and the *z*-axis is in nm, and **(b)** Si(100) scratched in one unidirectional scan cycle at a normal force of 80 μN and different scanning velocities (after [27.31])

the microscale and to evaluate the mechanical integrity (scratch resistance) of ultrathin films at low loads.

To study the effect of scanning velocity, unidirectional scratches 5 μm in length were generated at scanning velocities ranging from 1 to 100 μm/s at various normal loads ranging from 40 to 140 μN [27.31]. There is no effect of scanning velocity obtained at a given normal load. Representative scratch profiles at 80 μN are shown in Fig. 27.42b. This may be because of a small effect of frictional heating with the change in scanning velocity used here. Furthermore, for a small change in interface temperature, there is a large underlying volume to dissipate the heat generated during scratching.

Scratching can be performed under ramped loading to determine the scratch resistance of materials and coatings [27.58]. The coefficient of friction is measured during scratching, and the load at which the coefficient of friction increases rapidly is known as the *critical load*, which is a measure of scratch resistance. In addition, postscratch imaging can be performed in situ with the AFM in tapping mode to study failure mechanisms. Figure 27.43 shows data from a scratch test on Si(100) with a scratch length of 25 μm and a scratching velocity of 0.5 μm/s. At the beginning of the scratch, the coefficient of friction is 0.04, which indicates a typical value for silicon. At about 35 μN (indicated by the arrow in the figure), there is a sharp increase in the coefficient of friction, which indicates the critical load. Beyond the critical load, the coefficient of friction continues to increase steadily. In the postscratch image, we note that at the critical load, a clear groove starts to form. This implies that Si(100) was damaged by plowing at the critical load, associated with the plastic flow of the material. At and after the critical load, small and uniform debris is observed, and the amount of debris increases with increasing normal load. *Sundararajan* and *Bhushan* [27.58] have also used this technique to measure the scratch resistance of diamond-like carbon coatings ranging in thickness from 3.5–20 nm.

27.3.3 Microscale Wear

By scanning the sample in two dimensions with the AFM, wear scars are generated on the surface. Fig-

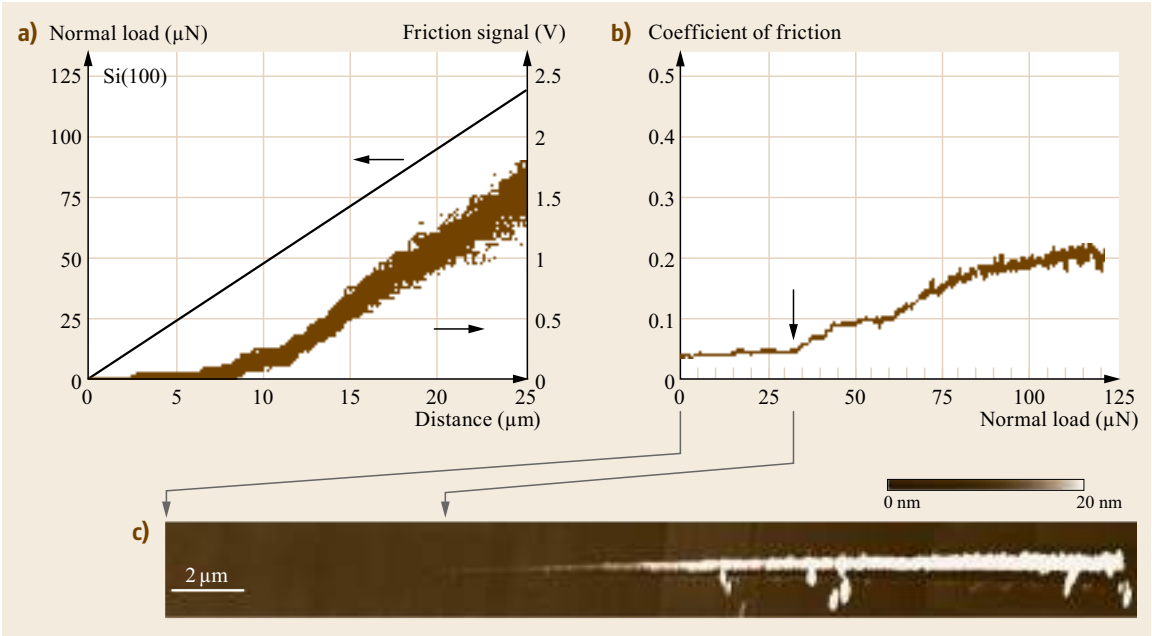


Fig. 27.43 **(a)** Applied normal load and friction signal measured during the microscratch experiment on Si(100) as a function of scratch distance, **(b)** friction data plotted in the form of coefficient of friction as a function of normal load, and **(c)** AFM surface height image of scratch obtained in tapping mode (after [27.58])

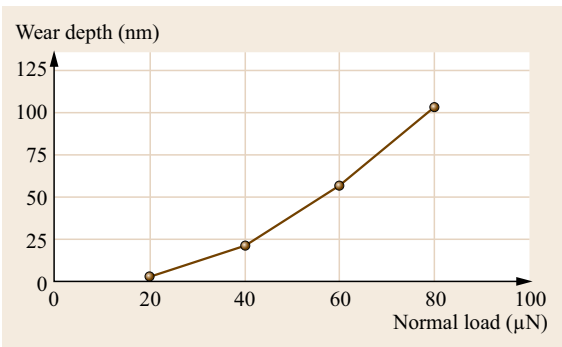


Fig. 27.44 Wear depth as a function of normal load for Si(100) after one cycle (after [27.138])

ure 27.44 shows the effect of normal load on wear depth on Si(100). We note that wear depth is very small below 20 μN of normal load [27.100, 138]. A normal load of 20 μN corresponds to contact stresses comparable to the hardness of silicon. Primarily, elastic deformation at loads below 20 μN is responsible for low wear [27.29, 30].

A typical wear mark of the size 2 μm × 2 μm generated at a normal load of 40 μN for one scan cycle and imaged using AFM with a scan size of 4 μm × 4 μm at 300 nN load is shown in Fig. 27.45a. The inverted map of wear marks shown in Fig. 27.45b indicates the uniform material removal at the bottom of the wear mark [27.100]. An AFM image of the wear mark shows debris at the edges, probably swiped during AFM scanning. This indicates that the debris is loose (not sticky) and can be removed during AFM scanning.

Next the mechanism of material removal on the microscale in AFM wear experiments is examined [27.31, 100, 138]. Figure 27.46 shows a secondary electron image of the wear mark and associated wear particles. The specimen used for the scanning electron microscope (SEM) was not scanned with the AFM after initial wear, in order to retain wear debris in the wear region. Wear debris is clearly observed. In the SEM micrographs, the wear debris appear to be agglomerated because of the high surface energy of the fine particles. Particles appear to be a mixture of rounded and so-called cutting type (feather-like or ribbon-like material). *Zhao* and *Bhushan* [27.138] reported an increase in the number and size of cutting-type particles with the normal load. The presence of cutting-type particles indicates that the material is removed primarily by plastic deformation.

To better understand the material removal mechanisms, *Zhao* and *Bhushan* [27.138] used transmission electron microscopy (TEM). The TEM micrograph of the worn region and associated diffraction pattern are shown in Fig. 27.47a,b. The bend contours are observed to pass through the wear mark in the micrograph. The bend contours around and inside the wear mark are indicative of a strain field, which in the absence of ap-

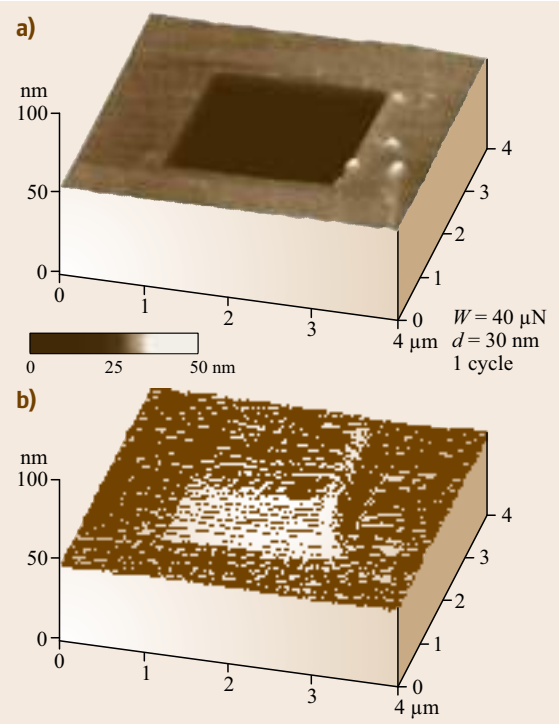


Fig. 27.45 **(a)** Typical gray-scale and **(b)** inverted AFM images of a wear mark created using a diamond tip at a normal load of 40 µN and one scan cycle on a Si(100) surface (after [27.100])

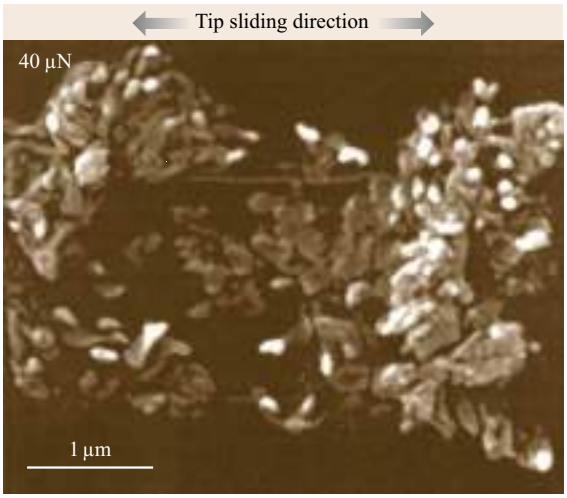


Fig. 27.46 Secondary electron image of wear mark and debris for Si(100) produced at a normal load of 40 µN and one scan cycle (after [27.138])

plied stresses can be interpreted as plastic deformation and/or elastic residual stresses. Often, localized plastic deformation during loading would lead to residual

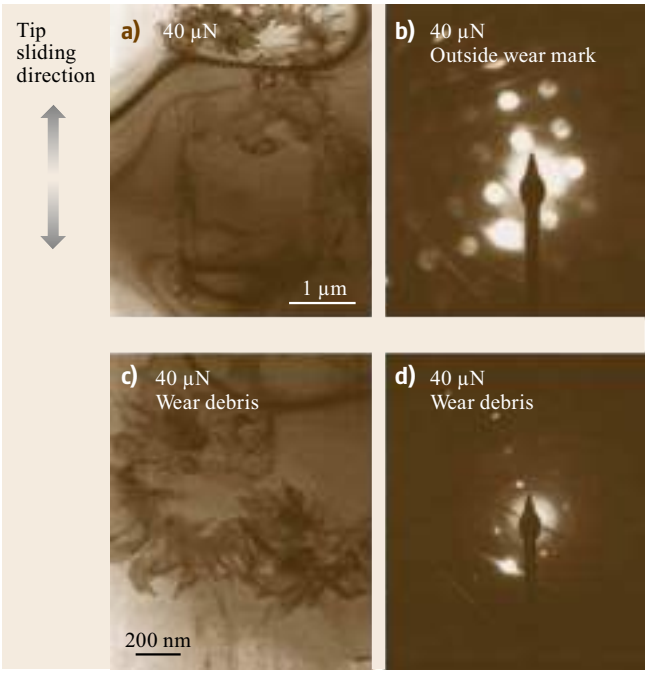


Fig. 27.47a–d Bright-field TEM micrographs **(a)** and diffraction patterns **(b)** of wear mark **(a,b)** and wear debris **(c,d)** in Si(100) produced at a normal load of 40 µN and one scan cycle. Bend contours around and inside wear mark are observed (after [27.138])

stresses during unloading; therefore, bend contours reflect a mix of elastic and plastic strains. The wear debris is observed outside the wear mark. The enlarged view of the wear debris in Fig. 27.47c shows that much of the debris is ribbon like, indicating that material is removed by a cutting process via plastic deformation, which is consistent with the SEM observations. The diffraction pattern from inside the wear mark was reported to be similar to that of virgin silicon, indicating no evidence of any phase transformation (amorphization) during wear. A selected area diffraction pattern of the wear debris shows some diffuse rings, which indicates the existence of amorphous material in the wear debris, confirmed as silicon oxide products from chemical analysis. It is known that plastic deformation occurs by generation and propagation of dislocations. No dislocation activity or cracking was observed at 40 µN. However, dislocation arrays could be observed at 80 µN. Figure 27.48 shows the TEM micrographs of the worn region at 80 µN; for better observation of the worn surface, wear debris was moved out of the wear mark by using AFM with a large area scan at 300 nN after the wear test. The existence of dislocation arrays confirms that material removal occurs by plastic deformation. This corroborates the observations made in scratch tests at ramped load in the previous section. It

is concluded that the material on the microscale at high loads is removed by plastic deformation with a small contribution from elastic fracture [27.138].

To understand wear mechanisms, evolution of wear can be studied using AFM. Figure 27.49 shows the evolution of wear marks of a DLC-coated disk sample. The data illustrate how the microwear profile for a load of 20 μN develops as a function of the number of scanning cycles [27.28]. Wear is not uniform, but is initiated at the nanoscratches. Surface defects (with high surface energy) present at the nanoscratches act as initiation sites for wear. Coating deposition also may not be uniform on and near nanoscratches which may lead to coating delamination. Thus, scratch-free surfaces will be relatively resistant to wear.

Wear precursors (precursors to measurable wear) can be studied by making surface potential measurements [27.81, 83, 84]. The contact potential difference, or simply the surface potential between two surfaces depends on a variety of parameters such as electronic work function, adsorption, and oxide layers. The surface potential map of an interface gives a measure of changes in the work function which is sensitive to both physical and chemical conditions of the surfaces including structural and chemical changes. Before material is actually removed in a wear process, the surface experiences stresses that result in surface and subsurface changes of structure and/or chemistry. These can cause changes in the measured potential of a surface. An AFM tip allows the mapping of the surface potential with nanoscale resolution. Surface height and change in surface potential maps of a polished single-crystal aluminum (100) sample, abraded using a diamond tip at loads of 1 and 9 μN, are shown in Fig. 27.50a. It is evident that both abraded regions show a large potential contrast (≈ 0.17 V), with respect to the nonabraded area. The black region in the lower right-hand part of the

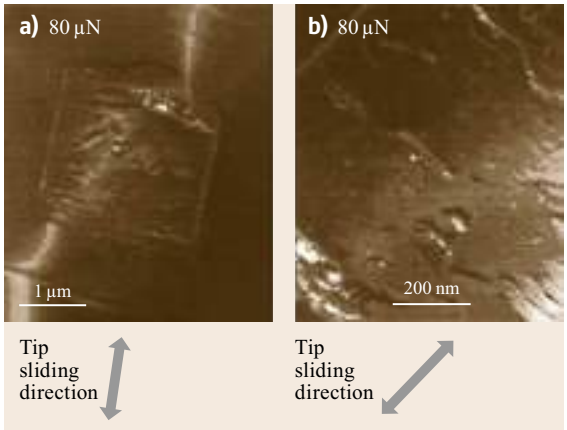


Fig. 27.48 (a) Bright-field and (b) weak-beam TEM micrographs of wear mark in Si(100) produced at a normal load of 80 μN and one scan cycle showing bend contours and dislocations (after [27.138])

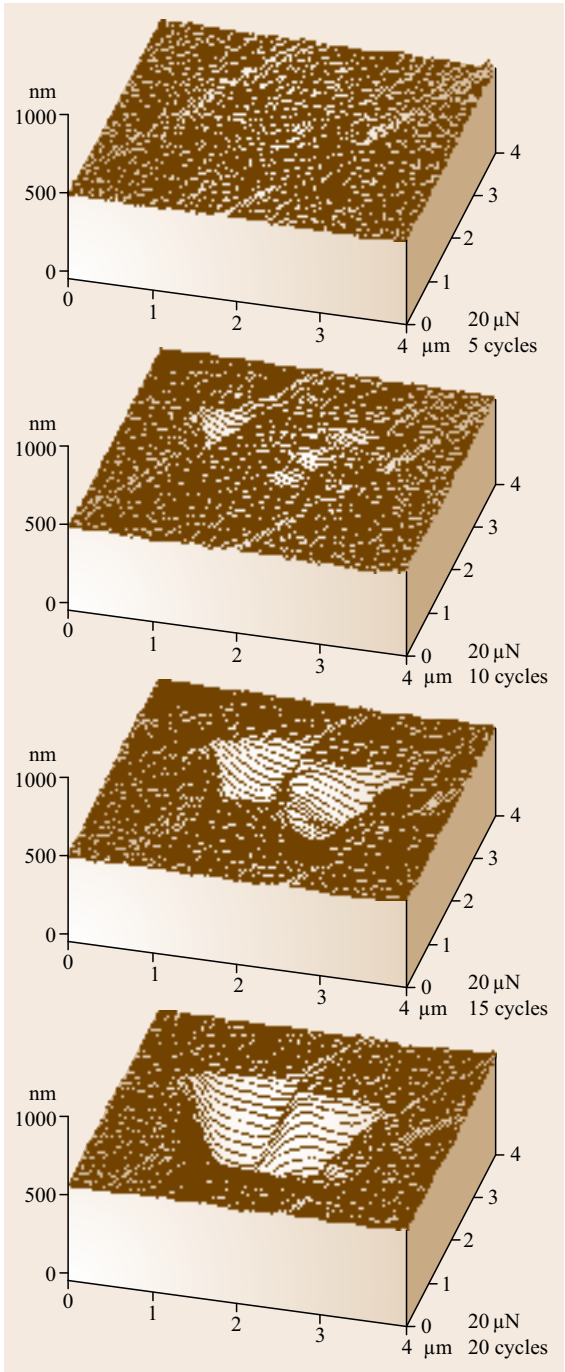


Fig. 27.49 Surface plots of a diamond-like carbon-coated thin-film disk showing the worn region; the normal load and number of test cycles are indicated (after [27.28])

topography scan shows a step that was created during the polishing phase. There is no potential contrast between the high region and the low region of the sample, indicating that the technique is independent of surface height. Figure 27.50b shows a close-up scan of the upper (low load) wear region in Fig. 27.50a. Notice that while there is no detectable change in the surface topography, there is nonetheless, a large change in the potential of the surface in the worn region. Indeed, the wear mark of Fig. 27.50b might not be visible at all in the topography map were it not for the noted absence of wear debris generated nearby and then swept off during the low load scan. Thus, even in the case of zero wear (no measurable deformation of the surface using AFM), there can be a significant change in the surface potential inside the wear mark which is useful for the study of wear precursors. It is believed that the removal of the thin contaminant layer including the natural oxide layer gives rise to the initial change in surface potential. The structural changes, which precede the generation of wear debris and/or measurable wear scars, occur under ultralow loads in the top few nanometers of the sample, and are primarily responsible for the subsequent changes in surface potential.

Surface potential measurements using the nano-Kelvin probe technique have been used to study wear and degradation for various applications such as for human skin [27.139], human hair [27.107, 140], Li-ion batteries [27.141] and MEMS switches [27.142].

27.3.4 In situ Characterization of Local Deformation

In situ surface characterization of local deformation of materials and thin films is carried out using a tensile stage inside an AFM [27.73–75, 85, 86]. As an example, failure mechanisms of coated polymeric thin films under tensile load were studied by *Bobji* and *Bhushan* [27.73, 74]. The specimens were strained at a rate of 4×10^{-3}%/s, and AFM images were captured at different strains up to about 10% to monitor generation and propagation of cracks and deformation bands. *Bobji* and *Bhushan* [27.73, 74] studied three magnetic tapes of thickness ranging from 7 to 8.5 μm. One of these had an acicular-shaped metal particle (MP) coating while the other two had metal-evaporated (ME) coatings with and without a thin diamond-like carbon (DLC) overcoat both on a polymeric substrate; all had a particulate back coating [27.16]. They also studied a polyethylene terephthalate (PET) substrate with 6 μm thickness. They reported that cracking of the coatings started at about 1% strain for all tapes much before the substrate starts to yield at about 2% strain. Figure 27.51 shows the topographical images of the MP tape at different strains. At 0.83% strain, a crack can be seen, originating at the marked point. As the tape is further stretched along the direction, as shown in Fig. 27.51, the crack propagates along the shorter boundary of the ellipsoidal particle. However, the general direction of crack propagation remains perpendicular to the direction of stretching. The length, width, and depth of the cracks increase with strain, and at the same time newer cracks keep nucleating and propagating with reduced crack spacing. At 3.75% strain, another crack can be seen nucleating. This crack continues to grow parallel to the first one. When the tape is unloaded after stretching up to a strain of about 2%, i. e., within the elastic limit of the substrate, the cracks close perfectly, and it is impossible to determine the difference from the unstrained tape.

Figure 27.52 shows topographical images of the three magnetic tapes and the PET substrate after being strained to 3.75%, which is well beyond the elastic

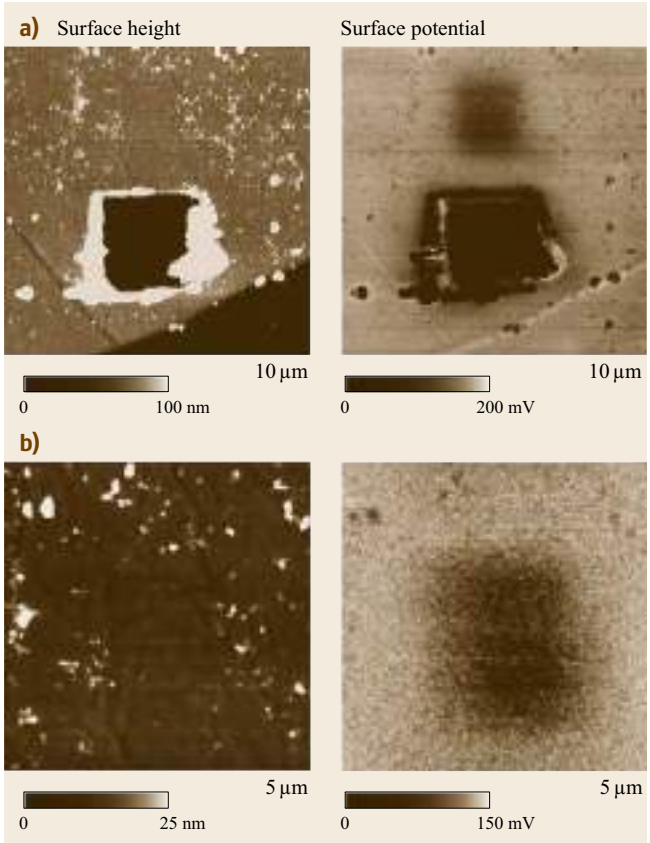


Fig. 27.50 (a) Surface height and change in surface potential maps of wear regions generated at 1 μN (*left*) and 9 μN (*right*) on a single-crystal aluminum sample showing bright contrast in the surface potential map on the worn regions. (b) Close up of upper (low load) wear region (after [27.83])

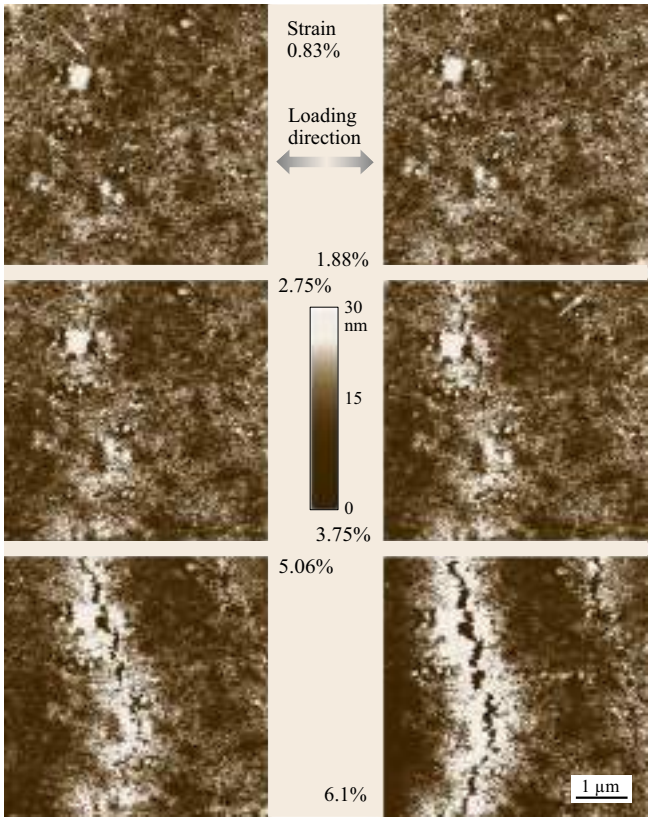


Fig. 27.51 Topographical images of the MP magnetic tape at different strains (after [27.73])

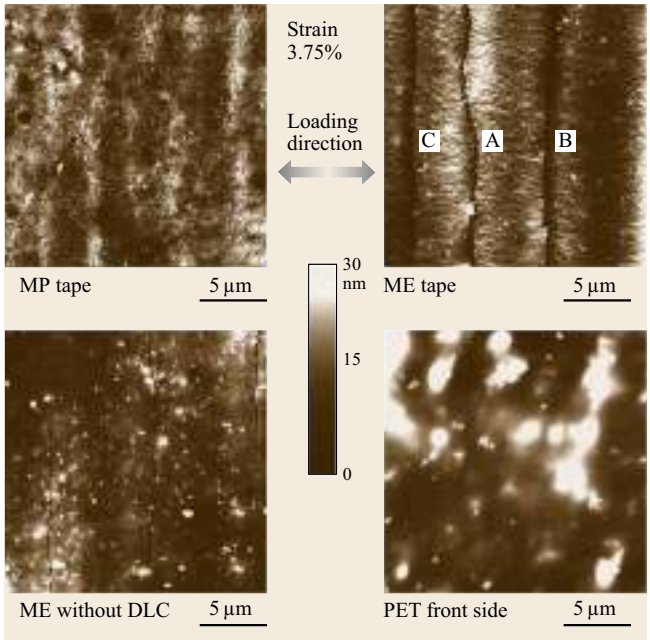


Fig. 27.52 Comparison of crack morphologies at 3.75% strain in three magnetic tapes and for a PET substrate. Cracks B and C, nucleated at higher strains, are more linear than crack A (after [27.74])

limit of the substrate. The MP tape develops short and numerous cracks perpendicular to the direction of loading. In tapes with metallic coating, the cracks extend throughout the tape width. In the ME tape with the DLC coating, there is a bulge in the coating around the primary cracks that are initiated when the substrate is still elastic, like crack A in the figure. The white band on the right-hand side of the figure is the bulge of another crack. The secondary cracks like B and C are generated at higher strains and are straighter compared to the primary cracks. In the ME tape which has a Co-O film on a PET substrate, with a thickness ratio of 0.03, both with and without DLC coating, no difference is observed in the rate of growth between primary and secondary cracks. The failure is cohesive with no bulging of the coating. This seems to suggest that the DLC coating has residual stresses that relax when the coating cracks, causing delamination. Since the stresses are already relaxed, the secondary crack does not result in delamination. The presence of the residual stress is confirmed by the fact that a free-standing ME tape curls up (in a cylindrical form with its axis perpendicular to the tape length) with a radius of curvature of about 6 mm, and the ME tape without the DLC does not curl. The magnetic coating side of the PET substrate is much smoother at smaller scan lengths. However, in the 20 μm scans it has a lot of bulging out, which appears as white spots in the figure. These spots change shape even when the samples are scanned in tapping mode at very low contact forces.

The variation of average crack width and average crack spacing with strain is plotted in Fig. 27.53. The crack width is measured at a spot along a given crack over a distance of 1 μm in the 5 μm scan image at different strains. The crack spacing is obtained by averaging the intercrack distance measured in five separate 50 μm scans at each strain. It can be seen that the cracks nucleate at a strain of about 0.7%–1.0%, well within the elastic limit of the substrate. There is a definite change in the slope of the load–displacement curve at the strain where cracks nucleate, and the slope after that is closer to the slope of the elastic portion of the substrate. This would mean that most of the load is supported by the substrate once the coating fails by cracking.

Fatigue experiments can be performed by applying a cyclic stress amplitude with a certain mean stress [27.75]. Fatigue life was determined by the first occurrence of cracks. Experiments were performed at various constant mean stresses and with a range of cyclic stress amplitudes for each mean stress value for various magnetic tapes. Number of cycles to failure was plotted as a function of stress state to obtain a so-called *S–N* (stress–life) diagram. As the stress is decreased,

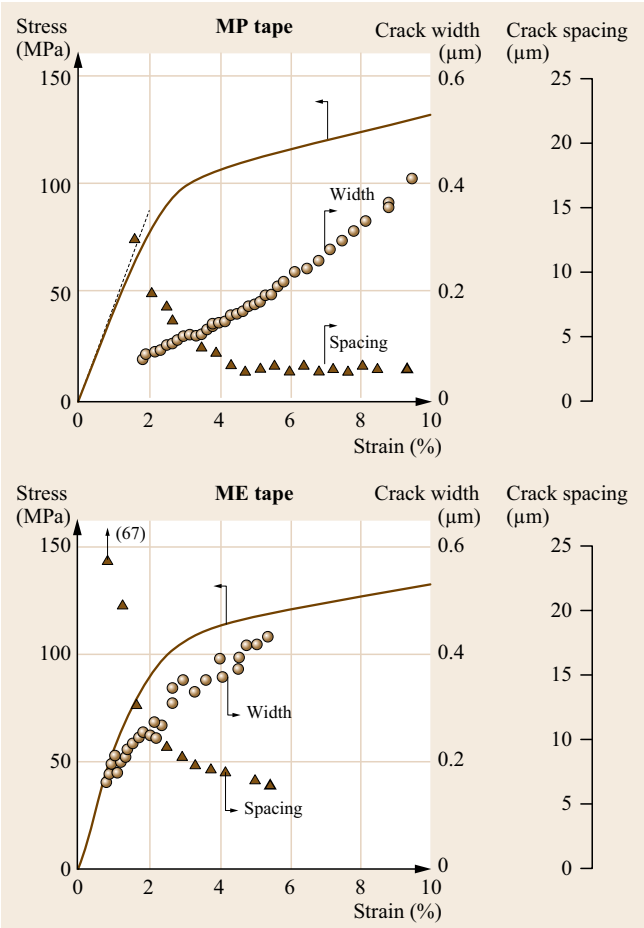


Fig. 27.53 Variation of stress, crack width, and crack spacing with strain in two magnetic tapes (after [27.73])

there is a stress value for which no failure occurs. This stress is termed the endurance limit or simply fatigue limit. Figure 27.54 shows the *S–N* curve for an ME tape and an ME tape without DLC. For the ME tape, the endurance limit is seen to go down with a decrease in mean stress. This is consistent with the literature, and is because at a lower mean stress the corresponding stress amplitude is relatively high and this causes failure. The endurance limit is found to be almost the same for all three mean stresses. In the case of the ME tape without DLC, the critical number of cycles is found to also be in the same range.

In situ surface characterization of unstretched and stretched films has been used to measure Poisson's ratio of polymeric thin films by *Bhushan* et al. [27.95]. Uniaxial tension is applied by the tensile stage. Surface height profiles obtained from the AFM images of unstretched and stretched samples are used to monitor the changes in displacements of the polymer films

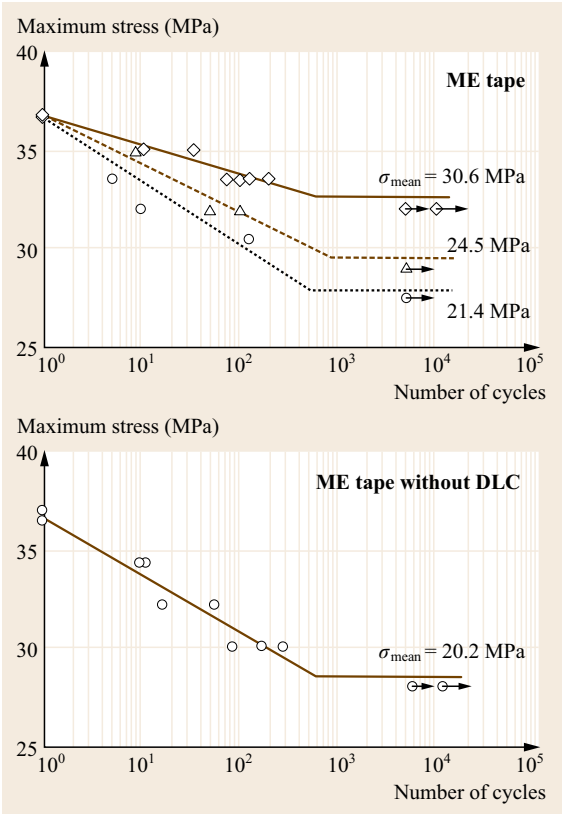


Fig. 27.54 *S–N* curve for two magnetic tapes with maximum stress plotted on the *ordinate* and number of cycles to failure on the *abscissa*. The data points marked with *arrows* indicate tests for which no failure (cracking) was observed in the scan area, even after a large number of cycles (10 000) (after [27.75])

in the longitudinal and lateral directions simultaneously.

27.3.5 Nanofabrication/Nanomachining

An AFM can be used for nanofabrication/nanomachining by extending the microscale scratching operation [27.2, 20, 28, 79]. Figure 27.55 shows two examples of nanofabrication. The patterns were created on a single-crystal silicon (100) wafer by scratching the sample surface with a diamond tip at specified locations and scratching angles. Each line is inscribed manually at a normal load of 15 μN and a writing speed of 0.5 μm/s. The separation between lines is about 50 nm, and the variation in line width is due to the tip asymmetry. Nanofabrication parameters – normal load, scanning speed, and tip geometry – can be controlled precisely to control the depth and length of the devices.

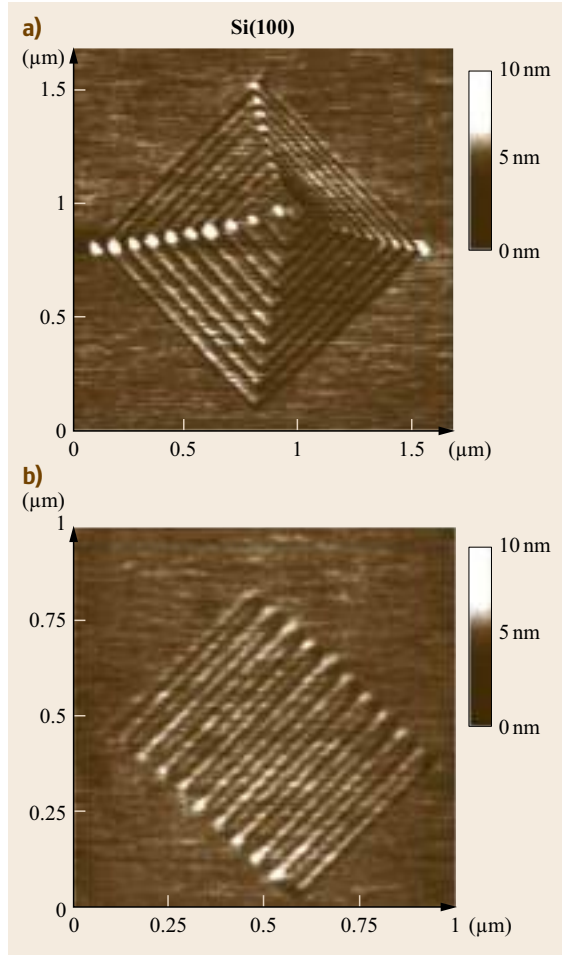


Fig. 27.55 **(a)** Trim and **(b)** spiral patterns generated by scratching an Si(100) surface using a diamond tip at a normal load of 15 μN and writing speed of 0.5 μm/s (after [27.2, 7])

Nanofabrication using mechanical scratching has several advantages over other techniques. Better control over the applied normal load, scan size, and scanning speed can be used for the nanofabrication of devices. Using the technique, nanofabrication can be performed on any engineering surface. Chemical etching or reactions is not required, and this dry nanofabrication process can be employed where the use of chemicals and electric field is prohibited. One disadvantage of this technique is the formation of debris during scratching. At light loads, debris formation is not a problem compared to high-load scratching. However, debris can be removed easily from the scan area at light loads during scanning.

27.4 Indentation

Mechanical properties on the relevant scales are needed for the analysis of friction and wear mechanisms. Mechanical properties, such as hardness and Young's modulus of elasticity, can be determined on the micro- to picoscales using the AFM [27.24, 28, 55, 61] and a depth-sensing indentation system used in conjunction with an AFM [27.60, 143–145].

27.4.1 Picoindentation

Indentability on the scale of subnanometers of soft samples can be studied in the force calibration mode (Fig. 27.7) by monitoring the slope of cantilever deflection as a function of the sample traveling distance after the tip is engaged and the sample is pushed against the tip. For a rigid sample, the cantilever deflection equals the sample traveling distance, but the former quantity is smaller if the tip indents the sample. In an example for a polymeric magnetic tape shown in Fig. 27.56, the line in the left portion of the figure is curved with a slope of less than 1 shortly after the sample touches the tip, which suggests that the tip has indented the sample [27.24]. Later, the slope is equal to 1, suggesting that the tip no longer indents the sample. This observation indicates that the tape surface is soft locally (polymer rich) but hard (as a result of magnetic particles) underneath. Since the curves in extending and retracting modes are identical, the indentation is elastic up to a maximum load of about 22 nN used in the measurements.

Detection of the transfer of material on a nanoscale is possible with the AFM. Indentation of C_{60}-rich fullerene films with an AFM tip has been shown [27.59] to result in the transfer of fullerene molecules to the AFM tip, as indicated by discontinuities in the cantilever deflection as a function of the sample traveling distance in subsequent indentation studies.

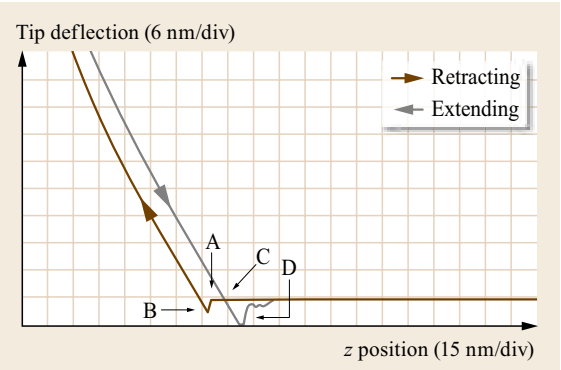


Fig. 27.56 Tip deflection (normal load) as a function of the z (separation distance) curve for a polymeric magnetic tape (after [27.24])

27.4.2 Nanoindentation

The indentation hardness of surface films with an indentation depth of as small as about 1 nm can be measured using an AFM [27.1, 16, 61]. Figure 27.57 shows the gray-scale plots of indentation marks made on Si(111) at normal loads of 60, 65, 70, and 100 μN. Triangular indents can be clearly observed with very shallow depths. At a normal load of 60 μN, indents are observed, and the depth of penetration is about 1 nm. As the normal load is increased, the indents become clearer, and indentation depth increases. For the case of hardness measurements at shallow depths on the same order as variations in surface roughness, it is desirable to subtract the original (unindented) map from the indent map for an accurate measurement of the indentation size and depth [27.28].

To make accurate measurements of hardness at shallow depths, a depth-sensing nano/picoindentation system (Fig. 27.9) is used [27.60]. Figure 27.58 shows the load–displacement curves at different peak loads for Si(100). Loading/unloading curves often exhibit sharp discontinuities, particularly at high loads. Discontinuities, also referred to as pop-ins, occurring during the initial loading part of the curve mark a sharp transition from pure elastic loading to a plastic deformation of the specimen surface, thus corresponding to an initial yield point. The sharp discontinuities in the unloading part of the curves are believed to be due to the formation of lateral cracks which form at the base of the median crack, which results in the surface of the specimen being thrust upward. Load–displacement data at residual depths as low as about 1 nm can be obtained. The indentation hardness of surface films has been measured for various materials at a range of loads including Si(100) up to a peak load of 500 μN and Al(100) up to a peak load of 2000 μN by *Bhushan* et al. [27.60] and *Kulkarni* and *Bhushan* [27.3, 16]. The hardnesses of single-crystal silicon and single-crystal aluminum at shallow depths on the order of a few nm (on a nanoscale) are found to be higher than at depths on the order of a few hundred nm (on a microscale) (Fig. 27.59). Microhardness has also been reported to be higher than that on the millimeter scale by several investigators. The data reported to date show that hardness exhibits a scale (size) effect.

During loading, the generation and propagation of dislocations are responsible for plastic deformation. A strain-gradient plasticity theory has been devel-

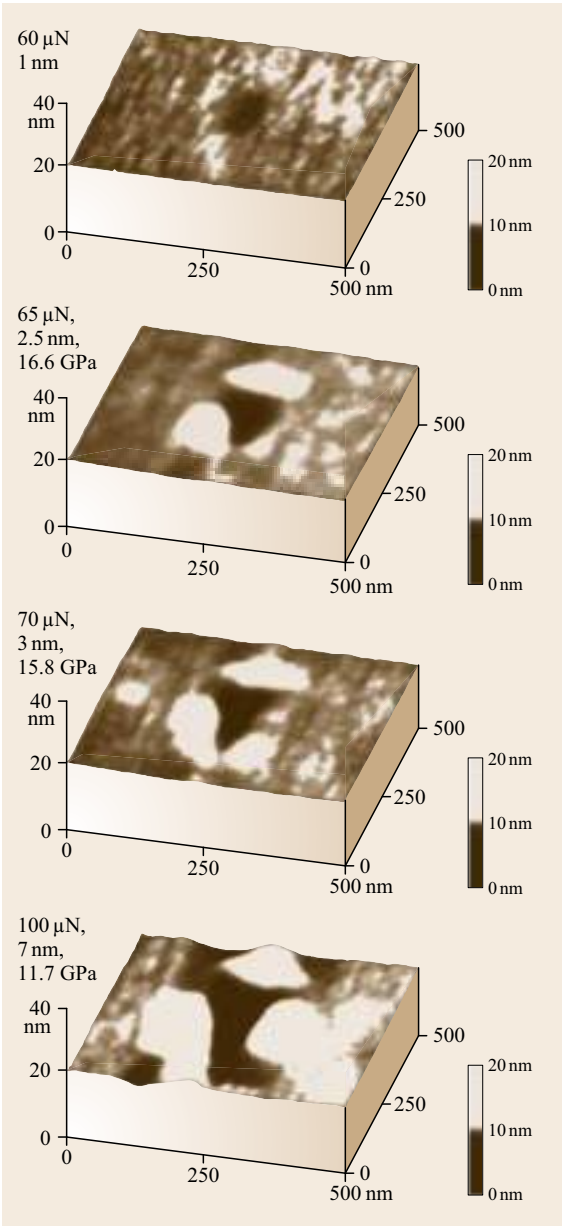


Fig. 27.57 Gray-scale plots of indentation marks on the Si(111) sample at various indentation loads. Loads, indentation depths and hardness values are listed in the figure (after [27.61])

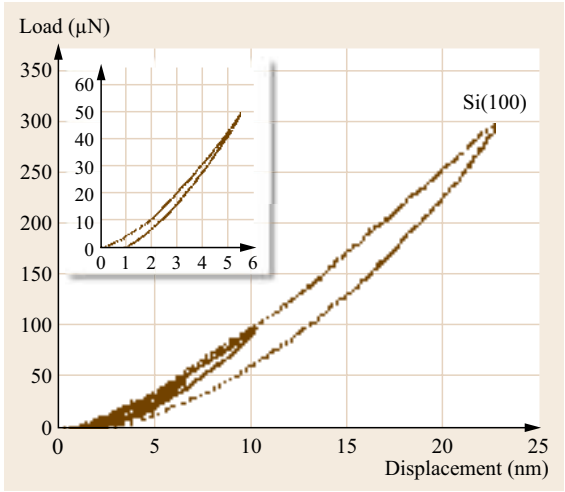


Fig. 27.58 Load–displacement curves at various peak loads for Si(100); *inset* shows the magnified curve for a peak load of 50 μN (after [27.60])

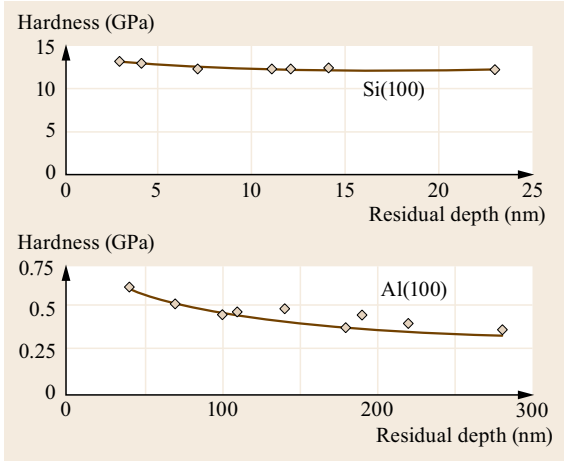


Fig. 27.59 Indentation hardness as a function of residual indentation depth for Si(100) (after [27.60]), and Al(100) (after [27.143])

oped for micro/nanoscale deformations, and is based on randomly created statistically stored and geometrically necessary dislocations [27.146, 147]. Large strain gradients inherent in small indentations lead to the accumulation of geometrically necessary dislocations, located in a certain sub-surface volume, for strain compatibility reasons. The large strain gradients in small indentations require these dislocations to account for the large slope at the indented surface. These dislocations become obstacles to other dislocations that cause enhanced hardening. These are a function of strain gradient, whereas statistically, stored dislocations are a function of strain. Based on this theory, scale-dependent hardness is given as

$$H = H_0 \sqrt{1 + \frac{\ell_d}{a}} \,, \tag{27.23}$$

where H_0 is the hardness in the absence of strain gradient or macrohardness, ℓ_d is the material-specific

characteristic length parameter, and a is the contact radius. In addition to the role of the strain-gradient plasticity theory, an increase in hardness with a decrease in indentation depth can possibly be rationalized on the basis that as the volume of deformed material decreases, there is a lower probability of encountering material defects.

Bhushan and *Koinkar* [27.55] used AFM measurements to show that ion implantation of silicon surfaces increases their hardness and thus their wear resistance. Formation of surface alloy films with improved mechanical properties by ion implantation is of growing technological importance as a means of improving the mechanical properties of materials. Hardness of 20 nm thick DLC films have been measured by *Kulkarni* and *Bhushan* [27.145].

To measure variations in mechanical properties of layered materials and nonhomogeneous composites, a so-called continuous stiffness technique is used [27.62]. The technique offers a direct measurement of dynamic contact stiffness during the loading portion of an indentation test. This allows the measurement of contact stiffness, elastic modulus, hardness, and fatigue properties. For example, the technique has been used in layered and nonhomogeneous magnetic tapes.

The creep and strain rate effects (viscoelastic effects) of ceramics can be studied using a depth-sensing indentation system. *Bhushan* et al. [27.60] and *Kulkarni* and *Bhushan* [27.143–145] have reported that ceramics (single-crystal silicon and diamond-like carbon) exhibit significant plasticity and creep on a nanoscale. Figure 27.60a shows the load–displacement curves for single-crystal silicon at various peak loads held at 180 s. To demonstrate the creep effects, the load–displacement curves for a 500 μN peak load held at 0 and 30 s are also shown as an inset. Note that significant creep occurs at room temperature. Nanoindenter experiments conducted by *Li* et al. [27.148] exhibited significant creep only at high temperatures (greater than or equal to 0.25 times the melting point of silicon). The mechanism of dislocation glide plasticity is believed to dominate the indentation creep process on the macroscale. To study the strain rate sensitivity of silicon, data at two different (constant) rates of loading are presented in Fig. 27.60b. Note that a change in the loading rate by a factor of about five results in a significant change in the load–displacement data. The viscoelastic effects observed here for silicon at ambient temperature could arise from the size effects mentioned earlier. Most likely, creep and strain rate experiments are being conducted on the hydrated films present on the silicon surface in an ambient environment, and these films are expected to be viscoelastic.

27.4.3 Localized Surface Elasticity and Viscoelasticity Mapping

The Young's modulus of elasticity can be calculated from the slope of the indentation curve during unloading. However, these measurements provide a single-point measurement. By using the force modulation technique, it is possible to obtain localized elasticity maps of soft and compliant materials of near-surface regions with nanoscale lateral resolution. This technique has been successfully used for polymeric magnetic tapes, which consist of magnetic and nonmagnetic ce-

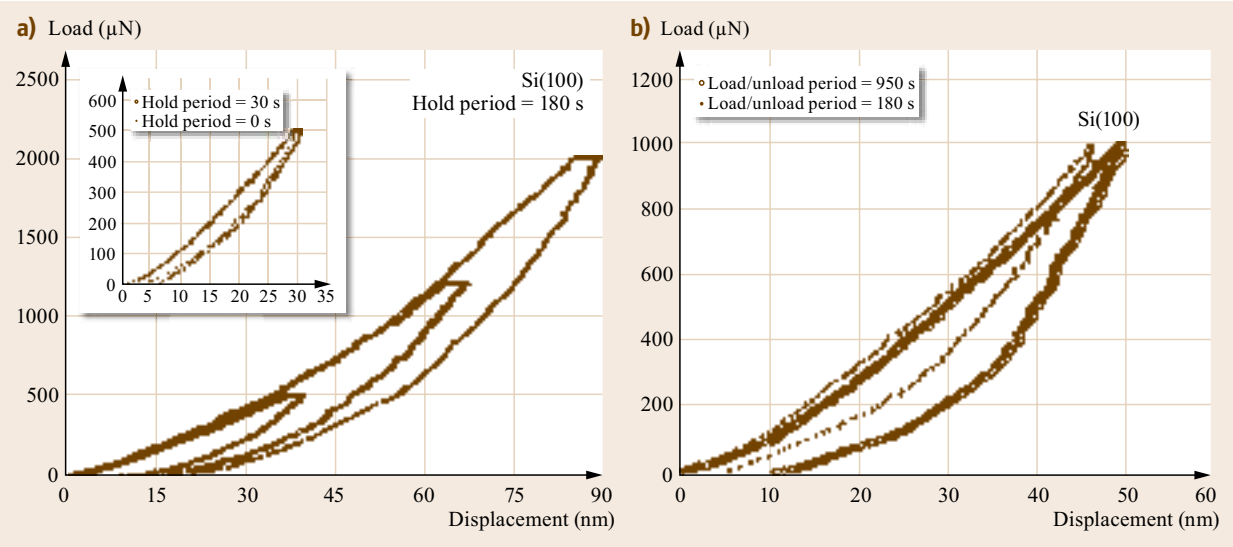


Fig. 27.60 **(a)** Creep behavior and **(b)** strain-rate sensitivity of Si(100) (after [27.60])

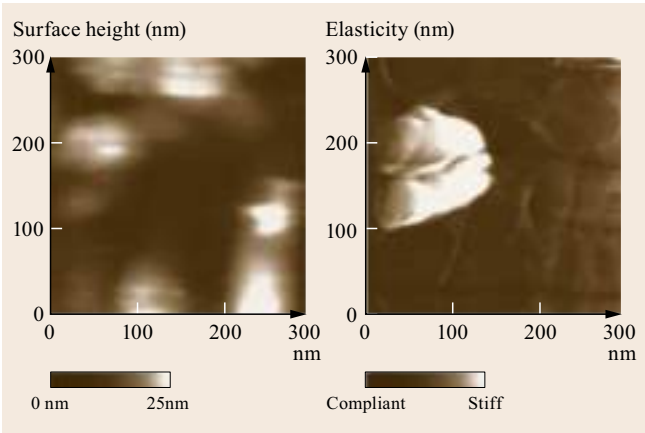


Fig. 27.61 Surface height and elasticity maps on a polymeric magnetic tape ($\sigma = 6.7$ nm and P–V $= 32$ nm; σ and P–V refer to standard deviation of surface heights and peak-to-valley distance, respectively). The gray scale on the elasticity map is arbitrary (after [27.66])

ramic particles in a polymeric matrix. Elasticity maps of a tape can be used to identify the relative distribution of the hard magnetic and nonmagnetic ceramic particles on the tape surface, which has an effect on friction and stiction at the head–tape interface [27.16]. Figure 27.61 shows the surface height and elasticity maps on a polymeric magnetic tape [27.66]. The elasticity image reveals sharp variations in the surface elasticity due to the composite nature of the film. As can be clearly seen, regions of high elasticity do not always correspond to high or low topography. On the basis of a Hertzian elastic-contact analysis, the static indentation depth of these samples during the force modulation scan is estimated to be about 1 nm. We conclude that the contrast seen is influenced most strongly by material properties in the top few nanometers, independent of the composite structure beneath the surface layer.

By using phase contrast microscopy, it is possible to obtain phase contrast maps or the contrast in viscoelastic properties of near surface regions with nanoscale lateral resolution. This technique has been successfully used for polymeric films and magnetic tapes which consist of ceramic particles in a polymeric matrix [27.69–72].

Figure 27.62 shows typical surface height, TR amplitude, and TR phase angle images for a MP tape using TR mode II, described earlier. The TR amplitude image provides contrast in lateral stiffness, and the TR phase angle image provides contrast in viscoelastic properties. In TR amplitude and phase angle images, the distribution of magnetic particles can be clearly seen and have better contrast than that in the TR surface height image. MP tape samples show a granular structure with elliptically shaped magnetic particle aggregates (50–100 nm in diameter). Studies by *Scott* and *Bhushan* [27.69], *Bhushan* and *Qi* [27.70], and *Kasai* et al. [27.71] have indicated that the phase shift can be related to the energy dissipation through the viscoelastic deformation process between the tip and the sample. Recent theoretical analysis has established a quantitative correlation between the lateral surface properties (stiffness and viscoelasticity) of materials and amplitude/phase angle shift in TR measurements [27.92]. The contrast in the TR amplitude and phase angle images is due to the in-plane (lateral) heterogeneity of the surface. The TR amplitude and phase angle images enable the lateral surface properties (lateral stiffness and viscoelasticity) of materials to be mapped.

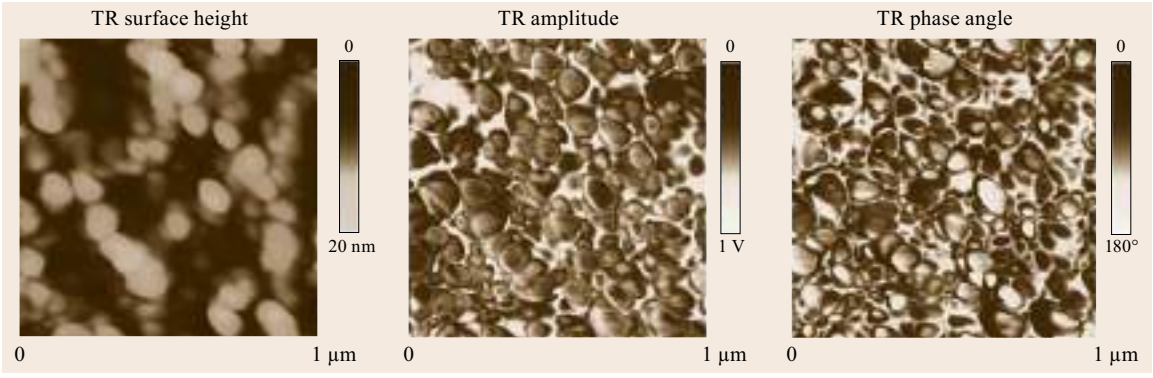


Fig. 27.62 Images of an MP tape obtained with TR mode II (constant deflection). TR mode II amplitude and phase angle images have the largest contrast among tapping, TR mode I, and TR mode II techniques (after [27.72])

27.5 Boundary Lubrication

27.5.1 Perfluoropolyether Lubricants

The classical approach to lubrication uses freely supported multimolecular layers of liquid lubricants [27.2, 9, 16, 149]. The liquid lubricants are sometimes chemically bonded to improve their wear resistance [27.2, 14, 16, 103, 104]. Partially chemically bonded, molecularly thick perfluoropolyether (PFPE) films are used for the lubrication of magnetic storage media because of their thermal stability and extremely low vapor pressure [27.16]. Chemically bonded lubricants are considered potential candidate lubricants for MEMS/NEMS. Molecularly thick PFPEs are well suited because of the following properties: low surface tension and a low contact angle which allows easy spreading on surfaces and provide hydrophobic properties; chemical and thermal stability which minimizes degradation under use; low vapor pressure which provides low out-gassing; high adhesion to substrate via organic functional bonds; and good lubricity which reduces contact surface wear.

For boundary lubrication studies, friction, adhesion, and durability experiments have been performed on virgin Si(100) surfaces and silicon surfaces lubricated with various PFPE lubricants [27.41–43, 47, 51, 53, 54]. More recently, there has been interest in selected ionic liquids for lubrication [27.150–152]. They possess efficient heat transfer properties. They are also electrically conducting, which is of interest in various MEMS/NEMS applications. Results of the following two PFPE lubricants will be presented here: Z-15 (with $-CF_3$ nonpolar end groups),

$$CF_3{-}O{-}(CF_2{-}CF_2{-}O)_m{-}(CF_2{-}O)_n{-}CF_3$$

($m/n \approx 2/3$) and Z-DOL (with $-OH$ polar end groups),

$$HO{-}CH_2{-}CF_2{-}O{-}(CF_2{-}CF_2{-}O)_m{-}(CF_2{-}O)_n{-}CF_2{-}CH_2{-}OH$$

($m/n \approx 2/3$). Z-DOL film was thermally bonded at 150 °C for 30 min, and the unbonded fraction was removed by a solvent (fully bonded) [27.16]. The thicknesses of Z-15 and Z-DOL films were 2.8 nm and 2.3 nm, respectively. The lubricant chain diameter of these molecules is about 0.6 nm, and molecularly thick films generally lie flat on surfaces with high coverage.

The adhesive forces of Si(100), Z-15, and Z-DOL (fully bonded) measured by a force calibration plot and a friction force versus normal load plot are summarized in Fig. 27.63 [27.47]. The data obtained by these two methods are in good agreement. Figure 27.63 shows that the presence of mobile Z-15 lubricant film increases the adhesive force as compared to that of Si(100) by meniscus formation. Whereas, the presence of the solid-like phase of the Z-DOL (fully bonded) film reduces the adhesive force as compared that of Si(100) because of the absence of mobile liquid. The schematic (bottom) in Fig. 27.63 shows the relative size and sources of the meniscus. It is well known that the native oxide layer (SiO_2) on top of the Si(100) wafer exhibits hydrophilic properties, and some water molecules can be adsorbed on this surface. The condensed water will form a meniscus as the tip approaches the sample surface. The larger adhesive force in Z-15 is not only caused by the Z-15 meniscus alone, the nonpolarized Z-15 liquid does not have good wettability and strong bonding with Si(100). Consequently, in the ambient environment, the condensed water molecules from the environment will permeate through the liquid Z-15 lubricant film and compete with the lubricant molecules present on the substrate. The interaction of the liquid lu-

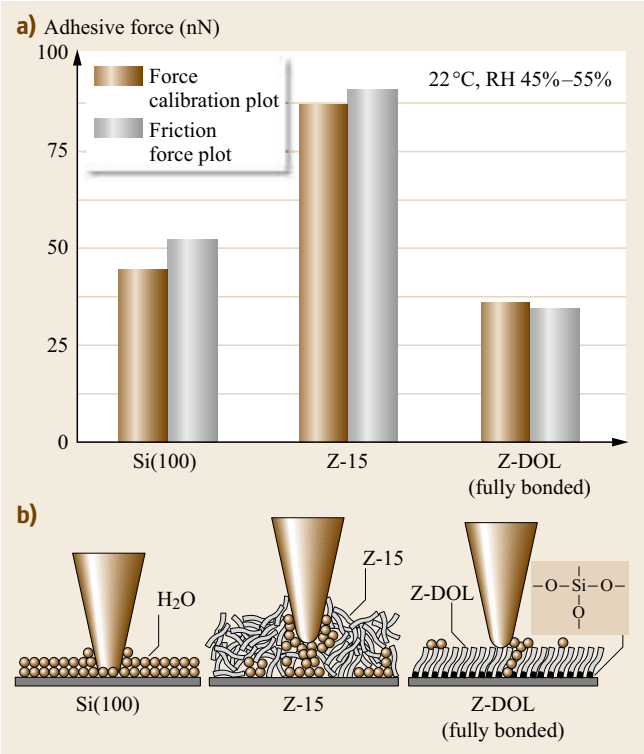


Fig. 27.63a,b Summary of the adhesive forces of Si(100) and Z-15 and Z-DOL (fully bonded) films measured by force calibration plots and friction force versus normal load plots in ambient air **(a)**. **(b)** Schematic showing the effect of the meniscus, formed between the AFM tip and the surface sample, on the adhesive and friction forces (after [27.47])

bricant with the substrate is weakened, and a boundary layer of the liquid lubricant forms puddles [27.42, 43]. This dewetting allows water molecules to be adsorbed on the Si(100) surface along with Z-15 molecules, and both of them can form a meniscus while the tip approaches the surface. Thus, the dewetting of the liquid Z-15 film results in a higher adhesive force and poorer lubrication performance. In addition, the Z-15 film is soft compared to the solid Si(100) surface, and penetration of the tip in the film occurs while pushing the tip down. This results in a large area of the tip being wetted by the liquid to form the meniscus at the tip–liquid (mixture of Z-15 and water) interface. It should also be noted that Z-15 has a higher viscosity compared to water, therefore the Z-15 film provides a higher resistance to motion and the coefficient of friction. In the case of the Z-DOL (fully bonded) film, both of the active groups of Z-DOL molecules are mostly bonded on the Si(100) substrate, thus the Z-DOL (fully bonded) film has low free surface energy and cannot be displaced readily by the water molecules or can readily adsorb the water molecules. Thus, the use of Z-DOL (fully bonded) can reduce the adhesive force.

To study the velocity effect on friction and adhesion, the variation of friction force, the adhesive force, and the coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of velocity are summarized in Fig. 27.64 [27.47]. It indicates that for a silicon wafer, the friction force decreases logarithmically with increasing velocity. For Z-15, the friction force decreases with increasing velocity up to 10 μm/s, after which it remains almost constant. The velocity has a very small effect on the friction force of Z-DOL (fully bonded); it reduced slightly only at very high velocity. Figure 27.64 also indicates that the adhesive force of Si(100) is increased when the velocity is higher than 10 μm/s. The adhesive force of Z-15 is reduced dramatically with a velocity increase up to 20 μm/s, after which it is reduced slightly, and the adhesive force of Z-DOL (fully bonded) is also decreased at high velocity. In the testing range of velocity, only the coefficient of friction of Si(100) decreases with velocity, but the coefficients of friction of Z-15 and Z-DOL (fully bonded) almost remain constant. This implies that the friction mechanisms of Z-15 and Z-DOL (fully bonded) do not change with the variation of velocity.

The mechanisms of the effect of velocity on adhesion and friction are explained based on schematics shown in Fig. 27.65 (right) [27.47]. For Si(100), a tribochemical reaction plays a major role. Although, at high velocity, the meniscus is broken and does not have enough time to rebuild, the contact stresses and high velocity lead to tribochemical reactions of the Si(100) wafer (which has native oxide (SiO_2)) and Si_3N_4 tip with water molecules and they form $Si(OH)_4$. The $Si(OH)_4$ is removed and continuously replenished during sliding. The $Si(OH)_4$ layer between the tip and the Si(100) surface is known to be of low shear strength and causes a decrease in friction force and coefficient of friction [27.103, 104]. The chemical bonds of Si–

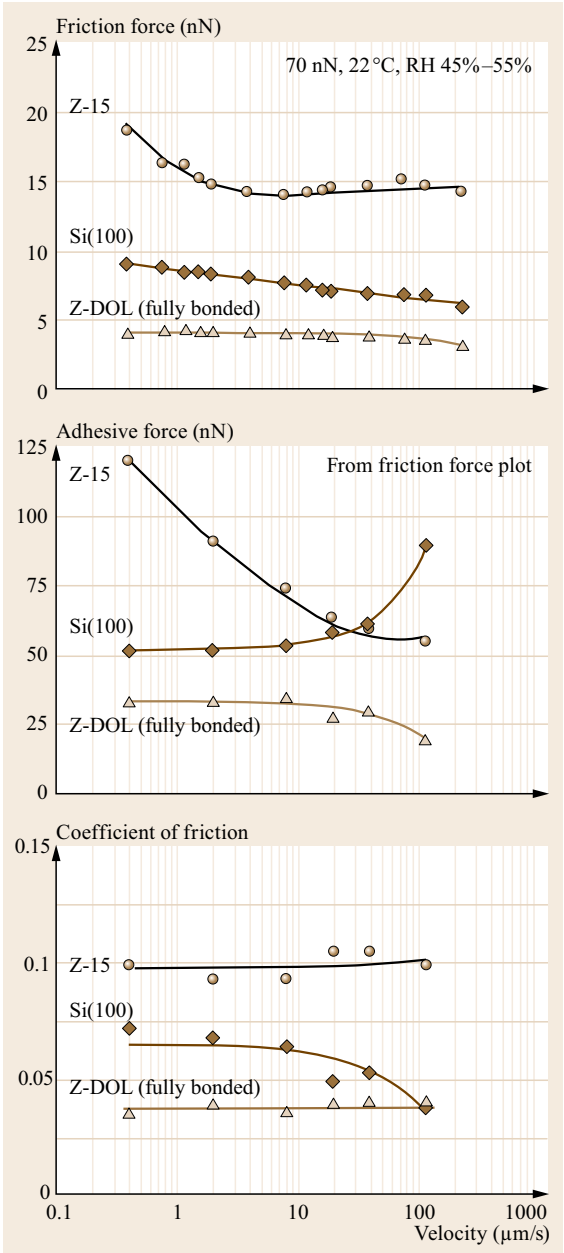


Fig. 27.64 The influence of velocity on the friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 and Z-DOL (fully bonded) films at 70 nN, in ambient air (after [27.47])

OH between the tip and Si(100) surface induce a large adhesive force. For the Z-15 film, at high velocity the meniscus formed by the condensed water and the Z-15 molecules is broken and does not have enough time to rebuild, therefore, the adhesive force and consequently friction force is reduced. The friction mechanisms for the Z-15 film still shear the same viscous liquid even at high velocity range, thus the coefficient of friction of the Z-15 film does not change with velocity. For the Z-DOL (fully bonded) film, the surface can adsorb a few water molecules in an ambient condition, and at high velocity these molecules are displaced, which is responsible for a slight decrease in friction force and adhesive force. *Koinkar* and *Bhushan* [27.42, 43] have suggested that in the case of samples with mobile films, such as condensed water and Z-15 films, alignment of the liquid molecules (shear thinning) is responsible for the drop in friction force with an increase in scanning velocity. This could be another reason for the decrease in friction force for the Si(100) and Z-15 films with velocity in this study.

To study the relative humidity effect on friction and adhesion, the variation of friction force, adhesive force, and the coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of relative humidity are shown in Fig. 27.66 [27.47]. It shows that for Si(100) and Z-15 film, the friction force increases with a relative humidity increase up to 45%, and then it shows a slight decrease with a further increase in the relative humidity. Z-DOL (fully bonded) has a smaller friction force than Si(100) and Z-15 in the whole testing

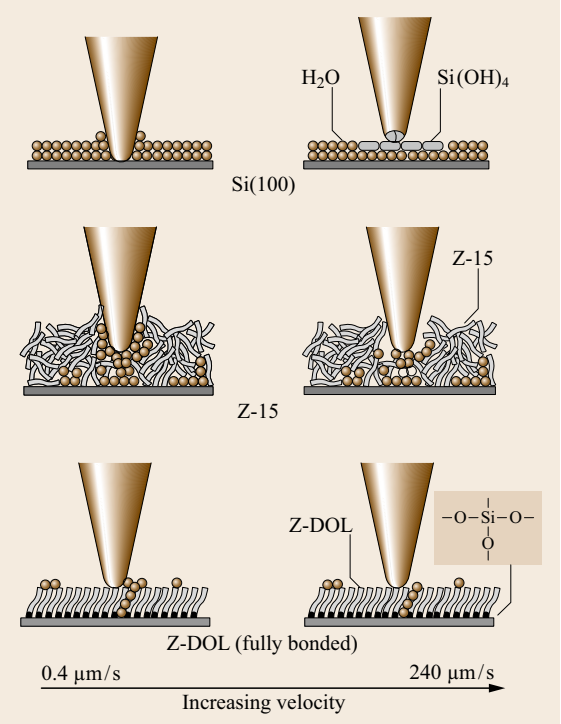


Fig. 27.65 The schematic shows the change of surface composition (by tribochemical reaction) and formation of a meniscus while increasing the velocity (after [27.47])

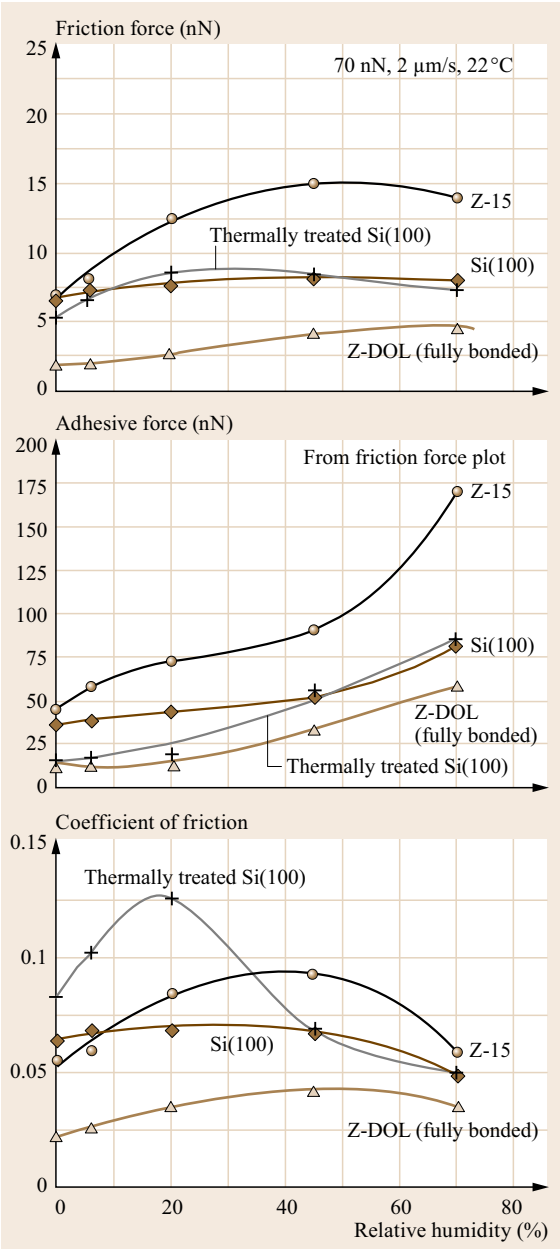


Fig. 27.66 The influence of relative humidity on the friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 and Z-DOL (fully bonded) films at 70 nN, 2 μm/s, and in 22 °C air (after [27.47])

range, and its friction force shows a relative apparent increase when the relative humidity is higher than 45%. For Si(100), Z-15, and Z-DOL (fully bonded), their adhesive forces increase with relative humidity, and their coefficients of friction increase with relative humidity up to 45%, after which they decrease with a further increase in relative humidity. It is also observed that the humidity effect on Si(100) really depends on the history of the Si(100) sample. As the surface of the Si(100) wafer readily adsorbs water in air, without any pretreatment the Si(100) used in our study almost reaches a saturated stage in terms of adsorbed water, and is responsible for less effect during increasing relative humidity. However, once the Si(100) wafer was thermally treated by baking at 150 °C for 1 h, a greater effect was observed.

The schematic (right) in Fig. 27.67 shows that because of its high free surface energy Si(100) can adsorb more water molecules during an increase in relative humidity [27.47]. As discussed earlier, for the Z-15 film in the humid environment, the condensed water from the humid environment competes with the lubricant film present on the sample surface, and the interaction of the liquid lubricant film with the silicon substrate is weakened and a boundary layer of the liquid lubricant forms puddles. This dewetting allows water molecules to be adsorbed on the Si(100) substrate mixed with Z-15 molecules [27.42, 43]. Obviously, more water molecules can be adsorbed on the

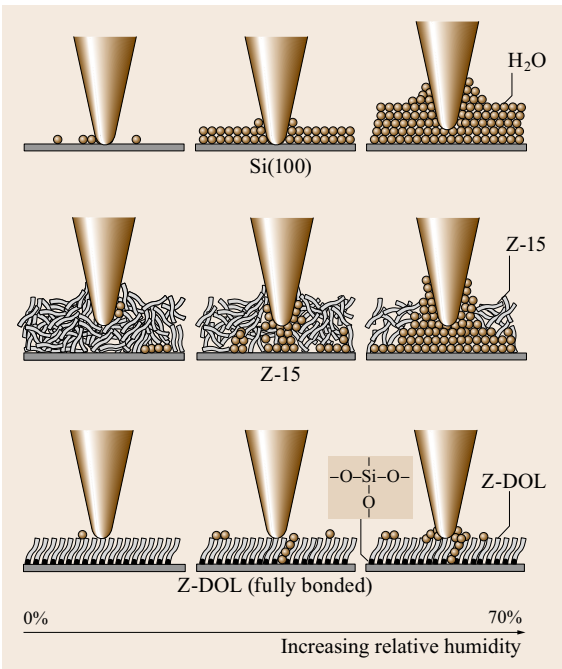


Fig. 27.67 Schematic shows the change of meniscus while increasing relative humidity. In this figure, the thermally treated Si(100) represents the Si(100) wafer that was baked at 150 °C for 1 h in an oven (in order to remove the adsorbed water) just before it was placed in the 0% RH chamber (after [27.47])

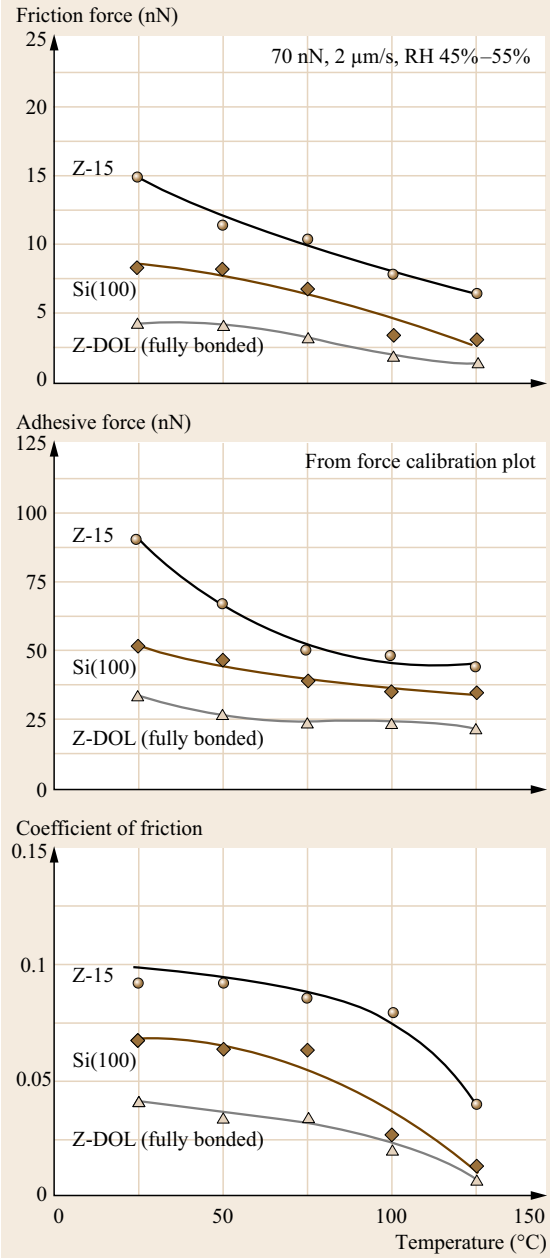


Fig. 27.68 The influence of temperature on the friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 and Z-DOL (fully bonded) films at 70 nN, 2 μm/s, and in RH 40–50% air (after [27.47])

Z-15 surface while increasing relative humidity. The more adsorbed water molecules in the case of Si(100), along with lubricant molecules in the Z-15 film case, form a bigger water meniscus which leads to an increase of friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 with humidity, but at a very high humidity of 70%, large quantities of adsorbed water can form a continuous water layer that separates the tip and sample surface and acts as a kind of lubricant, which causes a decrease in the friction force and the coefficient of friction. For the Z-DOL (fully bonded) film, because of their hydrophobic surface properties, water molecules can be adsorbed at humidity higher than 45%, and cause an increase in the adhesive force and friction force.

To study the temperature effect on friction and adhesion, the variation of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of temperature are summarized in Fig. 27.68 [27.47]. It shows that the increasing temperature causes a decrease of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded). The schematic (right) in Fig. 27.69 indicates that at high temperature, desorption of water leads to a decrease of the friction force, the adhesive forces, and the coefficient of friction for all samples. For the Z-15 film, the reduction of viscosity at high temperature also has a contribution to the decrease of friction force and coefficient of friction. In the case of the Z-DOL (fully bonded) film, molecules can be more easily oriented at high temperature, which

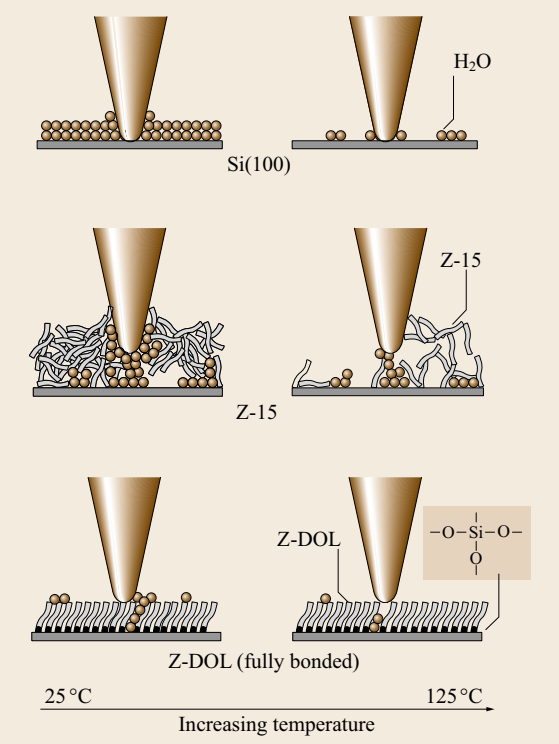


Fig. 27.69 The schematic shows that at high temperature, desorption of water decreases the adhesive forces. And the reduced viscosity of Z-15 leads to the decrease of coefficient of friction. High temperature facilitates orientation of molecules in the Z-DOL (fully bonded) film which results in a lower coefficient of friction (after [27.47])

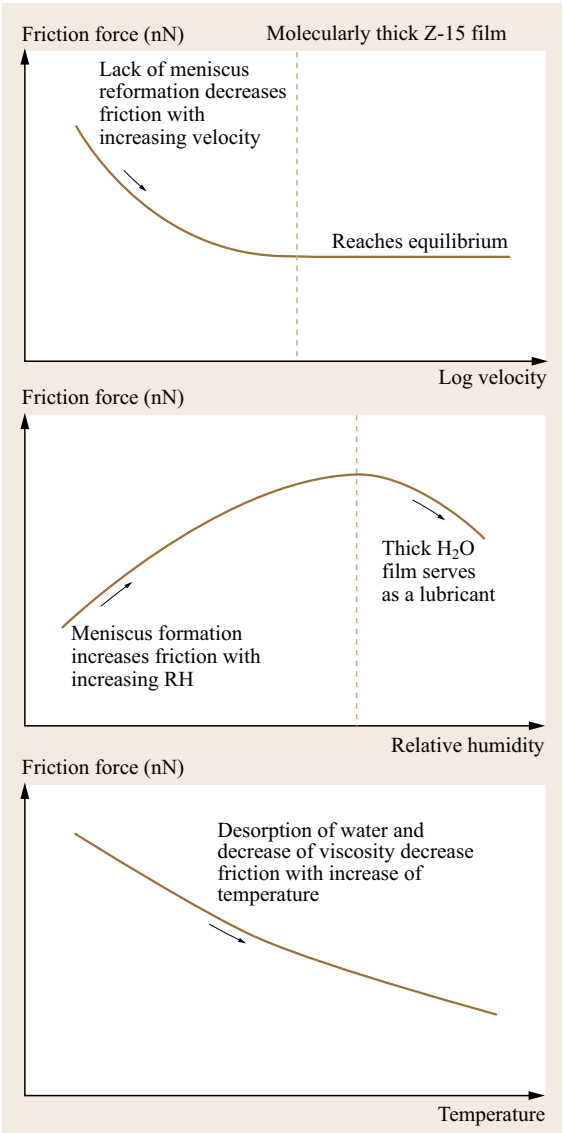


Fig. 27.70 Schematic shows the change of friction force of molecularly thick Z-15 films with log velocity, relative humidity, and temperature. The changing trends are also addressed in this figure (after [27.47])

may be partly responsible for the low friction force and coefficient of friction.

As a brief summary, the influence of velocity, relative humidity, and temperature on the friction force of a mobile Z-15 film is presented in Fig. 27.70 [27.47]. The changing trends are also addressed in this figure.

To study the durability of lubricant films at the nanoscale, the friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of the number of scanning cycles are shown in Fig. 27.71 [27.47]. As observed earlier, the friction force of Z-15 is higher than that of Si(100) with the lowest values for Z-DOL (fully bonded). During cycling, the friction force and the coefficient of friction of Si(100) show a slight decrease during the initial few cycles then remain constant. This is related to the removal of the native oxide. In the case of the Z-15 film, the friction force and coefficient of friction show an increase during the initial few cycles and then approach higher stable values. This is believed to be caused by the attachment of the Z-15 molecules to the tip. After several scans, the molecular interaction reaches equilibrium, and after that the friction force and coefficient of friction remain constant. In the case of the Z-DOL (fully bonded) film, the friction force and coefficient of friction start out low and remain low during the entire test for 100 cycles. This suggests that Z-DOL (fully bonded) molecules do not become attached or displaced as readily as Z-15.

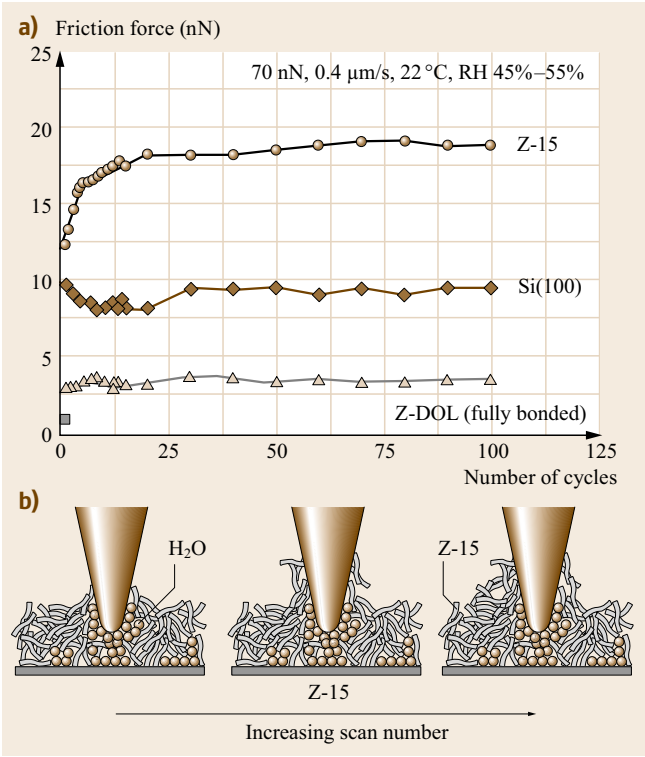


Fig. 27.71 **(a)** Friction force versus number of sliding cycles for Si(100) and Z-15 and Z-DOL (fully bonded) films at 70 nN, 0.8 μm/s, and in ambient air. **(b)** Schematic shows that some liquid Z-15 molecules can be attached to the tip. The molecular interaction between the molecules attached to the tip with the Z-15 molecules in the film results in an increase of friction force with multiscanning (after [27.47])

27.5.2 Self-Assembled Monolayers

For the lubrication of MEMS/NEMS, another effective approach involves the deposition of organized and dense molecular layers of long-chain molecules. Two common methods to produce monolayers and thin films are Langmuir–Blodgett (L–B) deposition and self-assembled monolayers (SAMs) by chemical grafting of molecules. LB films are physically bonded to the substrate by weak van der Waals attraction, while SAMs are chemically bonded via covalent bonds to the substrate. Because of the choice of chain length and terminal linking group that SAMs offer, they hold great promise for boundary lubrication of MEMS/NEMS. A number of studies have been conducted to study the tribological properties of various SAMs deposited on Si, Al, and Cu substrates [27.20, 39–41, 44–46, 48–50, 111, 153–159].

Bhushan and *Liu* [27.44] studied the effect of film compliance on adhesion and friction. They used hexadecane thiol (HDT), 1,1′-biphenyl-4-thiol (BPT), and crosslinked BPT (BPTC) solvent deposited on an Au(111) substrate (Fig. 27.72a). The average values and standard deviation of the adhesive force and coefficient of friction are presented in Fig. 27.72b. On the basis of this data, the adhesive force and coefficient of friction of SAMs are less than the corresponding substrates. Among various films, HDT exhibits the lowest values. Based on the stiffness measurements of various SAMs, HDT was most compliant, followed by BPT and BPTC. On the basis of friction and stiffness measurements, SAMs with high-compliance long carbon chains exhibit low friction; chain compliance is desirable for low friction. The friction mechanism of SAMs is explained by a so-called *molecular spring* model (Fig. 27.73). According to this model, the chemically adsorbed self-assembled molecules on a substrate are like assembled molecular springs anchored to the substrate. An asperity sliding on the surface of SAMs is like a tip sliding on the top of the *molecular springs or brush*. The molecular spring assembly has compliant features and can experience orientation and compression under load. The orientation of the molecular springs or brush under normal load reduces the shearing force at

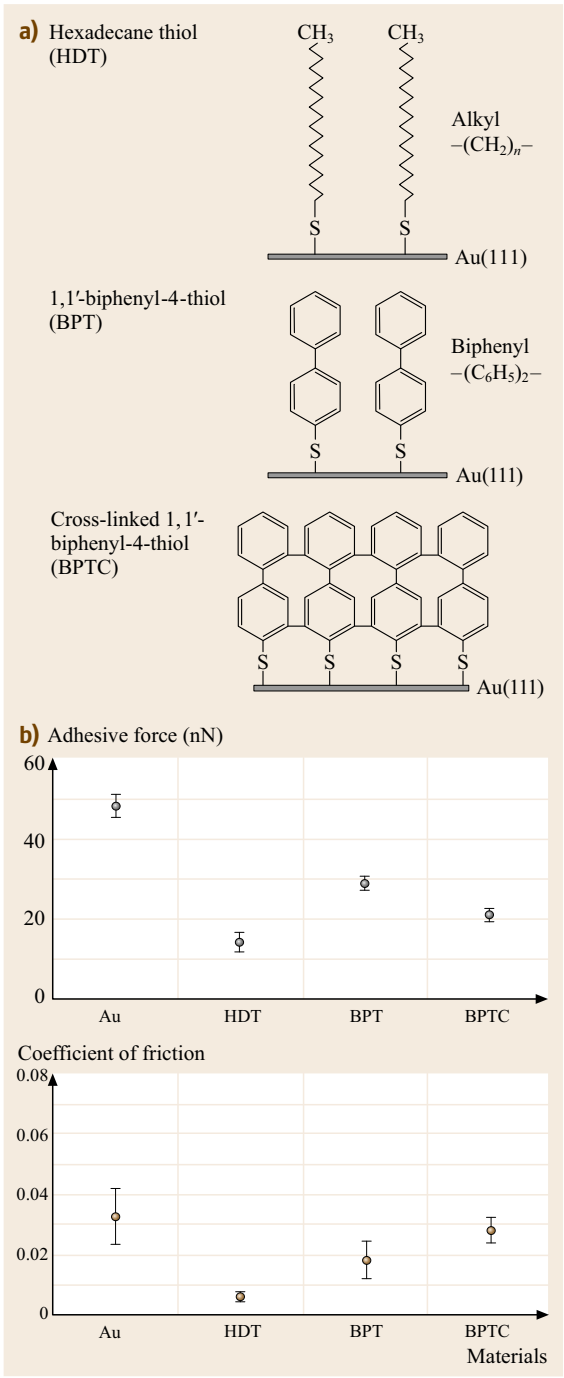


Fig. 27.72 **(a)** Schematics of structures of hexadecane thiol and biphenyl thiol SAMs on Au(111) substrates, and **(b)** adhesive force and coefficient of friction of Au(111) substrate and various SAMs (after [27.44]) ◄

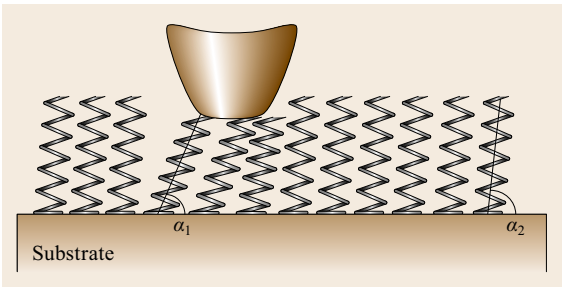


Fig. 27.73 Molecular spring model of SAMs. In this figure, $\alpha_1 < \alpha_2$, which is caused by further orientation under the normal load applied by an asperity tip (after [27.44])

the interface, which in turn reduces the friction force. The orientation is determined by the spring constant of a single molecule as well as by the interaction between the neighboring molecules, which can be reflected by the packing density or the packing energy. It should be noted that the orientation can lead to conformational defects along the molecular chains, which lead to energy dissipation.

An elegant way to demonstrate the influence of molecular stiffness on friction is to investigate SAMs with different structures on the same wafer. For this purpose, a micropatterned SAM was prepared. First, the biphenyldimethylchlorosilane (BDCS) was deposited on silicon by a typical self-assembly method [27.46]. Then the film was partially crosslinked using a mask technique with low-energy electron irradiation. Finally, the micropatterned BDCS films were realized, which had the as-deposited and crosslinked coating regions on the same wafer. The local stiffness properties of this micropatterned sample were investigated by the force-modulation AFM technique [27.66]. The variation in the deflection amplitude provides a measure of the relative local stiffness of the surface. Surface height, stiffness, and friction images of the micropatterned BDCS specimen were obtained and are presented in Fig. 27.74 [27.46]. The circular areas correspond to the as-deposited film, and the remaining area to the crosslinked film. Figure 27.74a indicates that crosslinking caused by the low-energy electron irradiation leads to about 0.5 nm decrease of the surface height of BDCS films. The corresponding stiffness images indicate that the crosslinked area has a higher stiffness than the as-deposited area. Figure 27.74b indicates that the as-deposited area (higher surface height area) has a lower friction force. Obviously, these data of the micropatterned sample prove that the local stiffness of SAMs has an influence on their friction performance. Higher stiffness leads to larger friction force. These results provide strong proof of the suggested molecular spring model.

The SAMs with high-compliance long carbon chains also exhibit the best wear resistance [27.44, 46].

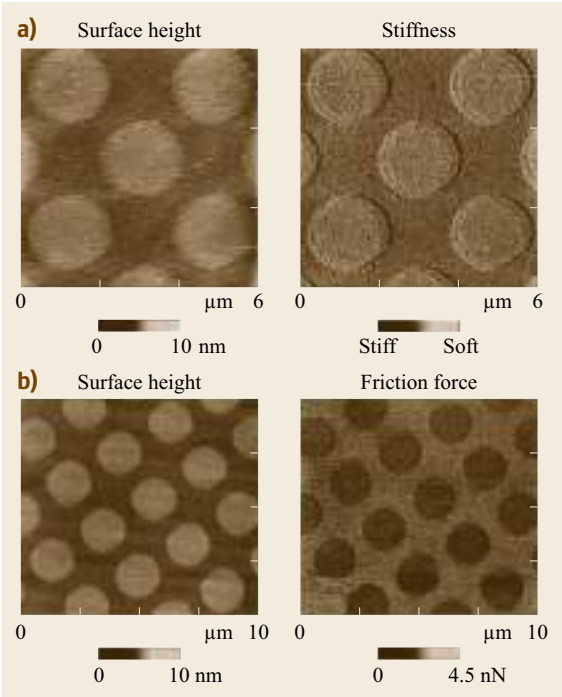


Fig. 27.74 **(a)** AFM gray-scale surface height and stiffness images, and **(b)** AFM gray-scale surface height and friction force images of micropatterned BDCS (after [27.46])

In wear experiments, the wear depth as a function of normal load curves shows a critical normal load, at which the film wears rapidly. A representative curve is shown in Fig. 27.75. Below the critical normal load, SAMs undergo orientation; at the critical load SAMs wear away from the substrate due to the relatively weak interface bond strengths, while above the critical normal load severe wear takes place on the substrate.

27.5.3 Liquid Film Thickness Measurements

Liquid film thickness mapping of ultrathin films (on the order of 2 nm) can be obtained using friction force microscopy [27.42] and adhesive force mapping [27.125]. Figure 27.76 shows the gray-scale plots of the surface topography and friction force obtained simultaneously for unbonded Demnum S-100-type PFPE lubricant film on silicon. Demnum-type PFPE lubricant (Demnum, Daikin, Japan) chains have $-CF_2-CH_2-OH$, a reactive end group on one end, whereas Z-DOL chains have the hydroxyl groups on both ends, as described earlier. The friction force plot shows well-distinguished low and high friction regions roughly corresponding to high and low regions in surface topography (thick and thin lubricant regions). A uniformly lubricated sample does not show such a variation in the friction.

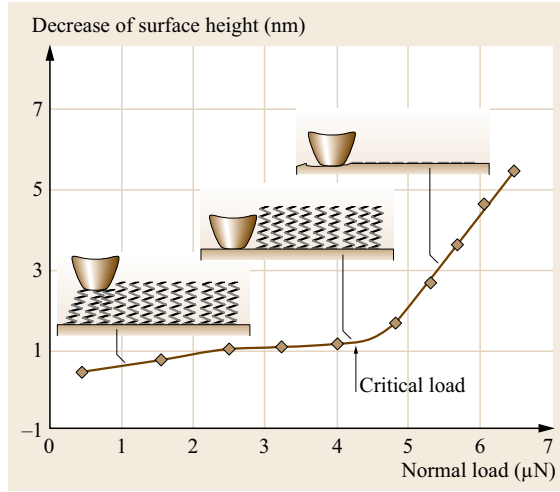


Fig. 27.75 Illustration of the wear mechanism of SAMs with increasing normal load (after [27.46])

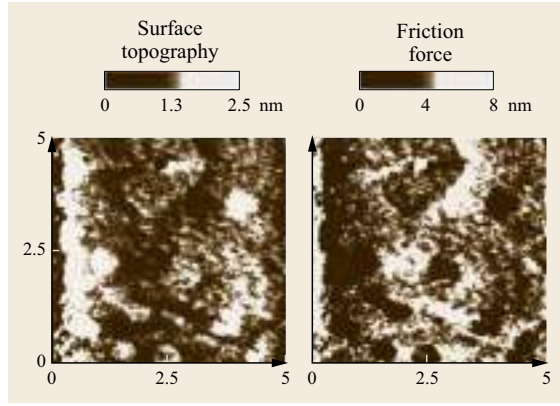


Fig. 27.76 Gray-scale plots of the surface topography and friction force obtained simultaneously for an unbonded Demnum-type perfluoropolyether lubricant film on silicon (after [27.42])

Friction force imaging can thus be used to measure the lubricant uniformity on the sample surface, which cannot be identified by surface topography alone. Figure 27.77 shows the gray-scale plots of the adhesive force distribution for silicon samples coated uniformly and nonuniformly with Z-DOL-type PFPE lubricant. It can be clearly seen that a region exists that has an adhesive force distinctly different from the other region for the nonuniformly coated sample. This implies that the liquid film thickness is nonuniform, giving rise to a difference in the meniscus forces.

Quantitative measurements of liquid film thickness of thin lubricant films (on the order of a few nm) with nanometer lateral resolution can be made with the AFM [27.2, 14, 72, 105]. The liquid film thickness is

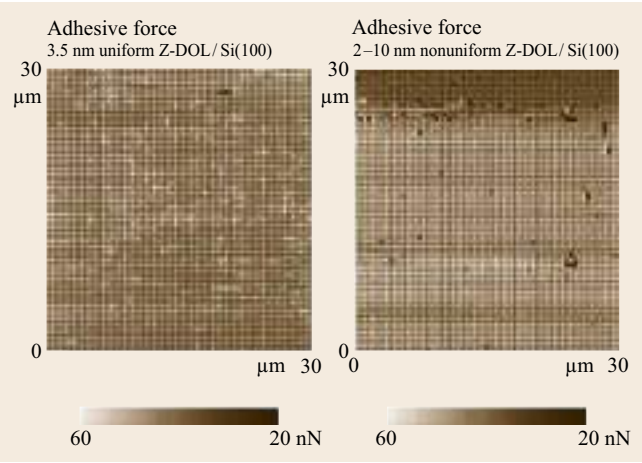


Fig. 27.77 Gray-scale plots of the adhesive force distribution of a uniformly coated, 3.5 nm thick unbonded Z-DOL film on silicon and 3–10 nm thick unbonded Z-DOL film on silicon that was deliberately coated nonuniformly by vibrating the sample during the coating process (after [27.125])

obtained by measuring the force on the tip as it approaches, contacts, and pushes through the liquid film and ultimately contacts the substrate. The distance between the sharp snap-in (owing to the formation of a liquid meniscus and van der Waals forces between the film and the tip) at the liquid surface and the hard repulsion at the substrate surface is a measure of the liquid film thickness. Figure 27.78 shows a plot of forces between the tip and both virgin hair and conditioner-treated hair. The hair sample was first brought into contact with the tip and then pulled away at a velocity of 400 nm/s. The zero tip–sample separation is defined to be the position where the force on the tip is zero, and the tip is not in contact with the sample. As the tip approaches the sample, a negative force exists which indicates an attractive force. The treated hair surface shows a much longer range of interaction with the tip compared to the very short range of interaction between virgin hair surfaces and the tip. Typically, the tip suddenly snaps into contact with the conditioner layer at a finite separation H (about 30 nm), which is proportional to conditioner thickness h. As the tip contacts the substrate, the tip travels with the sample. When the sample is withdrawn, the forces on the tip slowly decrease to zero once the meniscus of liquid is drawn out from the hair surface. It should be noted that the distance H between the sharp snap-in at the liquid surface and the hard wall contact with the substrate is not the real conditioner thickness h. Because of interaction of the liquid with the tip at some spacing distance, H tends to be thicker than the actual film thickness, but can still provide an estimate of the actual film thickness and upper limit of thickness.

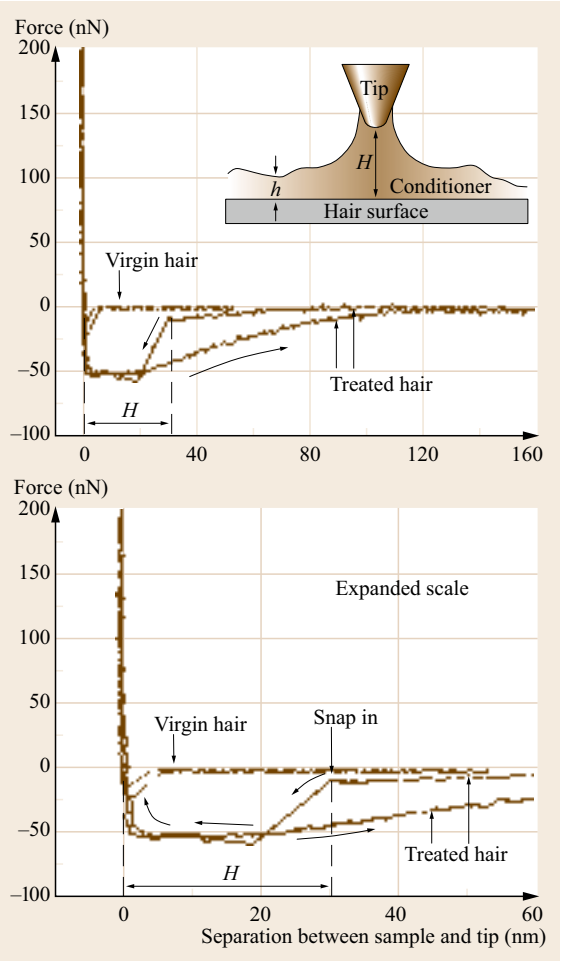


Fig. 27.78 Forces between tip and hair surface as a function of tip–sample separation for virgin hair and conditioner-treated hair. A schematic of measurement for localized conditioner thickness is shown in the *inset at the top*. The expanded scale view of force curve at small separation is shown *at the bottom* (after [27.72]) ◄

27.6 Conclusion

For most solid–solid interfaces of technological relevance, contact occurs at multiple asperities. A sharp AFM/FFM tip sliding on a surface simulates just one such contact. However, asperities come in all shapes and sizes. The effect of the radius of a single asperity (tip) on the friction/adhesion performance can be studied using tips of different radii. AFM/FFM are used to study various tribological phenomena including surface roughness, adhesion, friction, scratching, wear, indentation, detection of material transfer, and boundary lubrication. Measurement of the atomic-scale friction of a freshly cleaved, highly oriented, pyrolytic graphite exhibits the same periodicity as that of the corresponding topography. However, the peaks in friction and those in the corresponding topography are displaced relative to each other. Variations in atomic-scale friction and the observed displacement can be explained by the variation in interatomic forces in the normal and lateral directions. The relevant friction mechanism is atomic-scale stick-slip. Local variations in microscale friction occur and are found to correspond to the local slopes, suggesting that a ratchet mechanism and collision effects are responsible for this variation. Directionality in the friction is observed on both micro- and macroscales which results from the surface roughness and surface preparation. Anisotropy in surface roughness accentuates this effect. Friction contrast in conventional frictional measurements is based on interactions dependent upon interfacial material properties superimposed by roughness-induced lateral forces. To obtain roughness-independent friction, lateral or torsional modulation techniques can be used. These techniques also allow measurements over a small region. AFM/FFM experiments are generally conducted at relative velocities up to about 200 μm/s. High-velocity experiments can be performed by either mounting a sample on a shear wave transducer driven at very high frequencies or mounting a sample on a high-velocity piezo stage. By using these techniques, friction and wear experiments can be performed at a range of sliding velocities as well as normal loads, and the data have been used to develop nanoscale friction and wear maps. Relevant friction mechanisms are different for different ranges of sliding velocities and normal loads.

Adhesion and friction in a wet environment depend on the tip radius, surface roughness, and relative humidity. Superhydrophobic surfaces can be designed by roughness optimization.

Nanoscale friction is generally found to be smaller than microscale friction. There are several factors responsible for the differences including wear and contaminant particles, the transition from elasticity to plasticity, scale-dependent roughness and mechanical properties, and meniscus effects. Nanoscale friction values increase with an increase in the normal load above a certain critical load (pressure), approaching the macroscale friction. The critical contact pressure corresponds to the hardness of the softer of the two contacting materials.

The wear rate on the microscale for single-crystal silicon is negligible below 20 μN, and is much higher and remains approximately constant at higher loads. Elastic deformation at low loads is responsible for negligible wear. Most of the wear debris is loose. SEM and TEM studies of the wear region suggest that the material on the microscale is removed by plastic deformation with a small contribution from elastic fracture; this observation corroborates with the scratch data. Evolution of wear has also been studied using AFM. Wear is found to be initiated at nanoscratches. For a sliding interface requiring near-zero friction and wear, contact stresses should be below the hardness of the softer material to minimize plastic deformation and surfaces should be free of nanoscratches. Further, wear precursors can be detected at early stages of wear by using surface potential measurements. It is found that even in the case of zero wear (no measurable deformation of the surface using AFM), there can be a significant change in the surface potential inside the wear mark, which is useful for the study of wear precursors. Detection of material transfer on a nanoscale is possible with AFM.

In situ surface characterization of the local deformation of materials and thin coatings can be carried out using a tensile stage inside an AFM. An AFM can also be used for nanofabrication/nanomachining.

A modified AFM can be used to obtain load–displacement curves and for the measurement of nanoindentation hardness and Young's modulus of elasticity, with a depth of indentation as low as 1 nm. A continuous stiffness technique can be used to measure variations in mechanical properties of layered materials and nonhomogeneous composites. Mechanical properties are scale-dependent. For example, hardness of ceramics on nanoscales is found to be higher than that on the microscale. Ceramics exhibit significant plasticity and creep on a nanoscale. By using the force modulation technique, localized surface elasticity maps of composite materials with penetration depth as low as 1 nm can be obtained. By using phase contrast microscopy in tapping or torsional mode, it is possible to obtain phase contrast maps or the contrast in viscoelastic properties of near-surface regions. Scratching and indentation on the nanoscale are powerful ways to screen for adhesion and resistance to deformation of ultrathin films.

Boundary lubrication studies and measurement of lubricant-film thickness with a lateral resolution on the nanoscale can be conducted using AFM. Chemically bonded lubricant films and self-assembled monolayers are superior in friction and wear resistance. For chemically bonded lubricant films, the adsorption of water, formation of meniscus and its changes during sliding, and surface properties play an important role in regard to the adhesion, friction, and durability of these films. Sliding velocity, relative humidity, and temperature affect adhesion and friction. For SAMs, their friction mechanism is explained by a so-called *molecular spring* model. Films with high-compliance, long carbon chains exhibit low friction and wear. Also, perfluoroalkylsilane SAMs on Si appear to be more hydrophobic with lower adhesion than alkylsilane SAMs on Si.

Investigations of adhesion, friction, wear, scratching, and indentation on the nanoscale using an AFM can provide insight into the failure mechanisms of materials. Coefficients of friction, wear rates, and mechanical properties such as hardness have been found to be different on the nanoscale than on the macroscale; generally, the coefficients of friction and wear rates on micro- and nanoscales are smaller, whereas the hardness is greater. Therefore, nanotribological studies may help define the regimes for ultralow friction and near-zero wear. These studies also provide insight into the atomic origins of adhesion, friction, wear, and lubrication mechanisms.

References

27.1 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374**, 607–616 (1995)

27.2 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

27.3 B. Bhushan: *Micro/Nanotribology and its Applications*, Nato Science, Vol. 330 (Kluwer, Dordrecht 1997)

27.4 I.L. Singer, H.M. Pollock: *Fundamentals of Friction: Macroscopic and Microscopic Processes*, Nato Science Series, Vol. E220 (Kluwer, Dordrecht 1992)

27.5 H.J. Guntherodt, D. Anselmetti, E. Meyer: *Forces in Scanning Probe Methods*, Vol. E286 (Kluwer, Dordrecht 1995)

27.6 B.N.J. Persson, E. Tosatti: *Physics of Sliding Friction*, Nato Science, Vol. E311 (Kluwer, Dordrecht 1996)

27.7 B. Bhushan: Nanoscale tribophysics and tribomechanics, Wear **225–229**, 465–492 (1999)

27.8 B. Bhushan: Wear and mechanical characterisation on micro- to picoscales using AFM, Int. Mater. Rev. **44**, 105–117 (1999)

27.9 B. Bhushan: *Modern Tribology Handbook: Principles of Tribology*, Vol. 1 (CRC, Boca Raton 2001)

27.10 B. Bhushan: *Fundamentals of Tribology and Bridging the Gap Between the Macro- and Micro/Nanoscales*, NATO Science Series II, Vol. 10 (Kluwer, Dordrecht 2001)

27.11 B. Bhushan: Nano- to microscale wear and mechanical characterization studies using scanning probe microscopy, Wear **251**, 1105–1123 (2001)

27.12 B. Bhushan: Nanotribology and nanomechanics, Wear **259**, 1507–1531 (2005)

27.13 B. Bhushan: Nanotribology, nanomechanics and nanomaterials characterization, Philos. Trans. R. Soc. A **366**, 1351–1381 (2008)

27.14 B. Bhushan: *Nanotribology and Nanomechanics – An Introduction*, 4th edn. (Springer International, Cham 2017)

27.15 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer, Cham 2016)

27.16 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)

27.17 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

27.18 B. Bhushan: Chemical, mechanical and tribological characterization of ultra-thin and hard amorphous carbon coatings as thin as 3.5 nm: Recent developments, Diam. Relat. Mater. **8**, 1985–2015 (1999)

27.19 B. Bhushan: Adhesion and stiction: Mechanisms, measurement techniques, and methods for reduction, J. Vac. Sci. Technol. B **21**, 2262–2296 (2003)

27.20 B. Bhushan, A.V. Kulkarni, V.N. Koinkar, M. Boehm, L. Odoni, C. Martelet, M. Belin: Microtribological characterization of self-assembled and Langmuir–Blodgett monolayers by atomic and friction force microscopy, Langmuir **11**, 3189–3198 (1995)

27.21 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscopy, Phys. Rev. Lett. **56**, 930–933 (1986)

27.22 G. Binnig, C. Gerber, E. Stoll, T.R. Albrecht, C.F. Quate: Atomic resolution with atomic force microscope, Europhys. Lett. **3**, 1281–1286 (1987)

27.23 C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59**, 1942–1945 (1987)

27.24 B. Bhushan, J. Ruan: Atomic-scale friction measurements using friction force microscopy: Part II – Application to magnetic media, ASME J. Tribol. **116**, 389–396 (1994)

27.25 J. Ruan, B. Bhushan: Atomic-scale friction measurements using friction force microscopy: Part I – general principles and new measurement techniques, ASME J. Tribol. **116**, 378–388 (1994)

27.26 J. Ruan, B. Bhushan: Atomic-scale and microscale friction of graphite and diamond using friction force microscopy, J. Appl. Phys. **76**, 5022–5035 (1994)

27.27 J. Ruan, B. Bhushan: Frictional behavior of highly oriented pyrolytic graphite, J. Appl. Phys. **76**, 8117–8120 (1994)

27.28 B. Bhushan, V.N. Koinkar, J. Ruan: Microtribology of magnetic media, Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. **208**, 17–29 (1994)

27.29 B. Bhushan, A.V. Kulkarni: Effect of normal load on microscale friction measurements, Thin Solid Films **278**, 49–56 (1996)

27.30 B. Bhushan, A.V. Kulkarni: Erratum to "Effect of normal load on microscale friction measurement", Thin Solid Films **293**, 333 (1997)

27.31 B. Bhushan, S. Sundararajan: Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46**, 3793–3804 (1998)

27.32 V. Scherer, W. Arnold, B. Bhushan: Active friction control using ultrasonic vibration. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 463–469

27.33 V. Scherer, W. Arnold, B. Bhushan: Lateral force microscopy using acoustic friction force microscopy, Surf. Interface Anal. **27**, 578–587 (1999)

27.34 M. Reinstaedtler, U. Rabe, V. Scherer, U. Hartmann, A. Goldade, B. Bhushan, W. Arnold: On the nanoscale measurement of friction using atomic-force microscope cantilever torsional resonances, Appl. Phys. Lett. **82**, 2604–2606 (2003)

27.35 M. Reinstaedtler, U. Rabe, A. Goldade, B. Bhushan, W. Arnold: Investigating ultrathin lubricant layers using resonant friction force microscopy, Tribol. Int. **38**, 533–541 (2005)

27.36 M. Reinstaedtler, T. Kasai, U. Rabe, B. Bhushan, W. Arnold: Imaging and measurement of elasticity and friction using the TR mode, J. Phys. D **38**, R269–R282 (2005)

27.37 B. Bhushan, T. Kasai: A surface topography-independent friction measurement technique using torsional resonance mode in an AFM, Nanotechnology **15**, 923–935 (2004)

27.38 N.S. Tambe, B. Bhushan: A new atomic force microscopy based technique for studying nanoscale friction at high sliding velocities, J. Phys. D: Appl. Phys. **38**, 764–773 (2005)

27.39 B. Bhushan, T. Kasai, G. Kulik, L. Barbieri, P. Hoffmann: AFM study of perfluorosilane and alkylsilane self-assembled monolayers for anti-stiction in MEMS/NEMS, Ultramicroscopy **105**, 176–188 (2005)

27.40 B. Bhushan, D. Hansford, K.K. Lee: Surface modification of silicon and polydimethylsiloxane surfaces with vapor-phase-deposited ultrathin fluorosilane films for biomedical nanodevices, J. Vac. Sci. Technol. A **24**, 1197–1202 (2006)

27.41 B. Bhushan, M. Cichomski, Z. Tao, N.T. Tran, T. Ethen, C. Merton, R.E. Jewett: Nanotribological characterization and lubricant degradation studies of metal-film magnetic tapes using novel lubricants, ASME J. Tribol. **129**, 621–627 (2007)

27.42 V.N. Koinkar, B. Bhushan: Micro/nanoscale studies of boundary layers of liquid lubricants for magnetic disks, J. Appl. Phys. **79**, 8071–8075 (1996)

27.43 V.N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14**, 2378–2391 (1996)

27.44 B. Bhushan, H. Liu: Nanotribological properties and mechanisms of alkylthiol and biphenyl thiol self-assembled monolayers studied by AFM, Phys. Rev. B **63**, 245412-1–245412-11 (2001)

27.45 H. Liu, B. Bhushan, W. Eck, V. Staedtler: Investigation of the adhesion, friction, and wear properties of biphenyl thiol self-assembled monolayers by atomic force microscopy, J. Vac. Sci. Technol. A **19**, 1234–1240 (2001)

27.46 H. Liu, B. Bhushan: Investigation of nanotribological properties of self-assembled monolayers with alkyl and biphenyl spacer chains, Ultramicroscopy **91**, 185–202 (2002)

27.47 H. Liu, B. Bhushan: Nanotribological characterization of molecularly-thick lubricant films for applications to MEMS/NEMS by AFM, Ultramicroscopy **97**, 321–340 (2003)

27.48 T. Kasai, B. Bhushan, G. Kulik, L. Barbieri, P. Hoffmann: Nanotribological study of perfluorosilane SAMs for anti-stiction and low wear, J. Vac. Sci. Technol. B **23**, 995–1003 (2005)

27.49 K.K. Lee, B. Bhushan, D. Hansford: Nanotribological characterization of perfluoropolymer thin films for biomedical micro/nanoelectromechanical systems applications, J. Vac. Sci. Technol. A **23**, 804–810 (2005)

27.50 N.S. Tambe, B. Bhushan: Nanotribological characterization of self assembled monolayers deposited on silicon and aluminum substrates, Nanotechnology **16**, 1549–1558 (2005)

27.51 Z. Tao, B. Bhushan: Bonding, degradation, and environmental effects on novel perfluoropolyether lubricants, Wear **259**, 1352–1361 (2005)

27.52 Z. Tao, B. Bhushan: Degradation mechanisms and environmental effects on perfluoropolyether, self assembled monolayers, and diamondlike carbon films, Langmuir **21**, 2391–2399 (2005)

27.53 M. Palacio, B. Bhushan: Surface potential and resistance measurements for detecting wear of chemically-bonded and unbonded molecularly-thick perfluoropolyether lubricant films using atomic force microscopy, J. Colloid Interface Sci. **315**, 261–269 (2007)

27.54 M. Palacio, B. Bhushan: Wear detection of candidate MEMS/NEMS lubricant films using atomic force microscopy-based surface potential measurements, Scr. Mater. **57**, 821–824 (2007)

27.55 B. Bhushan, V.N. Koinkar: Tribological studies of silicon for magnetic recording applications, J. Appl. Phys. **75**, 5741–5746 (1994)

27.56 V.N. Koinkar, B. Bhushan: Microtribological studies of Al_2O_3-TiC, polycrystalline and single-crystal Mn-Zn ferrite and SiC head slider materials, Wear

202, 110–122 (1996)
27.57 V.N. Koinkar, B. Bhushan: Microtribological properties of hard amorphous carbon protective coatings for thin film magnetic disks and heads, Proc. Inst. Mech. Eng. Part J **211**, 365–372 (1997)
27.58 S. Sundararajan, B. Bhushan: Development of a continuous microscratch technique in an atomic force microscope and its application to study scratch resistance of ultra-thin hard amorphous carbon coatings, J. Mater. Res. **16**, 75–84 (2001)
27.59 J. Ruan, B. Bhushan: Nanoindentation studies of fullerene films using atomic force microscopy, J. Mater. Res. **8**, 3019–3022 (1993)
27.60 B. Bhushan, A.V. Kulkarni, W. Bonin, J.T. Wyrobek: Nano/picoindentation measurement using a capacitance transducer system in atomic force microscopy, Philos. Mag. **74**, 1117–1128 (1996)
27.61 B. Bhushan, V.N. Koinkar: Nanoindentation hardness measurements using atomic force microscopy, Appl. Phys. Lett. **64**, 1653–1655 (1994)
27.62 X. Li, B. Bhushan: A review of nanoindentation continuous stiffness measurement technique and its applications, Mater. Charact. **48**, 11–36 (2002)
27.63 B. Bhushan, X. Li: Nanomechanical characterisation of solid surfaces and thin films, Int. Mater. Rev. **48**, 125–164 (2003), (invited)
27.64 P. Maivald, H.J. Butt, S.A.C. Gould, C.B. Prater, B. Drake, J.A. Gurley, V.B. Elings, P.K. Hansma: Using force modulation to image surface elasticities with the atomic force microscope, Nanotechnology **2**, 103–106 (1991)
27.65 B. Anczykowski, D. Kruger, K.L. Babcock, H. Fuchs: Basic properties of dynamic force microscopy with the scanning force microscope in experiment and simulation, Ultramicroscopy **66**, 251–259 (1996)
27.66 D. DeVecchio, B. Bhushan: Localized surface elasticity measurements using an atomic force microscope, Rev. Sci. Instrum. **68**, 4498–4505 (1997)
27.67 U.D. Schwarz, O. Zwoerner, P. Koester, R. Wiesendanger: Friction force spectroscopy in the low-load regime with well-defined tips. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 233–238
27.68 S. Amelio, A.V. Goldade, U. Rabe, V. Scherer, B. Bhushan, W. Arnold: Measurements of elastic properties of ultra-thin diamond-like carbon coatings using atomic force acoustic microscopy, Thin Solid Films **392**, 75–84 (2001)
27.69 W.W. Scott, B. Bhushan: Use of phase imaging in atomic force microscopy for measurement of viscoelastic contrast in polymer nanocomposites and molecularly-thick lubricant films, Ultramicroscopy **97**, 151–169 (2003)
27.70 B. Bhushan, J. Qi: Phase contrast imaging of nanocomposites and molecularly-thick lubricant films in magnetic media, Nanotechnology **14**, 886–895 (2003)
27.71 T. Kasai, B. Bhushan, L. Huang, C. Su: Topography and phase imaging using the torsional resonance mode, Nanotechnology **15**, 731–742 (2004)
27.72 N. Chen, B. Bhushan: Morphological, nanomechanical and cellular structural characterization of human hair and conditioner distribution using torsional resonance mode in an AFM, J. Micros. **220**, 96–112 (2005)
27.73 M.S. Bobji, B. Bhushan: Atomic force microscopic study of the micro-cracking of magnetic thin films under tension, Scr. Mater. **44**, 37–42 (2001)
27.74 M.S. Bobji, B. Bhushan: In situ microscopic surface characterization studies of polymeric thin films during tensile deformation using atomic force microscopy, J. Mater. Res. **16**, 844–855 (2001)
27.75 N.S. Tambe, B. Bhushan: In situ study of nanocracking of multilayered magnetic tapes under monotonic and fatigue loading using an AFM, Ultramicroscopy **100**, 359–373 (2004)
27.76 M. Palacio, B. Bhushan: Normal and lateral force calibration techniques for AFM cantilevers, Crit. Rev. Solid State Mater. Sci. **35**, 73–104 (2010)
27.77 M. Palacio, B. Bhushan: Erratum to: Normal and lateral force calibration techniques for AFM cantilevers, Crit. Rev. Solid State Mater. Sci. **36**, 261 (2010)
27.78 B. Bhushan, T. Kasai, C.V. Nguyen, M. Meyyappan: Multiwalled carbon nanotube AFM probes for surface characterization of micro/nanostructures, Microsyst. Technol. **10**, 633–639 (2004)
27.79 B. Bhushan: Micro/nanotribology and its applications to magnetic storage devices and MEMS, Tribol. Int. **28**, 85–95 (1995)
27.80 A. Avila, B. Bhushan: Electrical measurement techniques in atomic force microscopy, Crit. Rev. Solid State Mater. Sci. **35**, 38–51 (2010)
27.81 B. Bhushan, A.V. Goldade: Kelvin probe microscopy measurements of surface potential change under wear at low loads, Wear **244**, 104–117 (2000)
27.82 D.T. Lee, J.P. Pelz, B. Bhushan: Instrumentation for direct, low frequency scanning capacitance microscopy, and analysis of position dependent stray capacitance, Rev. Sci. Instrum. **73**, 3523–3533 (2002)
27.83 D. DeVecchio, B. Bhushan: Use of a nanoscale Kelvin probe for detecting wear precursors, Rev. Sci. Instrum. **69**, 3618–3624 (1998)
27.84 B. Bhushan, A.V. Goldade: Measurements and analysis of surface potential change during wear of single crystal silicon (100) at ultralow loads using kelvin probe microscopy, Appl. Surf. Sci. **157**, 373–381 (2000)
27.85 I.P. Seshadri, B. Bhushan: In-situ tensile deformation characterization of human hair with atomic force microscopy, Acta Mater. **56**, 774–781 (2008)
27.86 I.P. Seshadri, B. Bhushan: Effect of ethnicity and treatments on in situ tensile response and morphological changes of human hair characterized by atomic force microscopy, Acta Mater. **56**, 3585–3597 (2008)
27.87 H.U. Krotil, T. Stifter, H. Waschipky, K. Weishaupt, S. Hild, O. Marti: Pulse force mode: A new method for the investigation of surface properties, Surf. Interface Anal. **27**, 336–340 (1999)

27.88 U. Rabe, K. Janser, W. Arnold: Vibrations of free and surface-coupled atomic force microscope cantilevers: Theory and experiment, Rev. Sci. Instrum. **67**, 3281–3293 (1996)

27.89 V. Scherer, B. Bhushan, U. Rabe, W. Arnold: Local elasticity and lubrication measurements using atomic force and friction force microscopy at ultrasonic frequencies, IEEE Trans. Magn. **33**, 4077–4079 (1997)

27.90 J. Tamayo, R. Garcia: Deformation, contact time, and phase contrast in tapping mode scanning force microscopy, Langmuir **12**, 4430–4435 (1996)

27.91 R. Garcia, J. Tamayo, M. Calleja, F. Garcia: Phase contrast in tapping-mode scanning force microscopy, Appl. Phys. A **66**, S309–S312 (1998)

27.92 Y. Song, B. Bhushan: Quantitative extraction of in-plane surface properties using torsional resonance mode in atomic force microscopy, J. Appl. Phys. **87**, 83533 (2005)

27.93 Z. Tao, B. Bhushan: Surface modification of AFM silicon probes for adhesion and wear reduction, Tribol. Lett. **21**, 1–16 (2006)

27.94 B. Bhushan, J. Ruan, B.K. Gupta: A scanning tunnelling microscopy study of fullerene films, J. Phys. D: Appl. Phys. **26**, 1319–1322 (1993)

27.95 B. Bhushan, P.S. Mokashi, T. Ma: A new technique to measure poisson's ratio of ultrathin polymeric films using atomic force microscopy, Rev. Sci. Instrum. **74**, 1043–1047 (2003)

27.96 G.A. Tomlinson: A molecular theory of friction, Philos. Mag. **7**, 905–939 (1929)

27.97 D. Tomanek, W. Zhong, H. Thomas: Calculation of an atomically modulated friction force in atomic force microscopy, Europhys. Lett. **15**, 887–892 (1991)

27.98 E. Meyer, R. Overney, R. Luthi, D. Brodbeck, L. Howald, J. Frommer, H.J. Guntherodt, O. Wolter, M. Fujihira, T. Takano, Y. Gotoh: Friction force microscopy of mixed Langmuir-Blodgett films, Thin Solid Films **220**, 132–137 (1992)

27.99 C.D. Frisbie, L.F. Rozsnyai, A. Noy, M.S. Wrighton, C.M. Lieber: Functional group imaging by chemical force microscopy, Science **265**, 2071–2074 (1994)

27.100 V.N. Koinkar, B. Bhushan: Scanning and transmission electron microscopies of single-crystal silicon microworn/machined using atomic force microscopy, J. Mater. Res. **12**, 3219–3224 (1997)

27.101 S. Sundararajan, B. Bhushan: Topography-induced contributions to friction forces measured using an atomic force/friction force microscope, J. Appl. Phys. **88**, 4825–4831 (2000)

27.102 V.N. Koinkar, B. Bhushan: Effect of scan size and surface roughness on microscale friction measurements, J. Appl. Phys. **81**, 2472–2479 (1997)

27.103 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)

27.104 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

27.105 B. Bhushan, G.S. Blackman: Atomic force microscopy of magnetic rigid disks and sliders and its applications to tribology, ASME J. Tribol. **113**, 452–458 (1991)

27.106 K. Yamanaka, E. Tomita: Lateral force modulation atomic force microscope for selective imaging of friction forces, Jpn. J. Appl. Phys. **34**, 2879–2882 (1995)

27.107 I.P. Seshadri, B. Bhushan: Effect of rubbing load on nanoscale charging characteristics of human hair characterized by AFM based Kelvin probe, J. Colloid Interface Sci. **325**, 580–587 (2008)

27.108 O. Marti, H.-U. Krotil: Dynamic friction measurement with the scanning force microscope. In: *Fundamentals of Tribology and Bridging the Gap Between the Macro- and Micro/Nanoscales*, ed. by B. Bhushan (Kluwer, Dordrecht 2001) pp. 121–135

27.109 Z. Tao, B. Bhushan: A new technique for studying nanoscale friction at sliding velocities up to 200 mm/s using atomic force microscope, Rev. Sci. Instrum. **77**, 103705 (2006)

27.110 N.S. Tambe, B. Bhushan: Scale dependence of micro/nano-friction and adhesion of MEMS/NEMS materials, coatings and lubricants, Nanotechnology **15**, 1561–1570 (2004)

27.111 N.S. Tambe, B. Bhushan: Friction model for the velocity dependence of nanoscale friction, Nanotechnology **16**, 2309–2324 (2005)

27.112 N.S. Tambe, B. Bhushan: Durability studies of micro/nanoelectromechanical system materials, coatings, and lubricants at high sliding velocities (up to 10 mm/s) using a modified atomic force microscope, J. Vac. Sci. Technol. A **23**, 830–835 (2005)

27.113 N.S. Tambe, B. Bhushan: Identifying materials with low friction and adhesion for nanotechnology applications, Appl. Phys. Lett. **86**, 061906 (2005)

27.114 Z. Tao, B. Bhushan: Velocity dependence and rest time effect in nanoscale friction of ultrathin films at high sliding velocities, J. Vac. Sci. Technol. A **25**, 1267–1274 (2007)

27.115 O. Zworner, H. Holscher, U.D. Schwarz, R. Wiesendanger: The velocity dependence of frictional forces in point-contact friction, Appl. Phys. A **66**, S263–S267 (1998)

27.116 E. Gnecco, R. Bennewitz, T. Gyalog, C. Loppacher, M. Bammerlin, E. Meyer, H.-J. Guntherodt: Velocity dependence of atomic friction, Phys. Rev. Lett. **84**, 1172–1175 (2000)

27.117 J.S. Helman, W. Baltensperger, J.A. Holyst: Simple-model for dry friction, Phys. Rev. B **49**, 3831–3838 (1994)

27.118 C. Fusco, A. Fasolino: Velocity dependence of atomic-scale friction: A comparative study of the one- and two-dimensional tomlinson model, Phys. Rev. B **71**, 045413 (2005)

27.119 N.S. Tambe, B. Bhushan: Nanoscale friction and wear maps, Philos. Trans. R. Soc. A **366**, 1405–1424 (2008)

27.120 N.S. Tambe, B. Bhushan: Nanoscale friction mapping, Appl. Phys. Lett. **86**, 193102 (2005)

27.121 N.S. Tambe, B. Bhushan: Nanoscale friction-induced phase transformation of diamond-like carbon, Scr. Mater. **52**, 751–755 (2005)

27.122 S.C. Lim, M.F. Ashby: Wear mechanism maps, Acta Metall. **35**, 1–24 (1987)

27.123 S.C. Lim, M.F. Ashby, J.H. Brunton: Wear-rate transitions and their relationship to wear mechanisms, Acta Metall. **35**, 1343–1348 (1987)

27.124 N.S. Tambe, B. Bhushan: Nanowear mapping: A novel atomic force microscopy based approach for studying nanoscale wear at high sliding velocities, Tribol. Lett. **20**, 83–90 (2005)

27.125 B. Bhushan, C. Dandavate: Thin-film friction and adhesion studies using atomic force microscopy, J. Appl. Phys. **87**, 1201–1210 (2000)

27.126 T. Stifter, O. Marti, B. Bhushan: Theoretical investigation of the distance dependence of capillary and van der Waals forces in scanning probe microscopy, Phys. Rev. B **62**, 13667–13673 (2000)

27.127 B. Bhushan, H. Liu, S.M. Hsu: Adhesion and friction studies of silicon and hydrophobic and low friction films and investigation of scale effects, ASME J. Tribol. **126**, 583–590 (2004)

27.128 H. Liu, B. Bhushan: Adhesion and friction studies of microelectromechanical systems/nanoelectromechanical systems materials using a novel microtriboapparatus, J. Vac. Sci. Technol. A **21**, 1528–1538 (2003)

27.129 B. Bhushan, B.K. Gupta: *Handbook of Tribology: Materials, Coatings and Surface Treatments* (Krieger, Malabar 1997)

27.130 B. Bhushan, S. Venkatesan: Mechanical and tribological properties of silicon for micromechanical applications: A review, Adv. Inf. Storage Syst. **5**, 211–239 (1993)

27.131 Properties of Silicon, EMIS Data Reviews Series No. 4 (INSPEC, Institution of Electrical Engineers, London 1988)

27.132 J.E. Field: *The Properties of Natural and Synthetic Diamond* (Academic, London 1992)

27.133 National Carbon Company: *The Industrial Graphite Engineering Handbook* (National Carbon Company, New York 1959)

27.134 M. Nosonovsky, B. Bhushan: Scale effects in dry friction during multiple-asperity contact, ASME J. Tribol. **127**, 37–46 (2005)

27.135 B. Bhushan, M. Nosonovsky: Scale effects in friction using strain gradient plasticity and dislocation-assisted sliding (microslip), Acta Mater. **51**, 4331–4345 (2003)

27.136 B. Bhushan, M. Nosonovsky: Comprehensive model for scale effects in friction due to adhesion and two- and three-body deformation (plowing), Acta Mater. **52**, 2461–2474 (2004)

27.137 B. Bhushan, M. Nosonovsky: Scale effects in dry and wet friction, wear, and interface temperature, Nanotechnology **15**, 749–761 (2004)

27.138 X. Zhao, B. Bhushan: Material removal mechanism of single-crystal silicon on nanoscale and at ultralow loads, Wear **223**, 66–78 (1998)

27.139 W. Tang, B. Bhushan, S. Ge: Triboelectrification studies of skin and skin cream using Kelvin probe microscopy, J. Vac. Sci. Technol. A **28**, 1018–1028 (2010)

27.140 R.A. Lodge, B. Bhushan: Effect of physical wear and triboelectric interaction on surface charges measured by kelvin probe microscopy, J. Colloid Interface Sci. **310**, 321–330 (2007)

27.141 S.C. Nagpure, B. Bhushan, S.S. Babu: Surface potential measurement of aged li-ion batteries using kelvin probe microscopy, J. Power Sources **196**, 1508–1512 (2011)

27.142 U. Zaghloul, B. Bhushan, F. Coccetti, P. Pons, R. Plana: Kelvin probe force microscopy based characterization techniques applied for electrostatic MEMS and thin dielectric films to investigate the dielectric substrate charging phenomena, J. Vac. Sci. Technol. A **29**, 051101 (2011)

27.143 A.V. Kulkarni, B. Bhushan: Nanoscale mechanical property measurements using modified atomic force microscopy, Thin Solid Films **290/291**, 206–210 (1996)

27.144 A.V. Kulkarni, B. Bhushan: Nano/picoindentation measurements on single-crystal aluminum using modified atomic force microscopy, Mater. Lett. **29**, 221–227 (1996)

27.145 A.V. Kulkarni, B. Bhushan: Nanoindentation measurement of amorphous carbon coatings, J. Mater. Res. **12**, 2707–2714 (1997)

27.146 N.A. Fleck, G.M. Muller, M.F. Ashby, J.W. Hutchinson: Strain gradient plasticity: Theory and experiment, Acta Metall. Mater. **42**, 475–487 (1994)

27.147 W.D. Nix, H. Gao: Indentation size effects in crystalline materials: A law for strain gradient plasticity, J. Mech. Phys. Solids **46**, 411–425 (1998)

27.148 W.B. Li, J.L. Henshall, R.M. Hooper, K.E. Easterling: The mechanism of indentation creep, Acta Metall. Mater. **39**, 3099–3110 (1991)

27.149 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids, Part 1* (Clarendon, Oxford 1950)

27.150 B. Bhushan, M. Palacio, B. Kinzig: AFM-based nanotribological and electrical characterization of ultrathin wear-resistant ionic liquid films, J. Colloid Interface. Sci. **317**, 275–287 (2008)

27.151 M. Palacio, B. Bhushan: Ultrathin wear-resistant ionic liquid films for novel MEMS/NEMS applications, Adv. Mater. **20**, 1194–1198 (2008)

27.152 M. Palacio, B. Bhushan: Molecularly thick dicationic liquid films for nanolubrication, J. Vac. Sci. Technol. A **27**, 986–995 (2009)

27.153 E. Hoque, J.A. DeRose, P. Hoffmann, H.J. Mathieu, B. Bhushan, M. Cichomski: Phosphonate self-assembled monolayers on aluminum surfaces, J. Chem. Phys. **124**, 174710 (2006)

27.154 E. Hoque, J.A. DeRose, G. Kulik, P. Hoffmann, H.J. Mathieu, B. Bhushan: Alkylphosphonate modified aluminum oxide surfaces, J. Phys. Chem. B **110**, 10855–10861 (2006)

27.155 E. Hoque, J.A. DeRose, P. Hoffmann, B. Bhushan, H.J. Mathieu: Alkylperfluorosilane self-assembled monolayers on aluminum: A comparison with alkylphosphonate self-assembled monolayers, J. Phys. Chem. C **111**, 3956–3962 (2007)

27.156 E. Hoque, J.A. DeRose, P. Hoffmann, B. Bhushan, H.J. Mathieu: Chemical stability of nonwetting, low adhesion self-assembled monolayer films

formed by perfluoroalkylsilazation of copper, J. Chem. Phys. **126**, 114706 (2007)

27.157 E. Hoque, J.A. DeRose, B. Bhushan, H.J. Mathieu: Self-assembled monolayers on aluminum and copper oxide surfaces: Surface and interface characteristics, nanotribological properties, and chemical stability. In: *Applied Scanning Probe Methods Vol. IX – Characterization*, ed. by B. Bhushan, H. Fuchs, M. Tomitori (Springer, Berlin, Heidelberg 2008) pp. 235–281

27.158 E. Hoque, J.A. DeRose, B. Bhushan, K.W. Hipps: Low adhesion, non-wetting phosphonate self-assembled monolayer films formed on copper oxide surfaces, Ultramicroscopy **109**, 1015–1022 (2009)

27.159 J.A. DeRose, E. Hoque, B. Bhushan, H.J. Mathieu: Characterization of perfluorodecanote self-assembled monolayers on aluminum and comparison of stability with phosphonate and siloxy self-assembled monolayers, Surf. Sci. **602**, 1360–1367 (2008)

28. Surface Forces and Nanorheology of Molecularly Thin Films

Dong Woog Lee, Marina Ruths, Jacob N. Israelachvili

In this chapter, we describe the static and dynamic normal forces that occur between surfaces in vacuum or liquid and the different modes of friction that can be observed between:

1. Bare surfaces in contact (dry or interfacial friction)
2. Surfaces separated by a thin liquid film (lubricated friction)
3. Surfaces coated with organic monolayers (boundary friction).

Experimental methods suitable for measuring normal surface forces, and adhesion and friction (lateral or shear) forces of different magnitude at the molecular level are described. We explain the molecular origin of van der Waals, electrostatic, solvation and polymer-mediated interactions, and basic models for the contact mechanics of adhesive and nonadhesive elastically deforming bodies. The effects of interaction forces, molecular shape, surface structure and roughness on adhesion and friction are discussed.

Simple models for the contributions of the adhesion force and external load to interfacial friction are illustrated with experimental data on both unlubricated and lubricated systems, as measured with the surface forces apparatus. We discuss rate-dependent adhesion (adhesion hysteresis) and how this is related to friction. Some examples of the transition from wearless friction to friction with wear are shown.

Lubrication in different lubricant thickness regimes is described together with explanations of nanorheological concepts. The occurrence of and transitions between smooth and stick–slip sliding in various types of dry (unlubricated and solid boundary lubricated) and liquid lubricated systems are discussed based on recent experimental results and models for stick–slip involving memory distance and dilatancy.

28.1 **Types of Surface Forces** 936

28.2 **Methods Used to Study Surface Forces** 938
28.2.1 Force Laws 938
28.2.2 Adhesion Forces 938
28.2.3 The SFA 940

28.3 **Normal Forces Between Dry (Unlubricated) Surfaces** 941
28.3.1 Van Der Waals Forces in Vacuum and Inert Vapors 941

28.4 **Normal Forces Between Surfaces in Liquids** 942
28.4.1 Van Der Waals Forces in Liquids 942
28.4.2 Electrostatic Forces 943
28.4.3 Solvation and Structural Forces 945
28.4.4 Hydration and Hydrophobic Forces 947
28.4.5 Polymer–Mediated Forces 949

28.5 **Adhesion and Capillary Forces** 951
28.5.1 Capillary Forces 951
28.5.2 Adhesion Mechanics 952
28.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch 952
28.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis 954

28.6 **Introduction: Different Modes of Friction** 956

28.7 **Relationship Between Adhesion and Friction** 957
28.7.1 Amontons' Law 957
28.7.2 Adhesion Force and Load Contribution to Interfacial Friction 958
28.7.3 Examples of Experimentally Observed Friction of Dry Surfaces 962
28.7.4 Transition from Interfacial to Normal Friction with Wear 965

28.8 **Liquid Lubricated Surfaces** 966
28.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime 966
28.8.2 Friction of Intermediate Thickness Films 967

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_28

28.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology 969
28.9 **Effects of Nanoscale Texture on Friction** 976
28.9.1 Role of the Shape of Confined Molecules 976
28.9.2 Effects of Surface Structure 978
References 978

28.1 Types of Surface Forces

In this chapter, we discuss the most important types of surface forces and the relevant equations for the force and friction laws. Several different attractive and repulsive forces operate between surfaces and particles. Some forces occur in vacuum, for example attractive van der Waals and repulsive hard-core interactions. Other types of forces can arise only when the interacting surfaces are separated by another condensed phase, which is usually a liquid. The most common types of surface forces and their main characteristics are listed in Table 28.1.

In *vacuum*, the two main long-range interactions are the attractive van der Waals and electrostatic (Coulomb) forces. At smaller surface separations (corresponding to molecular contact at surface separations of $D \approx 0.2\,\text{nm}$), additional attractive interactions can be found such as covalent or metallic bonding forces. These attractive forces are stabilized by the hard-core repulsion. Together they determine the surface and interfacial energies of planar surfaces, as well as the strengths of materials and adhesive junctions. Adhesion forces are often strong enough to elastically or plastically deform bodies or particles when they come into contact.

In *vapors* (e.g., atmospheric air containing water and organic molecules), solid surfaces in, or close to, contact will generally have a surface layer of chemisorbed or physisorbed molecules, or a capillary condensed liquid bridge between them. A surface layer usually causes the adhesion to decrease, but in the case of capillary condensation, the additional Laplace pressure or attractive *capillary* force may make the adhesion between the surfaces stronger than in an inert gas or vacuum.

When totally immersed in a *liquid*, the force between particles or surfaces is completely modified from that in vacuum or air (vapor). The van der Waals attraction is generally reduced, but other forces can now arise that can qualitatively change both the range and even the sign of the interaction. The attractive force in such a system can be either stronger or weaker than in the absence of the intervening liquid. Depending on the different forces that may be operating simultaneously in solution, the overall force law is not generally monotonically attractive even at long range; it can be repulsive, or the force can change sign at some finite surface separation. In such cases, the potential-energy minimum, which determines the adhesion force or energy, does not occur at the true molecular contact between the surfaces, but at some small distance further out.

The forces between surfaces in a liquid medium can be particularly complex at *short range*, i. e., at surface separations below a few nanometers or 4–10 molecular diameters. This is partly because, with increasing confinement, a liquid ceases to behave as a structureless continuum with bulk properties; instead, the size and shape of its molecules begin to determine the overall interaction. In addition, the surfaces themselves can no longer be treated as inert and structureless walls (i. e., mathematically flat) and their physical and chemical properties at the atomic scale must now be taken into account. The force laws will then depend on whether the surfaces are amorphous or crystalline (and whether the lattices of crystalline surfaces are matched or not), rough or smooth, rigid or soft (fluid-like), and hydrophobic or hydrophilic.

It is also important to distinguish between *static* (i. e., equilibrium) interactions and *dynamic* (i. e., nonequilibrium) forces such as viscous and friction forces. For example, certain liquid films confined between two contacting surfaces may take a surprisingly long time to equilibrate, as may the surfaces themselves, so that the short-range and adhesion forces appear to be time dependent, resulting in aging effects.

Table 28.1 Types of surface forces in vacuum versus in liquid (colloidal forces). Note: (v) applies only to interactions in vacuum, (s) applies only to interactions in solution (or to surfaces separated by a liquid), and (v&s) applies to interactions occurring both in vacuum and in solution

Type of force	Subclasses or alternative names	Main characteristics
	Attractive forces	
van der Waals	Debye induced dipole force (v&s) London dispersion force (v&s) Casimir force (v&s)	Ubiquitous, occurs both in vacuum and in liquids
Electrostatic	Ionic bond (v) Coulombic force (v&s) Hydrogen bond (v) Charge-exchange interaction (v&s) Acid–base interaction (s) *Harpooning* interaction (v)	Strong, long-range, arises in polar solvents; requires surface charging or charge-separation mechanism
Ion correlation	van der Waals force of polarizable ions (s)	Requires mobile charges on surfaces in a polar solvent
Quantum mechanical	Covalent bond (v) Metallic bond (v) Exchange interaction (v)	Strong, short-range, responsible for contact binding of crystalline surfaces
Solvation	Oscillatory force (s) Depletion force (s)	Mainly entropic in origin, the oscillatory force alternates between attraction and repulsion
Hydrophobic	Attractive hydration force (s)	Strong, apparently long-range; origin not yet understood
Specific binding	*Lock-and-key* or complementary binding (v&s) Receptor–ligand interaction (s) Antibody–antigen interaction (s)	Subtle combination of different noncovalent forces giving rise to highly specific binding; main recognition mechanism of biological systems
	Repulsive forces	
van der Waals	van der Waals disjoining pressure (s)	Arises only between dissimilar bodies interacting in a medium
Electrostatic	Coulombic force (v&s)	Arises only for certain constrained surface charge distributions
Quantum mechanical	Hard-core or steric repulsion (v) Born repulsion (v)	Short-range, stabilizing attractive covalent and ionic binding forces, effectively determine molecular size and shape
Solvation	Oscillatory solvation force (s) Structural force (s) Hydration force (s)	Monotonically repulsive forces, believed to arise when solvent molecules bind strongly to surfaces
Entropic	Osmotic repulsion (s) Double-layer force (s) Thermal fluctuation force (s) Steric polymer repulsion (s) Undulation force (s) Protrusion force (s)	Due to confinement of molecular or ionic species; requires mechanism that keeps trapped species between the surfaces
	Dynamic interactions	
Nonequilibrium	Hydrodynamic forces (s) Viscous forces (s) Friction forces (v&s) Lubrication forces (s)	Energy-dissipating forces occurring during relative motion of surfaces or bodies

28.2 Methods Used to Study Surface Forces

28.2.1 Force Laws

The full force law $F(D)$ between two surfaces, i. e., the force F as a function of surface separation D, can be measured in a number of ways [28.1–3]. The simplest is to move the base of a spring by a known amount ΔD_0. Figure 28.1 illustrates this method when applied to the interaction of two magnets. However, the method is also applicable at the microscopic or molecular level, and it forms the basis of all direct force-measuring apparatuses such as the surface forces apparatus (SFA; [28.1, 4]). If there is a detectable force between the surfaces, this will cause the force-measuring spring to deflect by ΔD_s, while the surface separation changes by ΔD. These three displacements are related by

$$\Delta D_s = \Delta D_0 - \Delta D \,. \tag{28.1}$$

The difference in force, ΔF, between the initial and final separations is given by

$$\Delta F = k_s \Delta D_s \,, \tag{28.2}$$

where k_s is the spring constant. The equations above provide the basis for measurements of the force difference between any two surface separations. For example, if a force-measuring apparatus with a known k_s can measure D (and thus ΔD), ΔD_0, and ΔD_s, the force difference ΔF can be measured between a large initial or reference separation D, where the force is zero ($F = 0$), and another separation $D - \Delta D$. By working one's way in increasing increments of $\Delta D = \Delta D_0 - \Delta D_s$, the full force law $F(D)$ can be constructed over any desired distance regime.

In order to measure an equilibrium force law, it is essential to establish that the two surfaces have stopped moving before the displacements are measured. When displacements are measured while two surfaces are still in relative motion, one also measures a viscous or frictional contribution to the total force. Such dynamic force measurements have enabled the viscosities of liquids near surfaces and in thin films to be accurately determined [28.6, 7]. In practice, it is difficult to measure the forces between two perfectly flat surfaces, because of the stringent requirement of perfect alignment for making reliable measurements at distances of a few tenths of a nanometer. It is far easier to measure the forces between curved surfaces, e.g., two spheres, a sphere and a flat surface, or two crossed cylinders. Furthermore, the force $F(D)$ measured between two curved surfaces can be directly related to the energy per unit area $E(D)$ between two flat surfaces at the same separation D by the so-called Derjaguin approximation [28.8]

$$E(D) = \frac{F(D)}{2\pi R} \,, \tag{28.3}$$

where R is the radius of the sphere (for a sphere and a flat surface) or the radii of the cylinders (for two crossed cylinders, Table 28.2).

28.2.2 Adhesion Forces

The most direct way to measure the adhesion of two solid surfaces (such as two spheres or a sphere on a flat surface) is to suspend one of them on a spring and measure the adhesion or pull-off force needed to separate

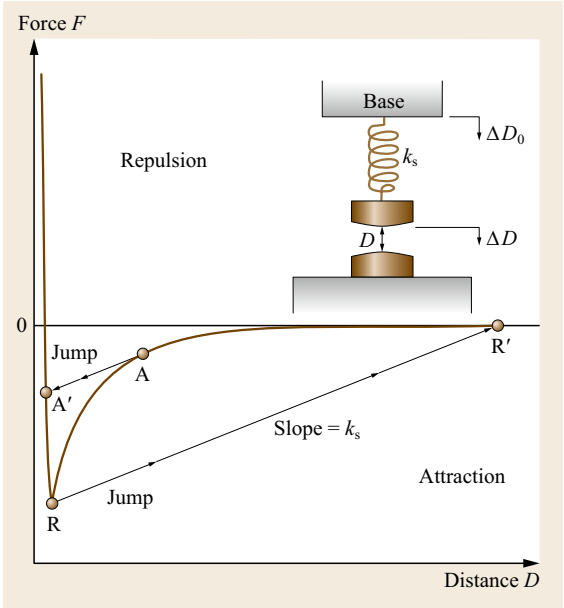
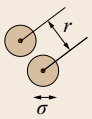
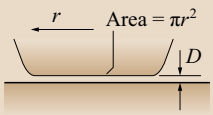
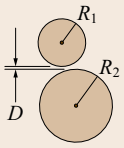
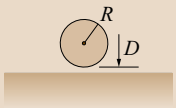
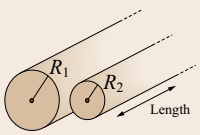
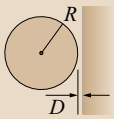
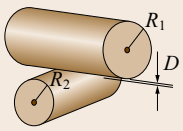


Fig. 28.1 Schematic attractive force law between two macroscopic objects such as two magnets, or between two microscopic objects such as the van der Waals force between a metal tip and a surface. On lowering the base supporting the spring, the latter will expand or contract such that, at any equilibrium separation D, the attractive force balances the elastic spring restoring force. If the gradient of the attractive force dF/dD exceeds the gradient of the spring's restoring force (defined by the spring constant k_s), the upper surface will jump from A into contact at A′ (A for advancing). On separating the surfaces by raising the base, the two surfaces will jump apart from R to R′ (R for receding). The distance R–R′ multiplied by k_s gives the adhesion force, i. e., the value of F at the point R (after [28.5] with permission)

Table 28.2 Van der Waals interaction energy and force between macroscopic bodies of different geometries

Geometry of bodies with surfaces D apart ($D \ll R$)			van der Waals interaction	
			Energy E	Force F
Two atoms or small molecules		$r \geq \sigma$	$\frac{-C_{vdW}}{r^6}$	$\frac{-6C_{vdW}}{r^7}$
Two flat surfaces (per unit area)		$r \gg D$	$\frac{-A_H}{12\pi D^2}$	$\frac{-A_H}{6\pi D^3}$
Two spheres or macromolecules of radii R_1 and R_2		$R_1, R_2 \gg D$	$\frac{-A_H}{6D}\left(\frac{R_1R_2}{R_1+R_2}\right)$	$\frac{-A_H}{6D^2}\left(\frac{R_1R_2}{R_1+R_2}\right)$
Sphere or macromolecule of radius R near a flat surface		$R \gg D$	$\frac{-A_H R}{6D}$	$\frac{-A_H R}{6D^2}$
Two parallel cylinders or rods of radii R_1 and R_2 (per unit length)		$R_1, R_2 \gg D$	$\frac{-A_H}{12\sqrt{2}D^{3/2}}\left(\frac{R_1R_2}{R_1+R_2}\right)^{1/2}$	$\frac{-A_H}{8\sqrt{2}D^{5/2}}\left(\frac{R_1R_2}{R_1+R_2}\right)^{1/2}$
Cylinder of radius R near a flat surface (per unit length)		$R \gg D$	$\frac{-A_H\sqrt{R}}{12\sqrt{2}D^{3/2}}$	$\frac{-A_H\sqrt{R}}{8\sqrt{2}D^{5/2}}$
Two cylinders or filaments of radii R_1 and R_2 crossed at 90°		$R_1, R_2 \gg D$	$\frac{-A_H\sqrt{R_1R_2}}{6D}$	$\frac{-A_H\sqrt{R_1R_2}}{6D^2}$

A negative force (A_H positive) implies attraction, a positive force means repulsion (A_H negative) (after [28.9], with permission)

the two bodies, using the deflection of the spring. If k_s is the stiffness of the force-measuring spring and ΔD is the distance the two surfaces jump apart when they separate, then the adhesion force F_{ad} is given by

$$F_{ad} = F_{max} = k_s \Delta D \,, \tag{28.4}$$

where we note that, in liquids, the maximum or minimum in the force may occur at some nonzero surface separation. From F_{ad} and a known surface geometry, and assuming that the surfaces were everywhere in molecular contact, one may also calculate the surface or interfacial energy γ. For an elastically deformable sphere of radius R on a flat surface, or for two crossed cylinders of radius R, we have [28.1, 10]

$$\gamma = \frac{F_{ad}}{3\pi R} \,, \tag{28.5}$$

while for two spheres of radii R_1 and R_2

$$\gamma = \frac{F_{ad}}{3\pi}\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{28.6}$$

where γ is in units of $\mathrm{J\,m^{-2}}$ (Sect. 28.5.2).

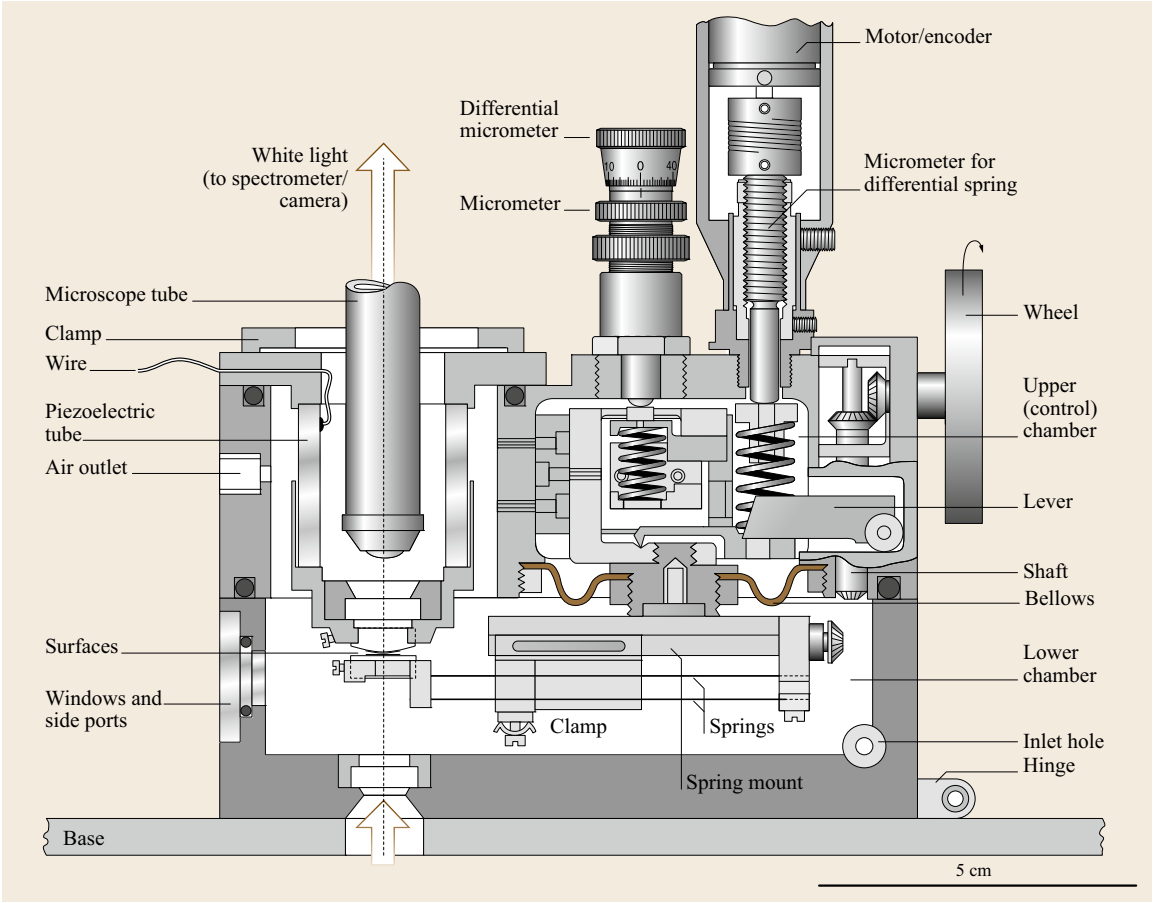


Fig. 28.2 An SFA where the intermolecular forces between two macroscopic, cylindrical surfaces of local radius R can be directly measured as a function of surface separation over a large distance regime from tenths of a nanometer to micrometers. Local or transient surface deformations can be detected optically. Various attachments for moving one surface laterally with respect to the other have been developed for friction measurements in different regimes of sliding velocity and sliding distance (after [28.11] with permission)

28.2.3 The SFA

The SFA (Fig. 28.2) is used for measurements of adhesion and force laws between two curved molecularly smooth surfaces immersed in liquids or controlled vapors [28.1, 4, 11]. In SFA experiments, ΔD_0 is changed by expanding or contracting a piezoelectric crystal by a known amount or by moving the base of the spring with sensitive motor-driven mechanical stages. The resulting change in surface separation ΔD is measured by multiple beam interferometry with an accuracy of ± 0.1 nm, and the spring deflection ΔD_s can then be obtained according to (28.1). From the shape of the interference fringes one also obtains the radius of the surfaces R and any surface deformation that arises during an interaction [28.12, 13]. The resolution in the lateral direction is about 1 μm. The surface separation can be independently controlled to within 0.1 nm, and the force sensitivity is about 10^{-10} N. For a typical surface radius of $R \approx 1$ cm, γ values can be measured to an accuracy of about 10^{-3} mJ m^{-2}.

Several different materials have been used to form the surfaces in the SFA, including mica [28.14, 15], silica [28.16], sapphire [28.17], and polymer sheets [28.18]. These materials can also be used as supporting substrates in experiments on the forces between adsorbed or chemically bound polymer layers [28.7, 19–22], surfactant and lipid monolayers and bilayers [28.9, 23–26], and metal and metal oxide layers [28.27–31]. The range of liquids and vapors that can be used is almost endless, and has thus far included aqueous solutions, organic liquids and solvents, polymer melts, various petroleum oils and lubricant liquids, dyes, and liquid crystals.

Friction attachments for the SFA [28.32–36] allow for the two surfaces to be sheared laterally past each other at varying sliding speeds or oscillating frequencies, while simultaneously measuring both the transverse (frictional or shear) force and the normal force (load) between them. The ranges of friction forces and sliding speeds that can be studied with such methods are currently 10^{-7}–10^{-1} N and 10^{-13}–10^{-2} m s^{-1} respectively [28.37]. The externally applied load L can be varied continuously, and both positive and negative loads can be applied. The distance between the surfaces D, their true molecular contact area, their elastic (or viscoelastic or elastohydrodynamic) deformation, and their lateral motion can all be monitored simultaneously by recording the moving interference fringe pattern. Equipment for dynamic measurements of normal forces has also been developed. Such measurements give information on the viscosity of the medium and the location of the shear or slip planes relative to the surfaces [28.6, 7].

28.3 Normal Forces Between Dry (Unlubricated) Surfaces

28.3.1 Van Der Waals Forces in Vacuum and Inert Vapors

Forces between macroscopic bodies (such as colloidal particles) across vacuum arise from interactions between the constituent atoms or molecules of each body across the gap separating them. These intermolecular interactions are electromagnetic forces between permanent or induced dipoles (van der Waals forces), and between ions (electrostatic forces). In certain systems, other types of interactions can also be observed between surfaces in vacuum, e.g., charge exchange interactions, sintering, and cold welding. In this section, we describe the van der Waals forces, which occur between all atoms and molecules and between all macroscopic bodies [28.1].

The interaction between two permanent dipoles with a fixed relative orientation can be attractive or repulsive. For the specific case of two freely rotating permanent dipoles in a liquid or vapor (orientational or Keesom interaction), and for a permanent dipole and an induced dipole in an atom or polar or nonpolar molecule (induction or Debye interaction), the interaction is on average always attractive. The third type of van der Waals interaction, the fluctuation or London dispersion interaction, arises from instantaneous polarization of one nonpolar or polar molecule due to fluctuations in the charge distribution of a neighboring nonpolar or polar molecule (Fig. 28.3). Correlation between these fluctuation-induced dipole moments gives an attraction that is present between any two molecules or surfaces across vacuum. At very small separations, the interaction will ultimately be repulsive as the electron clouds of atoms and molecules begin to overlap. The total interaction is thus a combination of a short-range repulsion and a relatively long-range attraction.

Except for in highly polar materials such as water, London dispersion interactions give the largest contribution (70%–100%) to the van der Waals attraction. The interaction energy of the van der Waals force between atoms or molecules depends on the separation r as

$$E(D) = -\frac{C_{\mathrm{vdW}}}{r^6}\,, \tag{28.7}$$

where the constant C_{vdW} depends on the dipole moments and polarizabilities of the molecules. At large separations (> 10 nm), the London interaction is weakened by a randomizing effect caused by the rapid fluctuations. That is, the induced temporary dipole moment of one molecule may have changed during the time needed for the transmission of the electromagnetic wave (photon) generated by its fluctuating charge density to another molecule and the return of the photon generated by the induced fluctuation in this second molecule. This phenomenon is called retardation and causes the interaction energy to decay as r^{-7} at large separations [28.38].

Dispersion interactions are to a first approximation additive, and their contribution to the interac-

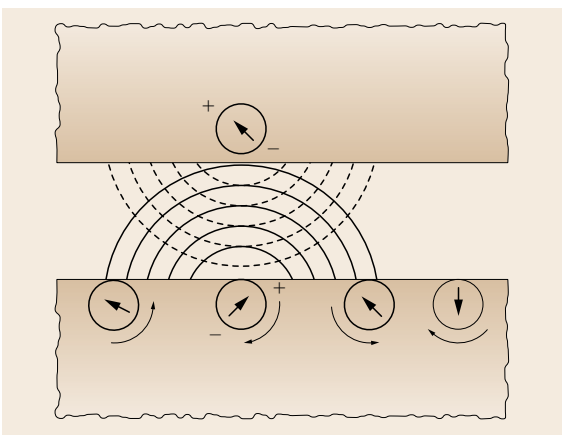


Fig. 28.3 Schematic representation of van der Waals interaction (dipole–induced dipole interaction)

tion energy between two macroscopic bodies (such as colloidal particles) across vacuum can be found by summing the pairwise interactions [28.39]. The interaction is generally described in terms of the Hamaker constant, A_H. Another approach is to treat the interacting bodies and an intervening medium as continuous phases and determine the strength of the interaction from bulk dielectric properties of the materials [28.40, 41]. Unlike the pairwise summation, this method takes into account the screening of the interactions between molecules inside the bodies by the molecules closer to the surfaces and the effects of the intervening medium. For the interaction between material 1 and material 3 across material 2, the nonretarded Hamaker constant given by the Lifshitz theory is approximately [28.1]

$$A_{H,123} = A_{H,\nu=0} + A_{H,\nu>0} \approx \frac{3}{4} k_B T \left(\frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2}\right)\left(\frac{\varepsilon_3 - \varepsilon_2}{\varepsilon_3 + \varepsilon_2}\right) + \frac{3h\nu_e}{8\sqrt{2}} \times \frac{\left(n_1^2 - n_2^2\right)\left(n_3^2 - n_2^2\right)}{\sqrt{\left(n_1^2 + n_2^2\right)}\sqrt{\left(n_3^2 + n_2^2\right)} \times \left(\sqrt{\left(n_1^2 + n_2^2\right)} + \sqrt{\left(n_3^2 + n_2^2\right)}\right)}, \quad (28.8)$$

where the first term ($\nu = 0$) represents the permanent dipole and dipole–induced dipole interactions and the second ($\nu > 0$) the London (dispersion) interaction. ε_i and n_i are the static dielectric constants and refractive indexes of the materials respectively. ν_e is the frequency of the lowest electron transition (around $3 \times 10^{15}\,s^{-1}$). Either one of the materials 1, 2, or 3 in (28.8) can be vacuum or air ($\varepsilon = n = 1$). A_H is typically 10^{-20}–10^{-19} J (the higher values are found for metals) for interactions between solids and liquids across vacuum or air.

The interaction energy between two macroscopic bodies is dependent on the geometry and is always attractive between two bodies of the same material (A_H positive, see (28.8)). The van der Waals interaction energy and force laws ($F = -dE(D)/dD$) for some common geometries are given in Table 28.2. Because of the retardation effect, the equations in Table 28.2 will lead to an overestimation of the dispersion force at large separations. It is, however, apparent that the interaction energy between macroscopic bodies decays more slowly with separation (i. e., has a longer range) than between two molecules. For inert nonpolar surfaces, e.g., consisting of hydrocarbons or van der Waals solids and liquids, the Lifshitz theory has been found to apply even at molecular contact, where it can be used to predict the surface energies (surface tensions) of such solids and liquids. For example, for hydrocarbon surfaces, $A_H = 5 \times 10^{-20}$ J. Inserting this value into the equation for two flat surfaces (Table 28.2) and using a cut-off distance of $D_0 \approx 0.165$ nm as an effective separation when the surfaces are in contact [28.1], we obtain for the surface energy γ (which is defined as half the interaction energy)

$$\gamma = \frac{E}{2} = \frac{A_H}{24\pi D_0^2} \approx 24\,\mathrm{mJ\,m^{-2}}, \quad (28.9)$$

a value that is typical for hydrocarbon solids and liquids [28.42]. If the adhesion force is measured between two crossed-cylindrical surfaces of $R = 1$ cm (a geometry equivalent to a sphere with $R = 1$ cm interacting with a flat surface, Table 28.2) using an SFA, we expect the adhesion force to be (Table 28.2) $F = A_H R/(6D_0^2) = 4\pi R\gamma \approx 3.0$ mN. Using a spring constant of $k_s = 100\,\mathrm{N\,m^{-1}}$, such an adhesive force will cause the two surfaces to jump apart by $\Delta D = F/k_s = 30\,\mu$m, which can be accurately measured. (For elastic bodies that deform in adhesive contact, R changes during the interaction and the measured adhesion force is 25% lower, see Sect. 28.5.2). Surface energies of solids can thus be directly measured with the SFA. The measured values are in good agreement with calculated values based on the known surface energies γ of the materials, and for nonpolar low-energy solids they are well accounted for by the Lifshitz theory [28.1].

28.4 Normal Forces Between Surfaces in Liquids

28.4.1 Van Der Waals Forces in Liquids

The dispersion interaction in a medium will be significantly lower than in vacuum, since the attractive interaction between two solute molecules in a medium (solvent) involves displacement and reorientation of the nearest-neighbor solvent molecules. Even though the surrounding medium may change the dipole moment and polarizability from that in vacuum, the interaction between two identical molecules remains attractive in a binary mixture. The extension of the interactions to the case of two macroscopic bodies is the same as described in Sect. 28.3.1. Typically, the Hamaker constants for interactions in a medium are an order of magnitude lower than in vacuum. Between macroscopic surfaces in liquids, van der Waals forces become im-

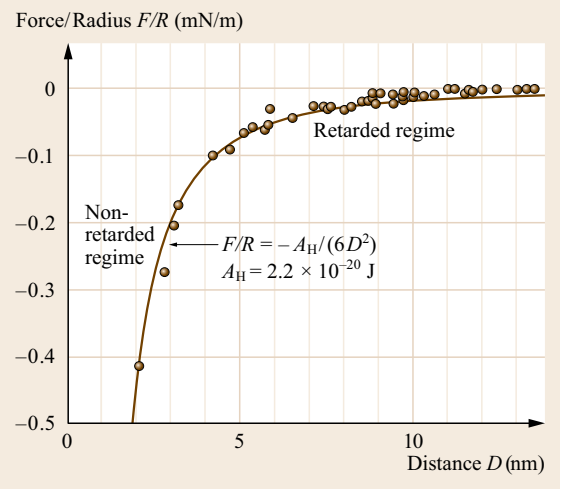


Fig. 28.4 Attractive van der Waals force F between two curved mica surfaces of radius $R \approx 1$ cm measured in water and various aqueous electrolyte solutions. The electrostatic interaction has been subtracted from the total measured force. The measured nonretarded Hamaker constant is $A_H = 2.2 \times 10^{-20}$ J. Retardation effects are apparent at distances larger than 5 nm, as expected theoretically (after [28.1], with permission from Elsevier Science)

portant at distances below 10−15 nm and may at these distances start to dominate interactions of different origin that have been observed at larger separations.

Figure 28.4 shows the measured van der Waals forces between two crossed-cylindrical mica surfaces in water and various salt solutions. Good agreement is obtained between experiment and theory. At larger surface separations, above about 5 nm, the measured forces fall off more rapidly than D^{-2}. This retardation effect (Sect. 28.3.1) is also predicted by the Lifshitz theory and is due to the time needed for propagation of the induced dipole moments over large distances.

From Fig. 28.4 we may conclude that, at separations above about 2 nm, or 8 molecular diameters of water, the *continuum* Lifshitz theory is valid. This would mean that water films as thin as 2 nm may be expected to have bulk-like properties, at least as far as their interaction forces are concerned. Similar results have been obtained with other liquids, where in general continuum properties are manifested, both as regards to their interactions and other properties such as viscosity, at a film thickness larger than five or ten molecular diameters. In the absence of a solvent (in vacuum), the agreement of measured van der Waals forces with the continuum Lifshitz theory is generally good at all separations down to molecular contact ($D = D_0$).

Van der Waals interactions in a system of three or more different materials (28.8) can be attractive or repulsive, depending on their dielectric properties. Numerous experimental studies show the attractive van der Waals forces in various systems [28.1], and repulsive van der Waals forces have also been measured directly [28.43]. A practical consequence of the repulsive interaction obtained across a medium with intermediate dielectric properties is that the van der Waals forces will give rise to preferential, nonspecific adsorption of molecules with an intermediate dielectric constant. It is possible to diminish the attractive interaction between dispersed colloidal particles by adsorption of a thin layer of material with dielectric properties close to those of the surrounding medium (matching of refractive index), or by adsorption of a polymer that gives a steric repulsive force that keeps the particles separated at a distance where the magnitude of the van der Waals attraction is negligible. Thermal motion will then keep the particles dispersed.

28.4.2 Electrostatic Forces

Most surfaces in contact with a highly polar liquid (such as water) acquire a surface charge, either by dissociation of ions from the surface into the solution or by preferential adsorption of certain ions from the solution. The surface charge is balanced by a layer of oppositely charged ions (counterions) in the solution at some small distance from the surface [28.1]. In dilute solution, this distance is the Debye length κ^{-1}, which is purely a property of the electrolyte solution. The Debye length falls with increasing ionic strength (i. e., with the molar concentration M_i and valency z_i) of the ions in solution

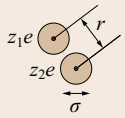
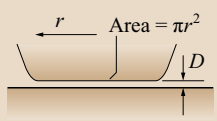
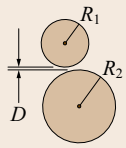
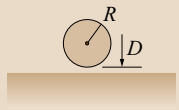
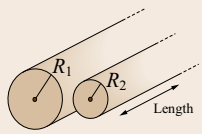
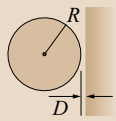
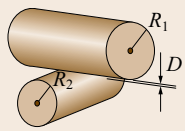
$$\kappa^{-1} = \left(\frac{\varepsilon \varepsilon_0 k_B T}{e^2 N_A \sum_i z_i^2 M_i} \right)^{1/2} , \qquad (28.10)$$

where e is the electronic charge. For example, for 1 : 1 electrolytes at 25 °C, $\kappa^{-1} = 0.304\,\text{nm}/\sqrt{M_{1:1}}$, where M_i is given in M (mol dm^{-3}). κ^{-1} is thus about 10 nm in a 1 mM NaCl solution and 0.3 nm in a 1 M solution. In totally pure water at pH 7, where $M_i = 10^{-7}$ M, κ^{-1} is 960 nm, or about 1 μm. The Debye length also relates the surface charge density σ of a surface to the electrostatic surface potential ψ_0 via the Grahame equation, which for 1 : 1 electrolytes can be expressed as

$$\sigma = \sqrt{8 \varepsilon \varepsilon_0 k_B T} \sinh \left(\frac{e \psi_0}{2 k_B T} \right) \sqrt{M_{1:1} N_A} . \qquad (28.11)$$

Since the Debye length is a measure of the thickness of the diffuse atmosphere of counterions near a charged surface, it also determines the range of the electrostatic double-layer interaction between two charged surfaces.

Table 28.3 Electrical double-layer interaction energy $E(D)$ and force ($F = -\mathrm{d}E/\mathrm{d}D$) between macroscopic bodies

Geometry of bodies with surfaces D apart ($D \ll R$)			Electric double-layer interaction	
			Energy E	Force F
Two ions or small molecules		$r \geq \sigma$	$\frac{\pm z_1 z_2 e^2}{4\pi\varepsilon\varepsilon_0 r} \frac{\mathrm{e}^{-\kappa(r-\sigma)}}{1+\kappa\sigma}$	$\frac{\pm z_1 z_2 e^2}{4\pi\varepsilon\varepsilon_0 r^2} \frac{(1+\kappa r)}{1+\kappa\sigma} \mathrm{e}^{-\kappa(r-\sigma)}$
Two flat surfaces (per unit area)		$r \gg D$	$\left(\frac{\kappa}{2\pi}\right) Z\mathrm{e}^{-\kappa D}$	$\left(\frac{\kappa^2}{2\pi}\right) Z\mathrm{e}^{-\kappa D}$
Two spheres or macromolecules of radii R_1 and R_2		$R_1, R_2 \gg D$	$\left(\frac{R_1 R_2}{R_1+R_2}\right) Z\mathrm{e}^{-\kappa D}$	$\kappa \left(\frac{R_1 R_2}{R_1+R_2}\right) Z\mathrm{e}^{-\kappa D}$
Sphere or macromolecule of radius R near a flat surface		$R \gg D$	$RZ\mathrm{e}^{-\kappa D}$	$\kappa RZ\mathrm{e}^{-\kappa D}$
Two parallel cylinders or rods of radii R_1 and R_2 (per unit length)		$R_1, R_2 \gg D$	$\frac{\kappa^{1/2}}{\sqrt{2\pi}} \left(\frac{R_1 R_2}{R_1+R_2}\right)^{1/2} Z\mathrm{e}^{-\kappa D}$	$\frac{\kappa^{3/2}}{\sqrt{2\pi}} \left(\frac{R_1 R_2}{R_1+R_2}\right)^{1/2} Z\mathrm{e}^{-\kappa D}$
Cylinder of radius R near a flat surface (per unit length)		$R \gg D$	$\kappa^{1/2} \sqrt{\frac{R}{2\pi}} Z\mathrm{e}^{-\kappa D}$	$\kappa^{3/2} \sqrt{\frac{R}{2\pi}} Z\mathrm{e}^{-\kappa D}$
Two cylinders or filaments of radii R_1 and R_2 crossed at 90°		$R_1, R_2 \gg D$	$\sqrt{R_1 R_2} Z\mathrm{e}^{-\kappa D}$	$\kappa \sqrt{R_1 R_2} Z\mathrm{e}^{-\kappa D}$

The interaction energy and force for bodies of different geometries is based on the Poisson–Boltzmann equation (a continuum, mean-field theory). Equation (28.14) gives the interaction constant Z (in terms of the surface potential ψ_0) for the interaction between similarly charged (ionized) surfaces in aqueous solutions of monovalent electrolyte. It can also be expressed in terms of the surface charge density σ by applying the Grahame equation (28.12) (after [28.9], with permission)

The electrostatic double-layer interaction is an entropic effect that arises upon decreasing the thickness of the liquid film containing the dissolved ions. Because of the attractive force between the dissolved ions and opposite charges on the surfaces, the ions stay between the surfaces, but an osmotic repulsion arises as their concentration increases. The long-range electrostatic interaction energy at large separations (weak overlap) between two similarly charged molecules or surfaces is typically repulsive and is roughly an exponentially decaying function of D

$$E(D) \approx +C_{\mathrm{ES}} \mathrm{e}^{-\kappa D} , \qquad (28.12)$$

where C_{ES} is a constant that depends on the geometry of the interacting surfaces, on their surface charge density, and on the solution conditions (Table 28.3). We see that the Debye length is the decay length of the interaction

energy between two surfaces (and of the mean potential away from one surface). C_{ES} can be determined by solving the so-called Poisson–Boltzmann equation or by using other theories [28.44, 45]. The equations in Table 28.3 are expressed in terms of a constant Z defined as

$$Z = 64\pi\varepsilon\varepsilon_0 \left(\frac{k_B T}{e}\right)^2 \tanh^2\left(\frac{ze\psi_0}{4k_B T}\right), \qquad (28.13)$$

which depends only on the properties of the surfaces.

The above approximate expressions are accurate only for surface separations larger than about one Debye length. At smaller separations one must use numerical solutions of the Poisson–Boltzmann equation to obtain the exact interaction potential, for which there are no simple expressions. In the limit of small D, it can be shown that the interaction energy depends on whether the surfaces remain at constant potential ψ_0 (as assumed in the above equations) or at constant charge σ (when the repulsion exceeds that predicted by the above equations), or somewhere between these limits. In the constant charge limit the total number of counterions in the compressed film does not change as D is decreased, whereas at constant potential the concentration of counterions is constant. The limiting pressure (or force per unit area) at constant charge is the osmotic pressure of the confined ions

$$F = k_B T \times \text{ion number density} = \frac{2\sigma k_B T}{zeD}, \text{ for } D \ll \kappa^{-1}. \qquad (28.14)$$

That is, as $D \to 0$ the double-layer pressure at constant surface charge becomes infinitely repulsive and independent of the salt concentration (at constant potential the force instead becomes a constant at small D). However, at small separations, the van der Waals attraction (which goes as D^{-2} between two spheres or as D^{-3} between two planar surfaces, see Table 28.2) wins out over the double-layer repulsion, unless some other short-range interaction becomes dominant (Sect. 28.4.4). This is the theoretical prediction that forms the basis of the so-called Derjaguin–Landau–Verwey–Overbeek (DLVO) theory [28.44, 46], illustrated in Fig. 28.5.

Because of the different distance dependences of the van der Waals and electrostatic interactions, the total force law, as described by the DLVO theory, can show several minima and maxima. Typically, the depth of the outer (secondary) minimum is a few $k_B T$, which is enough to cause reversible flocculation of particles from an aqueous dispersion. If the force barrier between the secondary and primary minimum is lowered, for example by increasing the electrolyte concentration, particles

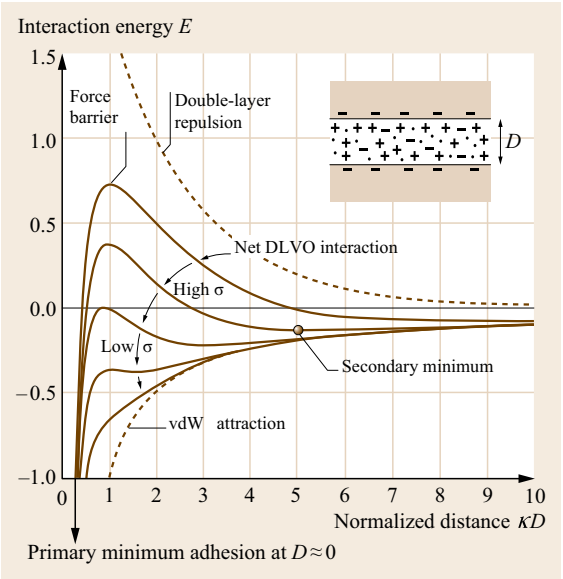


Fig. 28.5 Schematic plots of the DLVO interaction potential energy E between two flat, charged surfaces (or, according to the Derjaguin approximation, (28.3), the force F between two curved surfaces) as a function of the surface separation normalized by the Debye length κ^{-1}. The van der Waals attraction (inverse power-law dependence on D) together with the repulsive electrostatic double-layer force (roughly exponential) at different surface charge σ (or potential, see (28.11)) determine the net interaction potential in aqueous electrolyte solution (after [28.47] with permission)

can be irreversibly coagulated in the primary minimum. In practice, other forces (described in the following sections) often appear at very small separations, so that the full force law between two surfaces or colloidal particles in solution can be more complex than might be expected from the DLVO theory.

28.4.3 Solvation and Structural Forces

When a liquid is confined within a restricted space, for example a very thin film between two surfaces, it ceases to behave as a structureless continuum. At small surface separations (below about ten molecular diameters), the van der Waals force between two surfaces or even two solute molecules in a liquid (solvent) is no longer a smoothly varying attraction. Instead, there arises an additional solvation force that generally oscillates between attraction and repulsion with distance, with a periodicity equal to some mean dimension σ of the liquid molecules [28.51]. Figure 28.6a shows the force law between two smooth mica surfaces across the hydrocarbon liquid tetradecane, whose inert, chainlike

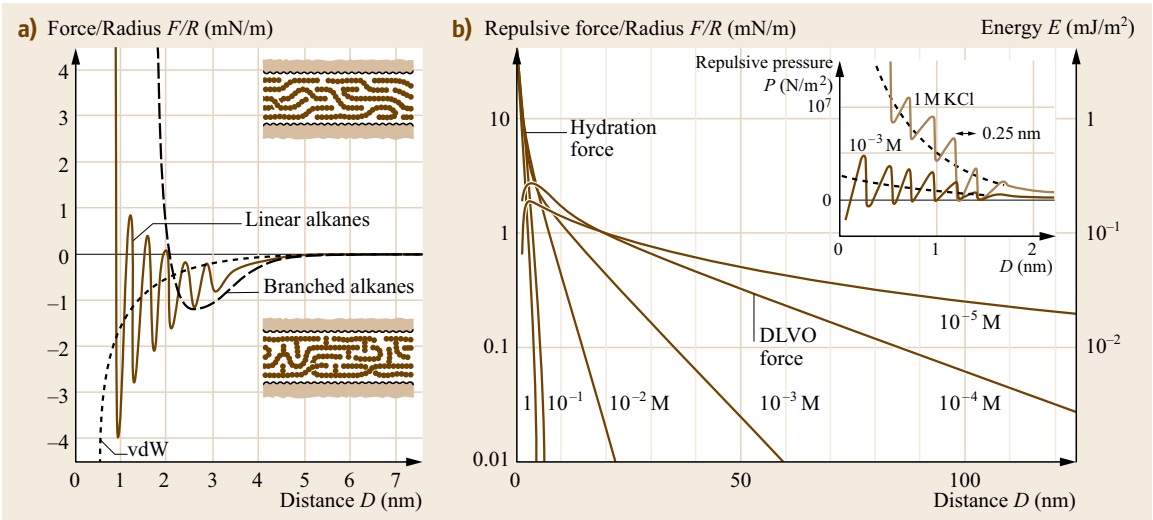


Fig. 28.6 **(a)** *Solid curve*: Forces measured between two mica surfaces across saturated linear chain alkanes such as *n*-tetradecane and *n*-hexadecane (after [28.48, 49]). The 0.4 nm periodicity of the oscillations indicates that the molecules are preferentially oriented parallel to the surfaces, as shown schematically in the *upper insert*. The theoretical continuum van der Waals attraction is shown as a *dotted curve*. *Dashed curve*: Smooth, nonoscillatory force law exhibited by irregularly shaped alkanes (such as 2-methyloctadecane) that cannot order into well-defined layers (*lower insert*) (after [28.48, 50]). Similar nonoscillatory forces are also observed between rough surfaces, even when these interact across a saturated linear chain liquid. This is because the irregularly shaped surfaces (rather than the liquid) now prevent the liquid molecules from ordering in the gap. **(b)** Forces measured between charged mica surfaces in KCl solutions of varying concentrations (after [28.14]). In dilute solutions (10^{-5} and 10^{-4} M), the measured forces are excellently described by the DLVO theory, based on exact solutions to the nonlinear Poisson–Boltzmann equation for the electrostatic forces and the Lifshitz theory for the van der Waals forces (using a Hamaker constant of $A_H = 2.2 \times 10^{-20}$ J). At higher concentrations, as more hydrated K^+ cations adsorb onto the negatively charged surfaces, an additional hydration force appears superimposed on the DLVO interaction at distances below 3–4 nm. This force has both an oscillatory and a monotonic component. *Insert*: Short-range hydration forces between mica surfaces shown as pressure versus distance. The *lower* and *upper curves* show surfaces 40% and 95% saturated with K^+ ions. At larger separations, the forces are in good agreement with the DLVO theory (after [28.1], with permission from Elsevier Science)

molecules have a width of $\sigma \approx 0.4$ nm. The short-range oscillatory force law is related to the density distribution function and potential of mean force characteristic of intermolecular interactions in liquids. These forces arise from the confining effects that the two surfaces have on liquid molecules, forcing them to order into quasidiscrete layers. Such layers are energetically or entropically favored and correspond to the minima in the free energy, whereas fractional layers are disfavored (energy maxima). This effect is quite general and arises in all simple liquids when they are confined between two smooth, rigid surfaces, both flat and curved.

Oscillatory forces are now well understood theoretically, at least for simple liquids, and a number of theoretical studies and computer simulations of various confined liquids (including water) that interact via some form of Lennard-Jones potential have invariably led to an oscillatory solvation force at surface separations below a few molecular diameters [28.52–55]. In a first approximation, the oscillatory force law may be described by an exponentially decaying cosine function of the form

$$E \approx E_0 \cos\left(\frac{2\pi D}{\sigma}\right) \mathrm{e}^{-D/\sigma} , \qquad (28.15)$$

where both theory and experiments show that the oscillatory period and the characteristic decay length of the envelope are close to σ.

Once the solvation zones of the two surfaces overlap, the mean liquid density in the gap is no longer the same as in the bulk liquid. Since the van der Waals interaction depends on the optical properties of the liquid, which in turn depend on the density, the van der Waals and the oscillatory solvation forces are not strictly additive. It is more correct to think of the solvation force as the van der Waals force at small separations with the molecular properties and density variations of the

medium taken into account. It is also important to appreciate that solvation forces do not arise simply because liquid molecules tend to structure into semi-ordered layers at surfaces. They arise because of the disruption or *change* of this ordering during the approach of a second surface. The two effects are related: the greater the tendency toward structuring at an isolated surface the greater the solvation force between two such surfaces; but there is a real distinction between the two phenomena that should be borne in mind.

Oscillatory forces lead to different adhesion values depending on the energy minimum from which two surfaces are being separated. For an interaction energy described by (28.15), quantized adhesion energies will be E_0 at $D = 0$ (primary minimum), E_0/e at $D = \sigma$, E_0/e^2 at $D = 2\sigma$, etc. E_0 can be thought of as a depletion force (Sect. 28.4.5) that is approximately given by the osmotic limit $E_0 \approx -k_B T/\sigma^2$, which can exceed the contribution to the adhesion energy in contact from the van der Waals forces (at $D_0 \approx 0.15{-}0.20\,\text{nm}$, as discussed in Sect. 28.3.1, keeping in mind that the Lifshitz theory fails to describe the force law at *intermediate distances*).

It is easy to understand how oscillatory forces arise between two flat, plane parallel surfaces. Between two curved surfaces, e.g., two spheres, one might imagine the molecular ordering and oscillatory forces to be smeared out in the same way that they are smeared out between two randomly rough surfaces (Sect. 28.5.3); however, this is not the case. Ordering can occur as long as the curvature or roughness is itself regular or uniform, i. e., not random. This is due to the Derjaguin approximation (28.3). If the energy between two flat surfaces is given by a decaying oscillatory function (for example, a cosine function as in (28.15)), then the force (and energy) between two curved surfaces will also be an oscillatory function of distance with some phase shift. Likewise, two surfaces with regularly curved regions will also retain their oscillatory force profile, albeit modified, as long as the corrugations are truly regular, i. e., periodic. On the other hand, surface roughness, even on the nanometer scale, can smear out oscillations if the roughness is random and the confined molecules are smaller than the size of the surface asperities [28.56, 57]. If an organic liquid contains small amounts of water, the expected oscillatory force can be replaced by a strongly attractive capillary force (Sect. 28.5.1).

28.4.4 Hydration and Hydrophobic Forces

There are many aqueous systems in which the DLVO theory fails and where there is an additional short-range force that is not oscillatory but monotonic. Between hydrophilic surfaces this force is exponentially repulsive and is commonly referred to as the hydration, or structural, force. The origin and nature of this force has long been controversial, especially in the colloidal and biological literature. Repulsive hydration forces are believed to arise from strongly hydrogen-bonding surface groups, such as hydrated ions or hydroxyl (−OH) groups, which modify the hydrogen-bonding network of liquid water adjacent to them. Because this network is quite extensive in range [28.58], the resulting interaction force is also of relatively long range.

Repulsive hydration forces were first extensively studied between clay surfaces [28.59]. More recently, they have been measured in detail between mica and silica surfaces [28.14–16], and even biological membranes [28.60, 61], where they have been found to decay exponentially with decay lengths of about 1 nm. Their effective range is 3−5 nm, which is about twice the range of the oscillatory solvation force in water. Empirically, the hydration repulsion between two hydrophilic surfaces appears to follow the simple equation

$$E = E_0 \mathrm{e}^{-D/\lambda_0}\,, \tag{28.16}$$

where $\lambda_0 \approx 0.6{-}1.1\,\text{nm}$ for 1 : 1 electrolytes and $E_0 = 3{-}30\,\text{mJ m}^{-2}$ depending on the hydration (hydrophilicity) of the surfaces, higher E_0 values generally being associated with lower λ_0 values.

The interactions between molecularly smooth mica surfaces in dilute electrolyte solutions obey the DLVO theory (Fig. 28.6b). However, at higher salt concentrations, specific to each electrolyte, hydrated cations bind to the negatively charged surfaces and give rise to a repulsive hydration force [28.14, 15]. This is believed to be due to the energy needed to dehydrate the bound cations, which presumably retain some of their water of hydration on binding. This conclusion was arrived at after noting that the strength and range of the hydration forces increase with the known hydration numbers of the electrolyte cations in the order: $Mg^{2+} > Ca^{2+} > Li^+ \approx Na^+ > K^+ > Cs^+$. Similar trends are observed with other negatively charged colloidal surfaces.

While the hydration force between two mica surfaces is overall repulsive below a distance of 4 nm, it is not always monotonic below about 1.5 nm but exhibits oscillations of mean periodicity of $0.25 \pm 0.03\,\text{nm}$, roughly equal to the diameter of the water molecule. This is shown in the insert in Fig. 28.6b, where we may note that the first three minima at $D = 0$, 0.28, and 0.56 nm occur at negative energies, a result that rationalizes observations on certain colloidal systems.

Surface force experiments have shown that hydration forces can be modified or *regulated* by exchanging ions of different hydration on surfaces, an effect that has

important practical applications in controlling the stability of colloidal dispersions. It has long been known that colloidal particles can be precipitated (coagulated or flocculated) by increasing the electrolyte concentration, an effect that was traditionally attributed to the reduced screening of the electrostatic double-layer repulsion between the particles due to the reduced Debye length. However, there are many examples where colloids are stabilized at high salt concentrations, not at low concentrations. This effect is now recognized as being due to the increased hydration repulsion experienced by certain surfaces when they bind highly hydrated ions at higher salt concentrations. Hydration regulation of adhesion and interparticle forces is an important practical method for controlling various processes such as clay swelling [28.62], ceramic processing and rheology [28.63], material fracture [28.63], and colloidal particle and bubble coalescence [28.64].

Water appears to be unique in having a solvation (hydration) force that exhibits both a monotonic and an oscillatory component. Between hydrophilic surfaces the monotonic component is repulsive (Fig. 28.6b), but between hydrophobic surfaces it is attractive and the final adhesion is much greater than expected from the Lifshitz theory.

A hydrophobic surface is one that is inert to water in the sense that it cannot bind to water molecules via ionic or hydrogen bonds. Hydrocarbons and fluorocarbons are hydrophobic, as is air, and the strongly attractive hydrophobic force has many important manifestations and consequences, such as the low solubility or miscibility of water and oil molecules, micellization, protein folding, strong adhesion and rapid coagulation of hydrophobic surfaces, nonwetting of water on hydrophobic surfaces, and hydrophobic particle attachment to rising air bubbles (the basic principle of froth flotation).

In recent years, there has been a steady accumulation of experimental data on the force laws between various hydrophobic surfaces in aqueous solution [28.65–76]. These studies have found that the force law between two macroscopic hydrophobic surfaces is of surprisingly long range, decaying exponentially with a characteristic decay length of 1–2 nm in the separation range of 0–10 nm, and then more gradually further out. The hydrophobic force can be far stronger than the van der Waals attraction, especially between hydrocarbon surfaces in water, for which the Hamaker constant is quite small.

For two surfaces in water the purely hydrophobic interaction energy (ignoring DLVO and oscillatory forces) in the range 0–10 nm is given by

$$E = -2\gamma e^{-D/\lambda_0} , \qquad (28.17)$$

where typically $\lambda_0 = 1–2$ nm, and $\gamma = 10–50\,\mathrm{mJ\,m^{-2}}$. The higher value corresponds to the interfacial energy of a pure hydrocarbon–water interface.

At a separation below 10 nm, the hydrophobic force appears to be insensitive or only weakly sensitive to changes in the type and concentration of electrolyte ions in the solution. The absence of a screening effect by ions attests to the nonelectrostatic origin of this interaction. In contrast, some experiments have shown that, at separations greater than 10 nm, the attraction does depend on the intervening electrolyte, and that in dilute solutions, or solutions containing divalent ions, it can continue to exceed the van der Waals attraction out to separations of 80 nm [28.68, 77]. Recent research suggests that the interactions at very long range might not represent a hydrophobic force since they are influenced by the presence of dissolved gas in the solution [28.73, 74], the stability of the hydrophobic surface [28.75], and on some types of surfaces, bridging submicroscopic bubbles [28.78, 79].

The long-range nature of the hydrophobic interaction has a number of important consequences. It accounts for the rapid coagulation of hydrophobic particles in water and may also account for the rapid folding of proteins. It also explains the ease with which water films rupture on hydrophobic surfaces. In this case, the van der Waals force across the water film is repulsive and therefore favors wetting, but this is more than offset by the attractive hydrophobic interaction acting between the two hydrophobic phases across water. Hydrophobic forces are increasingly being implicated in the adhesion and fusion of biological membranes and cells. It is known that both osmotic and electric-field stresses enhance membrane fusion, an effect that may be due to the concomitant increase in the hydrophobic area exposed between two adjacent surfaces [28.80].

From the previous discussion we can infer that even though highly related to each other [28.76], hydration and hydrophobic forces are not of a simple nature. These interactions are probably the most important, yet the least understood, of all the forces in aqueous solutions. The unusual properties of water and the nature of the surfaces (including their homogeneity and stability) appear to be equally important [28.80].

How do these exponentially decaying repulsive or attractive forces arise? Theoretical work and computer simulations [28.53, 54] suggest that the solvation forces in water should be purely oscillatory, whereas other theoretical studies [28.81–83] suggest a monotonically exponential repulsion or attraction, possibly superimposed on an oscillatory force. The latter is consistent with experimental findings, as shown in the *inset* in Fig. 28.6b, where it appears that the oscillatory force is simply additive with the monotonic hydration and

DLVO forces, suggesting that these arise from essentially different mechanisms.

Like the repulsive hydration force, the origin of the hydrophobic force is still unknown. A Monte Carlo simulation [28.83] of the interaction between two hydrophobic surfaces across water at separations below 1.5 nm showed a decaying oscillatory force superimposed on a monotonically attractive curve. In more recent computational and experimental work [28.84–86], it has been suggested that hydrophobic surfaces generate a depleted region of water around them, and that a long-range attractive force due to depletion arises between two such surfaces. Such a difference in density might also cause boundary slip of water at hydrophobic surfaces [28.87].

It is questionable whether the hydration or hydrophobic force should be viewed as an ordinary type of solvation or structural force that reflects the packing of water molecules. The energy (or entropy) associated with the hydrogen-bonding network, which extends over a much larger region of space than the molecular correlations, is probably at the root of the long-range interactions of water. The situation in water appears to be governed by much more than the molecular packing effects that dominate the interactions in simpler liquids.

28.4.5 Polymer-Mediated Forces

Polymers or macromolecules are chainlike molecules consisting of many identical segments (monomers or repeating units) held together by covalent bonds. The size of a polymer coil in solution or in the melt is determined by a balance between van der Waals attraction (and hydrogen bonding, if present) between polymer segments, and the entropy of mixing, which causes the polymer coil to expand. In polymer melts above the glass transition temperature, and at certain conditions in solution, the attraction between polymer segments is exactly balanced by the entropy effect. The polymer solution will then behave virtually ideally, and the density distribution of segments in the coil is Gaussian. This is called the theta (θ) condition, and it occurs at the theta or Flory temperature for a particular combination of polymer and solvent or solvent mixture. At lower temperatures (in a poor or bad solvent), the polymer-polymer interactions dominate over the entropic, and the coil will shrink or precipitate. At higher temperatures (good solvent conditions), the polymer coil will expand.

High-molecular-weight polymers form large coils, which significantly affect the properties of a solution even when the total mass of polymer is very low. The radius of the polymer coil is proportional to the segment length a and the number of segments n. At theta conditions, the hydrodynamic radius of the polymer coil (the root-mean-square separation of the ends of one polymer chain) is theoretically given by $R_h = an^{1/2}$, and the unperturbed radius of gyration (the average root-mean-square distance of a segment from the center of mass of the molecule) is $R_g = a(n/6)^{1/2}$. In a good solvent the perturbed size of the polymer coil, the Flory radius R_F, is proportional to $n^{3/5}$.

Polymers interact with surfaces mainly through van der Waals and electrostatic interactions. The physisorption of polymers containing only one type of segment is reversible and highly dynamic, but the rate of exchange of adsorbed chains with free chains in the solution is low, since the polymer remains bound to the surface as long as one segment along the chain is adsorbed. The adsorption energy per segment is on the order of k_BT. In a good solvent, the conformation of a polymer on a surface is very different from the coil conformation in bulk solution. Polymers adsorb in *trains*, separated by *loops* extending into the solution and dangling *tails* (the ends of the chain). Compared to adsorption at lower temperatures, good solvent conditions favor more of the polymer chain being in the solvent, where it can attain its optimum conformation. As a result, the extension of the polymer is longer, even though the total amount of adsorbed polymer is lower. In a good solvent, the polymer chains can also be effectively repelled from a surface, if the loss in conformational entropy close to the surface is not compensated for by a gain in enthalpy from adsorption of segments. In this case, there will be a layer of solution (thickness $\approx R_g$) close to the surfaces that is depleted of polymer.

The interaction forces between two surfaces across a polymer solution will depend on whether the polymer adsorbs onto the surfaces or is repelled from them, and also on whether the interaction occurs at *true* or *restricted* thermodynamic equilibrium. At true or full equilibrium, the polymer between the surfaces can equilibrate (exchange) with polymer in the bulk solution at all surface separations. Some theories [28.92, 93] predict that, at full equilibrium, the polymer chains would move from the confined gap into the bulk solution where they could attain entropically more favorable conformations, and that a monotonic attraction at all distances would result from bridging and depletion interactions (which will be discussed below). Other theories suggest that the interaction at small separations would be ultimately repulsive, since some polymer chains would remain in the gap due to their attractive interactions with many sites on the surface (enthalpic) – more sites would be available to the remaining polymer chains if some others desorbed and diffused out from the gap [28.94].

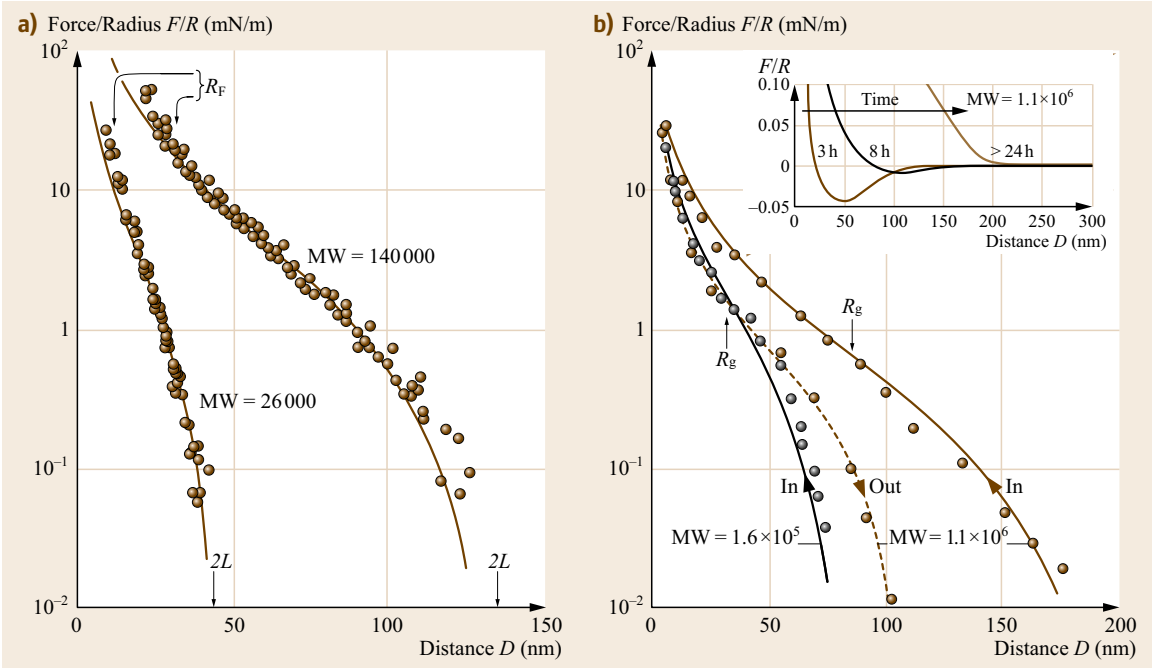


Fig. 28.7a,b Experimentally determined forces in systems of two interacting polymer layers: **(a)** Polystyrene brush layers grafted via an adsorbing chain-end group onto mica surfaces in toluene (a good solvent for polystyrene). *Left curve*: MW = 26 000 g/mol, R_F = 12 nm. *Right curve*: MW = 140 000 g/mol, R_F = 32 nm. Both force curves were reversible on approach and separation. The *solid curves* are theoretical fits using the Alexander–de Gennes theory with the following measured parameters: spacing between attachments sites s = 8.5 nm, brush thickness L = 22.5 and 65 nm (adapted from [28.88]). **(b)** Polyethylene oxide layers physisorbed onto mica from 150 µg/ml solution in aqueous 0.1 M KNO_3 (a good solvent for polyethylene oxide). *Main figure*: Equilibrium forces at full coverage after ≈ 16 h adsorption time. *Left curve*: MW = 160 000 g/mol, R_g = 32 nm. *Right curve*: MW = 1 100 000 g/mol, R_g = 86 nm. Note the hysteresis (irreversibility) on approach and separation for this physisorbed layer, in contrast to the absence of hysteresis with grafted chains in the case of **(a)**. The *solid curves* are based on a modified form of the Alexander–de Gennes theory. *Insert* in **(b)**: evolution of the forces with the time allowed for the higher MW polymer to adsorb from solution. Note the gradual reduction in the attractive bridging component (adapted from [28.89–91], after [28.1], with permission from Elsevier Science)

At restricted equilibrium, the polymer is kinetically trapped, and the adsorbed amount is thus constant as the surfaces are brought toward each other, but the chains can still rearrange on the surfaces and in the gap. Experimentally, the true equilibrium situation is very difficult to attain, and most experiments are done at restricted equilibrium conditions. Even the equilibration of conformations assumed in theoretical models for restricted equilibrium conditions can be so slow that this condition is difficult to reach experimentally.

In systems of adsorbing polymers, bridging of chains from one surface to the other can give rise to a long-range attraction, since the bridging chains would gain conformational entropy if the surfaces were closer together. In poor solvents, both bridging and intersegment interactions contribute to an attraction [28.20]. However, regardless of solvent and equilibrium conditions, a strong repulsion due to the osmotic interactions is seen at small surface separations in systems of adsorbing polymers at restricted equilibrium.

In systems containing high concentrations of nonadsorbing polymer, the difference in solute concentration in the bulk and between the surfaces at separations smaller than the approximate polymer coil diameter ($2R_g$, i. e., when the polymer has been squeezed out from the gap between the surfaces) may give rise to an attractive osmotic force (the depletion attraction) [28.95–98]. In addition, if the polymer coils become initially compressed as the surfaces approach each other, this can give rise to a repulsion (depletion stabilization) at large separations [28.97]. For a system of two cylindrical surfaces of radius R, the maximum depletion force F_{dep} is expected to occur when the surfaces are in contact and is given by multiplying the

depletion (osmotic) pressure, $P_{dep} = \rho k_B T$, by the contact area πr^2, where r is given by the chord theorem $r^2 = (2R - R_g)R_g \approx 2RR_g$ [28.1]

$$\frac{F_{dep}}{R} = -2\pi R_g \rho k_B T \,, \tag{28.18}$$

where ρ is the number density of the polymer in the bulk solution.

If a part of the polymer (typically an end group) is different from the rest of the chain, this part may preferentially adsorb to the surface. An end-adsorbed polymer is attached to the surface at only one point, and the extension of the chain is dependent on the grafting density, i. e., the average distance s between adsorbed end groups on the surface (Fig. 28.7). One distinguishes between different regions of increasing overlap of the chains (stretching) called pancake, mushroom, and brush regimes [28.99]. In the mushroom regime, where the coverage is sufficiently low so that there is no overlap between neighboring chains, the thickness of the adsorbed layer is proportional to $n^{1/2}$ (i. e., to R_g) at theta conditions and to $n^{3/5}$ in a good solvent.

Several models [28.99–103] have been developed for the extension and interactions between two brushes (strongly stretched grafted chains). They are based on a balance between osmotic pressure within the brush layers (uncompressed and compressed) and the elastic energy of the chains, and differ mainly in the assumptions of the segment density profile, which can be a step function or parabolic. At high coverage (in the brush regime), where the chains will avoid overlapping each other, the thickness of the layer is proportional to n.

Experimental work on both monodisperse [28.21, 104] and polydisperse [28.22] systems at different solvent conditions has confirmed the expected range and magnitude of the repulsive interactions resulting from compression of densely packed grafted layers.

28.5 Adhesion and Capillary Forces

28.5.1 Capillary Forces

When considering the adhesion of two solid surfaces or particles in air or in a liquid, it is easy to overlook or underestimate the important role of capillary forces, i. e., forces arising from the Laplace pressure of curved menisci formed by condensation of a liquid between and around two adhering surfaces (Fig. 28.8).

The adhesion force between a nondeformable spherical particle of radius R and a flat surface in an inert atmosphere (Fig. 28.8a) is

$$F_{ad} = 4\pi R \gamma_{SV} \,. \tag{28.19}$$

But in an atmosphere containing a condensible vapor, the expression above is replaced by

$$F_{ad} = 4\pi R(\gamma_{LV} \cos\theta + \gamma_{SL}) \,, \tag{28.20}$$

where the first term is due to the Laplace pressure of the meniscus and the second is due to the direct adhesion of the two contacting solids within the liquid. Note that the above equation does not contain the radius of curvature r of the liquid meniscus (Fig. 28.8b). This is because for smaller r the Laplace pressure γ_{LV}/r increases, but the area over which it acts decreases by the same amount, so the two effects cancel out. In experiments with inert liquids, such as hydrocarbons, condensing between two mica surfaces indicates that (28.20) is valid for values of r as small as 1–2 nm, corresponding to vapor pressures as low as 40% of saturation [28.105, 106]. Capillary condensation also occurs in binary liquid systems, e.g., when water dissolved in hydrocarbon liquids condenses around two contacting hydrophilic surfaces or when a vapor cavity forms in water around two hydrophobic surfaces. In the case of water condensing from vapor or from oil, it also appears that the bulk value of γ_{LV} is applicable for meniscus radii as small as 2 nm.

The capillary condensation of liquids, especially water, from vapor can have additional effects on the physical state of the contact zone. For example, if the surfaces contain ions, these will diffuse and build up within the liquid bridge, thereby changing the chemical composition of the contact zone, as well as influencing the adhesion. In the case of surfaces covered with surfactant or polymer molecules (amphiphilic surfaces), the molecules can turn over on exposure to humid air, so that the surface nonpolar groups become replaced by polar groups, which renders the surfaces hydrophilic. When two such surfaces come into contact, water will condense around the contact zone and the adhesion force will also be affected – generally increasing well above the value expected for inert hydrophobic surfaces. It is apparent that adhesion in vapor or a solvent is often largely determined by capillary forces arising from the condensation of liquid that may be present only in very small quantities, e.g., 10–20% of saturation in the vapor, or 20 ppm in the solvent.

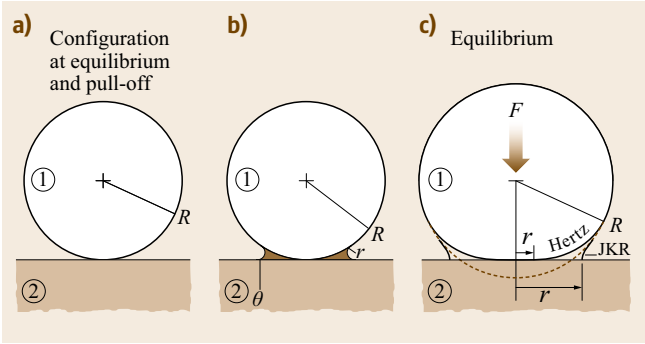


Fig. 28.8a–c Adhesion and capillary forces between materials 1 and 2 : **(a)** a nondeforming sphere on a rigid, flat surface in an inert atmosphere and **(b)** in a vapor that can capillary-condense around the contact zone. At equilibrium, the concave radius r of the liquid meniscus is given by the Kelvin equation. For a concave meniscus to form, the contact angle θ has to be less than 90°. In the case of hydrophobic surfaces surrounded by water, a vapor cavity can form between the surfaces. As long as the surfaces are perfectly smooth, the contribution of the meniscus to the adhesion force is independent of r (after [28.5] with permission). **(c)** Elastically deformable sphere on a rigid flat surface in the absence (Hertz) and presence (JKR) of adhesion (**(a,c)** after [28.1], with permission from Elsevier Science)

28.5.2 Adhesion Mechanics

Two bodies in contact deform as a result of surface forces and/or applied normal forces. For the simplest case of two interacting elastic spheres (a model that is easily extended to an elastic sphere interacting with an undeformable surface, or vice versa) and in the absence of attractive surface forces, the vertical central displacement (compression) was derived by *Hertz* [28.107] (Fig. 28.8c). In this model, the displacement and the contact area are equal to zero when no external force (load) is applied, i. e., at the points of contact and of separation. The contact area A increases with normal force or load as $L^{2/3}$.

In systems where attractive surface forces are present between the surfaces, the deformations are more complicated. Modern theories of the adhesion mechanics of two contacting solid surfaces are based on the Johnson–Kendall–Roberts (JKR) theory [28.10], or on the Derjaguin–Muller–Toporov (DMT) theory [28.108–110]. The JKR theory is applicable to easily deformable, large bodies with high surface energy, whereas the DMT theory better describes very small and hard bodies with low surface energy [28.111]. The intermediate regime has also been described [28.112].

In the JKR theory, two spheres of radii R_1 and R_2, bulk elastic modulus K, and surface energy γ will flatten due to attractive surface forces when in contact at no external load. The contact area will increase under an external load L or normal force F, such that at mechanical equilibrium the radius of the contact area r is given by

$$r^3 = \frac{R}{K}\left[F + 6\pi R\gamma + \sqrt{12\pi R\gamma F + (6\pi R\gamma)^2}\right], \tag{28.21}$$

where $R = R_1R_2/(R_1 + R_2)$. In the absence of surface energy γ, (28.21) is reduced to the expression for the radius of the contact area in the Hertz model. Another important result of the JKR theory gives the adhesion force or pull-off force

$$F_{\text{ad}} = -3\pi R\gamma_{\text{S}}\,, \tag{28.22}$$

where the surface energy γ_{S} is defined through $W = 2\gamma_{\text{S}}$, where W is the reversible work of adhesion. Note that, according to the JKR theory, a finite elastic modulus K, while having an effect on the load–area curve, has no effect on the adhesion force, an interesting and unexpected result that has nevertheless been verified experimentally [28.10, 113, 114]. Equations (28.21) and (28.22) provide the framework for analyzing results of adhesion measurements (Fig. 28.9) of contacting solids, known as contact mechanics, and for studying the effects of surface conditions and time on adhesion energy hysteresis (Sect. 28.5.4).

The JKR theory has been extended [28.115, 116] to consider rigid or elastic substrates separated by thin compliant layers of very different elastic moduli, a situation commonly encountered in SFA and atomic force microscopy (AFM) experiments. The deformation of the system is then strongly dependent on the ratio of r to the thickness of the confined layer. At small r (low L), the deformation occurs mostly in the thin confined layer, whereas at large r (large L), it occurs mainly in the substrates. Because of the changing distribution of traction across the contact, the adhesion force in a layered system is also modified from that of isotropic systems (28.22) so that it is no longer independent of the elastic moduli.

28.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch

In a contact between two rough surfaces, the real area of contact varies with the applied load in a different manner than between smooth surfaces [28.117, 118]. For nonadhering surfaces exhibiting an exponential distribution of *elastically* deforming asperities (spherical caps of equal radius), it has been shown that the contact area for rough surfaces increases approximately

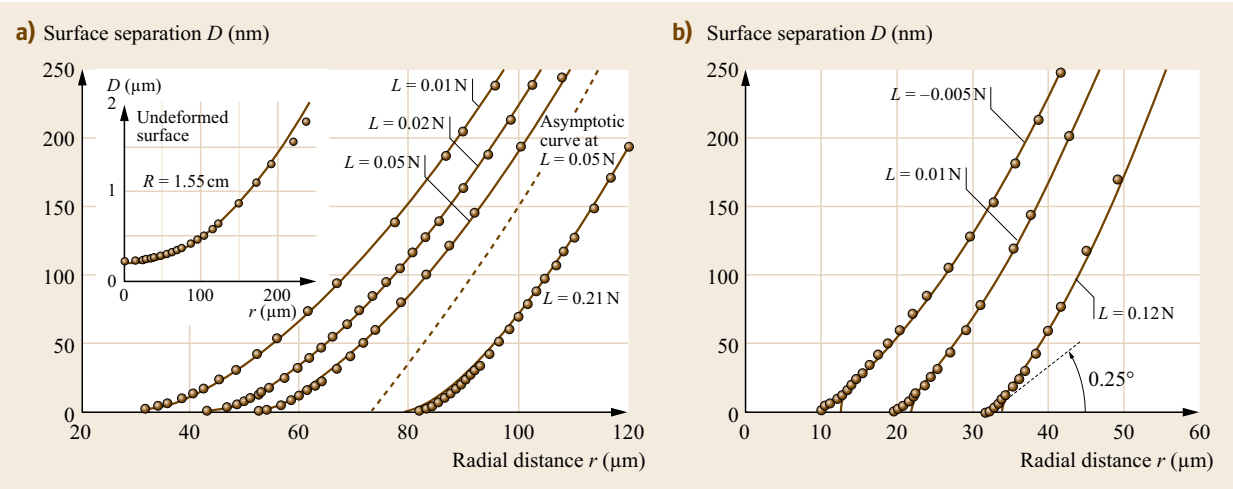


Fig. 28.9a,b Experimental data on contact mechanics for ideal Hertz and JKR contacts. **(a)** Measured profiles of surfaces in nonadhesive contact (*circles*) compared with Hertz profiles (*continuous curves*). The system was mica surfaces in a concentrated KCl solution in which the surfaces do not adhere. When not in contact, the surface shape is accurately described by a sphere of radius $R = 1.55\,\text{cm}$ (*insert*). The applied loads were 0.01, 0.02, 0.05, and 0.21 N. The last profile was measured in a different region of the surfaces where the local radius of curvature was 1.45 cm. The Hertz profiles correspond to central displacements of $\delta = 66.5, 124, 173,$ and 441 nm. The *dashed line* shows the shape of the undeformed sphere corresponding to the curve at a load of 0.05 N; it fits the experimental points at larger distances (not shown). **(b)** Surface profiles measured with adhesive contact (mica surfaces adhering in dry nitrogen gas) at applied loads of -0.005, 0.01, and 0.12 N. The *continuous lines* are JKR profiles obtained by adjusting the central displacement in each case to get the best fit to points at larger distances. The values are $\delta = -4.2, 75.6,$ and 256 nm. Note that the scales of this figure exaggerate the apparent angle at the junction of the surfaces. This angle, which is insensitive to load, is only about 0.25°. (after [28.113], with permission from Elsevier Science)

linearly with the applied normal force (load) L instead of as $L^{2/3}$ for smooth surfaces [28.117]. It has also been shown that for *plastically* deforming metal microcontacts the real contact area increases with load as $A \propto L$ [28.119, 120]. In systems with attractive surface forces, there is a competition between this attraction and repulsive forces arising from compression of high asperities. As a result, the adhesion in such systems can be very low, especially if the surfaces are not easily deformed [28.121, 122]. The opposite is possible for soft (viscoelastic) surfaces where the real (molecular) contact area might be larger than for two perfectly smooth surfaces [28.123]. The size of the real contact area at a given normal force is also an important issue in studies of nanoscale friction, both of single-asperity contacts (Sect. 28.7) and of contacts between rough surfaces (Sect. 28.9.2).

Adhesion forces may also vary depending on the commensurability of the crystallographic lattices of the interacting surfaces. The adhesion between two mica surfaces has been studied as a function of the orientation (twist angle) of their surface lattices [28.124]. The forces were measured in air, water, and an aqueous salt solution where oscillatory structural forces were present. In humid air, the adhesion was found to be relatively independent of the twist angle θ due to the adsorption of a 0.4 nm-thick amorphous layer of organics and water at the interface. In contrast, in water, sharp adhesion peaks (energy minima) occurred at $\theta = 0°$, $\pm 60°$, $\pm 120°$ and 180°, corresponding to the coincidence angles of the surface lattices (Fig. 28.10). As little as $\pm 1°$ away from these peaks, the energy decreased by 50%. In aqueous KCl solution, due to potassium ion adsorption, the water between the surfaces becomes ordered, resulting in an oscillatory force profile where the adhesive minima occur at discrete separations of about 0.25 nm, corresponding to integral numbers of water layers. The whole interaction potential was then found to depend on the orientation of the surface lattices, and the effect extended at least four molecular layers.

It has also been appreciated that the structure of the confining surfaces is just as important as the nature of the liquid for determining the solvation forces [28.56, 57, 126–128]. Between two surfaces that are completely flat but unstructured, the liquid molecules will order into layers, but there will be no lateral ordering within the layers. In other words, there will be positional ordering normal but not parallel to the surfaces. If the surfaces have a crystalline (periodic) lattice, this may

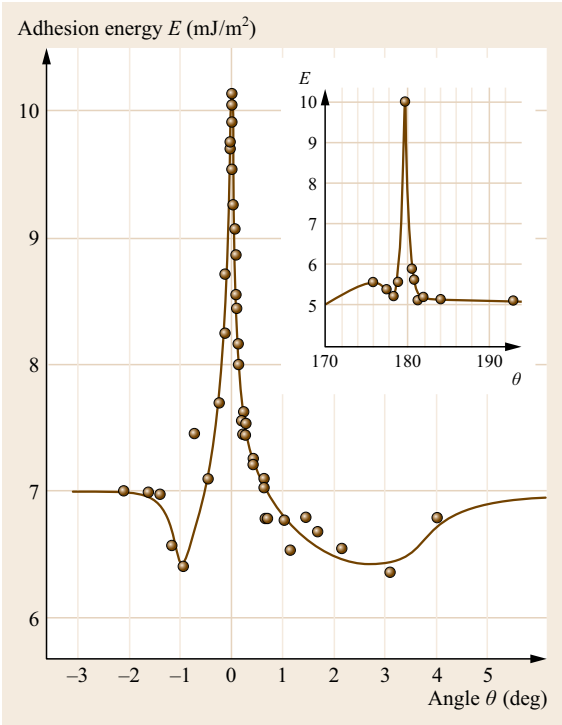


Fig. 28.10 Adhesion energy for two mica surfaces in contact in water (in the primary minimum of an oscillatory force curve) as a function of the mismatch angle θ about $\theta = 0°$ and 180° between the mica surface lattices (after [28.125] with permission)

induce ordering parallel to the surfaces as well, and the oscillatory force then also depends on the structure of the surface lattices. Further, if the two lattices have different dimensions (mismatched or incommensurate lattices), or if the lattices are similar but are not in register relative to each other, the oscillatory force law is further modified [28.124, 129] and the tribological properties of the film are also influenced, as discussed in Sect. 28.9 [28.129, 130].

As shown by the experiments, these effects can alter the magnitude of the adhesive minima found at a given separation within the last one or two nanometers of a thin film by a factor of two. The force barriers (maxima) may also depend on orientation. This could be even more important than the effects on the minima. A high barrier could prevent two surfaces from coming closer together into a much deeper adhesive well. Thus the maxima can effectively contribute to determining not only the final separation of two surfaces, but also their final adhesion. Such considerations should be particularly important for determining the thickness and strength of intergranular spaces in ceramics, the adhesion forces between colloidal particles in concentrated electrolyte solution, and the forces between two surfaces in a crack containing capillary-condensed water.

For surfaces that are randomly rough, oscillatory forces become smoothed out and disappear altogether, to be replaced by a purely monotonic solvation force [28.48, 56, 57]. This occurs even if the liquid molecules themselves are perfectly capable of ordering into layers. The situation of symmetric liquid molecules confined between rough surfaces is therefore not unlike that of asymmetric molecules between smooth surfaces (Sect. 28.4.3 and Fig. 28.6a). To summarize, for there to be an oscillatory solvation force, the liquid molecules must be able to be correlated over a reasonably long range. This requires that both the liquid molecules and the surfaces have a high degree of order or symmetry. If either is missing, so will the oscillations be. Depending on the size of the molecules to be confined, a roughness of only a few tenths of a nanometer is often sufficient to eliminate any oscillatory component of the force law [28.31, 56].

28.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis

Under ideal conditions the adhesion energy is a well-defined thermodynamic quantity. It is normally denoted by E or W (the work of adhesion) or γ (the surface tension, where $W = 2\gamma$) and gives the reversible work done on bringing two surfaces together or the work needed to separate two surfaces from contact. Under ideal, equilibrium conditions these two quantities are the same, but under most realistic conditions they are not; the work needed to separate two surfaces is always greater than that originally gained by bringing them together. An understanding of the molecular mechanisms underlying this phenomenon is essential for understanding many adhesion phenomena, energy dissipation during loading-unloading cycles, contact angle hysteresis, and the molecular mechanisms associated with many frictional processes.

Adhesion hysteresis may be thought of as being due to mechanical effects such as instabilities, or chemical effects such as interdiffusion, interdigitation, molecular reorientations and exchange processes occurring at an interface after contact, as illustrated in Fig. 28.11. Such processes induce roughness and chemical heterogeneity even though initially (and after separation and reequilibration) both surfaces are perfectly smooth and chemically homogeneous. In general, if the energy change, or work done, on separating two surfaces from adhesive contact is not fully recoverable on bringing the two surfaces back into contact again, the adhesion hys-

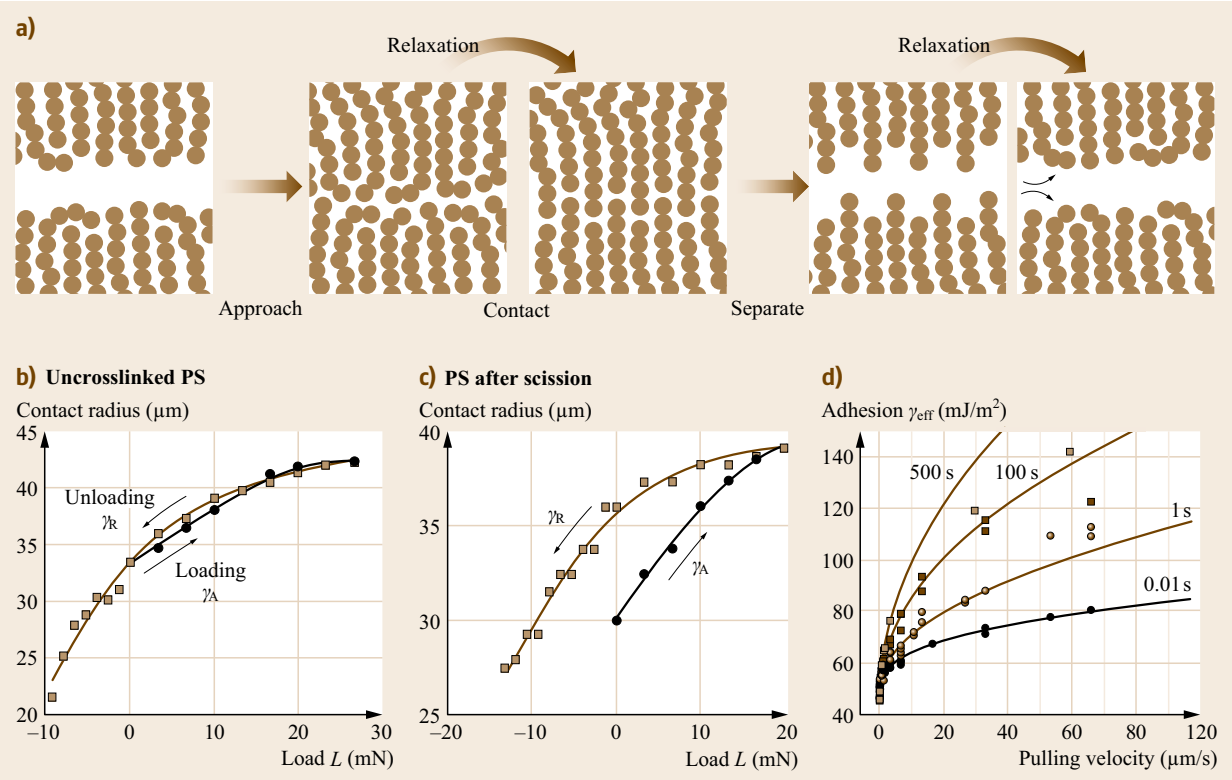


Fig. 28.11 **(a)** Schematic representation of interpenetrating chains. **(b,c)** JKR plots (contact radius r as a function of applied load L) showing small adhesion hysteresis for uncross-linked polystyrene and larger adhesion hysteresis after chain scission at the surfaces after 18 h irradiation with ultraviolet light in an oxygen atmosphere. The adhesion hysteresis continues to increase with the irradiation time. **(d)** Rate-dependent adhesion of hexadecyl trimethyl ammonium bromide (CTAB) surfactant monolayers. The *solid curves* (after [28.131]) are fits to experimental data on CTAB adhesion after different contact times (after [28.132]) using an approximate analytical solution for a JKR model, including crack tip dissipation. Due to the limited range of validity of the approximation, the fits rely on the part of the experimental data with low effective adhesion energy only. From the fits one can determine the thermodynamic adhesion energy, the characteristic dissipation velocity, and the intrinsic dissipation exponent of the model ((**a**) after [28.133], American Chemical Society), ((**b,c**) after [28.134], American Association for the Advancement of Science), ((**d**) after [28.131] American Chemical Society)

teresis may be expressed as

$$W_R > W_A \quad \text{or} \quad \Delta W = (W_R - W_A) > 0\,, \qquad (28.23)$$

where W_R and W_A are the adhesion or surface energies for receding (separating) and advancing (approaching) two solid surfaces respectively.

Hysteresis effects are also commonly observed in wetting/dewetting phenomena [28.135]. For example, when a liquid spreads and then retracts from a surface the advancing contact angle θ_A is generally larger than the receding angle θ_R. Since the contact angle θ is related to the liquid–vapor surface tension γ_L and the solid–liquid adhesion energy W by the Dupré equation

$$(1 + \cos\theta)\gamma_L = W\,, \qquad (28.24)$$

we see that wetting hysteresis or contact angle hysteresis ($\theta_A > \theta_R$) actually implies adhesion hysteresis, $W_R > W_A$, as given by (28.23).

Energy-dissipating processes such as adhesion and contact angle hysteresis arise because of practical constraints of the *finite time* of measurements and the *finite elasticity* of materials. This prevents many loading-unloading or approach-separation cycles from being thermodynamically reversible, even though they would be if carried out infinitely slowly. By thermodynamically irreversible one simply means that one cannot go through the approach-separation cycle via a continuous series of equilibrium states, because some of these are connected via spontaneous – and therefore thermodynamically irreversible – instabilities or transitions where energy is liberated and therefore *lost*

via heat or phonon release [28.136]. This is an area of much current interest and activity, especially regarding the fundamental molecular origins of adhesion and friction in polymer and surfactant systems, and the relationships between them [28.131, 132, 134, 137].

28.6 Introduction: Different Modes of Friction

Most frictional processes occur with the sliding surfaces becoming damaged in one form or another [28.138]. This may be referred to as normal friction. In the case of brittle materials, the damaged surfaces slide past each other while separated by relatively large, micrometer-sized wear particles. With more ductile surfaces, the damage remains localized to nanometer-sized, plastically deformed asperities. Some features of the friction between damaged surfaces will be described in Sect. 28.7.4.

There are also situations in which sliding can occur between two perfectly smooth, undamaged surfaces. This may be referred to as interfacial sliding or boundary friction and is the focus of the following sections. The term boundary lubrication is more commonly used to denote the friction of surfaces that contain a thin protective lubricating layer such as a surfactant monolayer, but here we shall use the term more broadly to include any molecularly thin solid, liquid, surfactant, or polymer film.

Experiments have shown that as a liquid film becomes progressively thinner its physical properties change, at first quantitatively and then qualitatively [28.33, 35, 139, 140]. The quantitative changes are manifested by an increased viscosity, non-Newtonian flow behavior, and the replacement of normal melting by a glass transition, but the film remains recognizable as a liquid (Fig. 28.12). In tribology, this regime is commonly known as the mixed lubrication regime, where the rheological properties of a film are intermediate between the bulk and boundary properties. One may also refer to it as the intermediate regime (Table 28.4).

For even thinner films, the changes in behavior are more dramatic, resulting in a qualitative change in properties. Thus first-order phase transitions can now occur to solid or liquid-crystalline phases [28.34, 128, 133, 141–145], whose properties can no longer be characterized even qualitatively in terms of bulk or continuum liquid properties such as viscosity. These films now exhibit yield points (characteristic of fracture in solids) and their molecular diffusion and relaxation times can be ten orders of magnitude longer than in the bulk liquid or even in films that are just slightly thicker. The three friction regimes are summarized in Table 28.4.

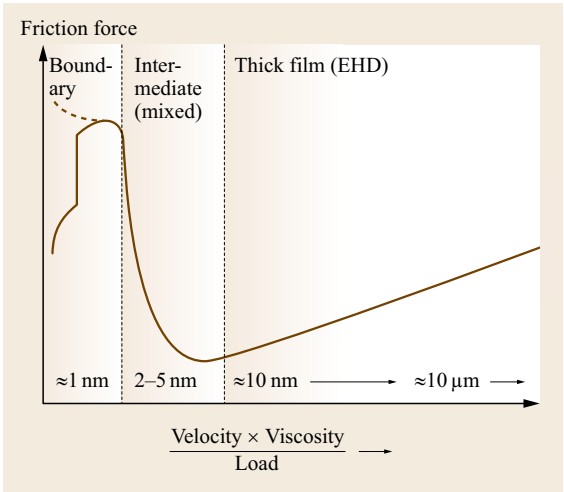


Fig. 28.12 Stribeck curve: an empirical curve giving the trend generally observed in the friction forces or friction coefficients as a function of sliding velocity, the bulk viscosity of the lubricating fluid, and the applied load (normal force). The three friction/lubrication regimes are known as the boundary lubrication regime (Sect. 28.7), the intermediate or mixed lubrication regime (Sect. 28.8.2), and the thick film or elastohydrodynamic (EHD) lubrication regime (Sect. 28.8.1). The film thicknesses believed to correspond to each of these regimes are also shown. For thick films, the friction force is purely viscous, e.g., Couette flow at low shear rates, but may become complicated at higher shear rates where EHD deformations of surfaces can occur during sliding (after [28.5], with permission)

Table 28.4 The three main tribological regimes characterizing the changing properties of liquids subjected to increasing confinement between two solid surfaces[a]. Based on work by *Granick* et al. [28.139], and others [28.28, 133, 141], on the dynamic properties of short-chain molecules such as alkanes and polymer melts confined between surfaces

Regime	Conditions for getting into this regime	Static/equilibrium properties[b]	Dynamic properties[c]
Bulk		Bulk (continuum) properties:	Bulk (continuum) properties:
	• Thick films (> 10 molecular diameters, $\gg R_g$ for polymers) • Low or zero loads • High shear rates	• Bulk liquid density • No long-range order	• Newtonian viscosity • Fast relaxation times • No glass temperature • No yield point • Elastohydrodynamic lubrication
Intermediate mixed		Modified fluid properties include:	Modified rheological properties include:
	• Intermediately thick films (4−10 molecular diameters, $\approx R_g$ for polymers) • Low loads or pressure	• Modified positional and orientational order[a] • Medium- to long-range molecular correlations • Highly entangled states	• Non-Newtonian flow • Glassy states • Long relaxation times • Mixed lubrication
Boundary		Onset of nonfluid-like properties:	Onset of tribological properties:
	• Molecularly thin films (< 4 molecular diameters) • High loads or pressure • Low shear rates • Smooth surfaces or asperities	• Liquid-like to solid-like phase transitions • Appearance of new liquid-crystalline states • Epitaxially induced long-range ordering	• No flow until yield point or critical shear stress reached • Solid-like film behavior characterized by defect diffusion, dislocation motion, shear melting • Boundary lubrication

[a] Confinement can lead to an increased or decreased order in a film, depending both on the surface lattice structure and the geometry of the confining cavity
[b] In each regime both the static and dynamic properties change. The static properties include the film density, the density distribution function, the potential of mean force, and various positional and orientational order parameters
[c] Dynamic properties include viscosity, viscoelastic constants, and tribological yield points such as the friction coefficient and critical shear stress

28.7 Relationship Between Adhesion and Friction

28.7.1 Amontons' Law

Early theories and mechanisms for the dependence of friction on the applied normal force or load L were developed by da Vinci, Amontons, Coulomb and Euler [28.146]. For the macroscopic objects investigated, the friction was found to be directly proportional to the load, with no dependence on the contact area. This is described by the so-called Amontons' law

$$F = \mu L\,, \tag{28.25}$$

where F is the shear or friction force and μ is a constant defined as the coefficient of friction. This friction law is still the principal means of quantitatively describing the friction between surfaces. However, particularly in the case of adhering surfaces, Amontons' law does not adequately describe the friction behavior with load, because of the finite friction force measured at zero and even negative applied loads.

When a lateral force, or shear stress, is applied to two surfaces in adhesive contact, the surfaces initially remain *pinned* to each other until some critical shear force is reached. At this point, the surfaces begin to slide past each other either smoothly or in jerks. The frictional force needed to initiate sliding from rest is known as the static friction force, denoted by F_s, while the force needed to maintain smooth sliding is referred to as the kinetic or dynamic friction force, denoted by F_k. In general, $F_s > F_k$. Two sliding surfaces may

also move in regular jerks, known as stick–slip sliding, which is discussed in more detail in Sect. 28.8.3. Such friction forces cannot be described by models used for thick films that are viscous (Sect. 28.8.1) and, therefore, shear as soon as the smallest shear force is applied.

28.7.2 Adhesion Force and Load Contribution to Interfacial Friction

In an adhesive contact, the contact area as a function of load is thus generally well described by the JKR (28.21). The friction force between two molecularly smooth surfaces sliding in adhesive contact is not simply proportional to the applied load L as might be expected from Amontons' law. There is an additional adhesion contribution that is proportional to the area of contact, A. Thus, in general, the interfacial friction force of dry, unlubricated surfaces sliding smoothly past each other in adhesive contact is given by

$$F = F_{\mathrm{k}} = S_{\mathrm{c}}A + \mu L\,, \tag{28.26}$$

where S_{c} is the critical shear stress (assumed to be constant), $A = \pi r^2$ is the contact area of radius r given by (28.21), and μ is the coefficient of friction. For low loads we have

$$F = S_{\mathrm{c}}A = S_{\mathrm{c}}\pi r^2 = S_{\mathrm{c}}\pi\left[\frac{R}{K}\left(L + 6\pi R\gamma + \sqrt{12\pi R\gamma L + (6\pi R\gamma)^2}\right)\right]^{2/3}, \tag{28.27}$$

whereas for high loads (or high μ), or when γ is very low [28.147–151], (28.26) reduces to Amontons' law: $F = \mu L$. Depending on whether the friction force in (28.26) is dominated by the first or second term, one may refer to the friction as adhesion-controlled or load-controlled respectively.

The following friction model (Fig. 28.13a), first proposed by *Tabor* [28.152] and developed further by others [28.32, 153, 154], has been quite successful at explaining the interfacial and boundary friction of two solid crystalline surfaces sliding past each other in the absence of wear. The surfaces may be unlubricated, or they may be separated by a monolayer or more of some boundary lubricant or liquid molecules. In this model, the values of the critical shear stress S_{c}, and the coefficient of friction μ, in (28.26) are calculated in terms of the energy needed to overcome the attractive intermolecular forces and compressive externally applied load as one surface is raised and then slid across the molecular-sized asperities of the other.

This model is similar to pushing a cart over a road of cobblestones where the cartwheels (which represent the molecules of the upper surface or film) must be made to roll over the cobblestones (representing the molecules of the lower surface) before the cart can move. When at rest, the cartwheels find grooves between the cobblestones where they sit in potential-energy minima, and so the cart is at some stable mechanical equilibrium. A certain lateral force (the push) is required to raise the cartwheels against the force of gravity in order to initiate motion. Motion will continue as long as the cart is pushed, and rapidly stops once it is no longer pushed. Energy is dissipated by the liberation of heat every time a wheel hits the next cobblestone. The cobblestone model is not unlike the Coulomb and interlocking asperity models of friction [28.146] except that it is being applied at the molecular level and for a situation where the external load is augmented by attractive intermolecular forces.

There are thus two contributions to the force pulling two surfaces together: the externally applied load or pressure, and the (internal) attractive intermolecular forces that determine the adhesion between the two surfaces. Each of these contributions affects the friction force in a different way, which we will discuss in more detail below.

Adhesion Force Contribution

When the two surfaces are initially in adhesive contact, the surface molecules will adjust themselves to fit snugly together [28.156], in an analogous manner to the self-positioning of the cartwheels on the cobblestone road. A small tangential force applied to one surface will therefore not result in the sliding of that surface relative to the other. The attractive van der Waals forces between the surfaces must first be overcome by having the surfaces separate by a small amount. To initiate motion, let the separation between the two surfaces increase by a small amount ΔD, while the lateral distance moved is $\Delta\sigma$. The energy put into the system by the force F acting over a lateral distance $\Delta\sigma$ is

$$\text{Input energy: } F \times \Delta\sigma\,. \tag{28.28}$$

This energy may be equated with the change in interfacial or surface energy associated with separating the surfaces by ΔD, i. e., from the equilibrium separation $D = D_0$ to $D = (D_0 + \Delta D)$. Since $\gamma \propto D^{-2}$ for two flat surfaces (Sect. 28.3.1, Table 28.2), the energy cost may be approximated by

$$\text{Surface energy change} \times \text{area:}\quad 2\gamma A\left[1 - \frac{D_0^2}{(D_0 + \Delta D)^2}\right] \approx 4\gamma A\left(\frac{\Delta D}{D_0}\right), \tag{28.29}$$

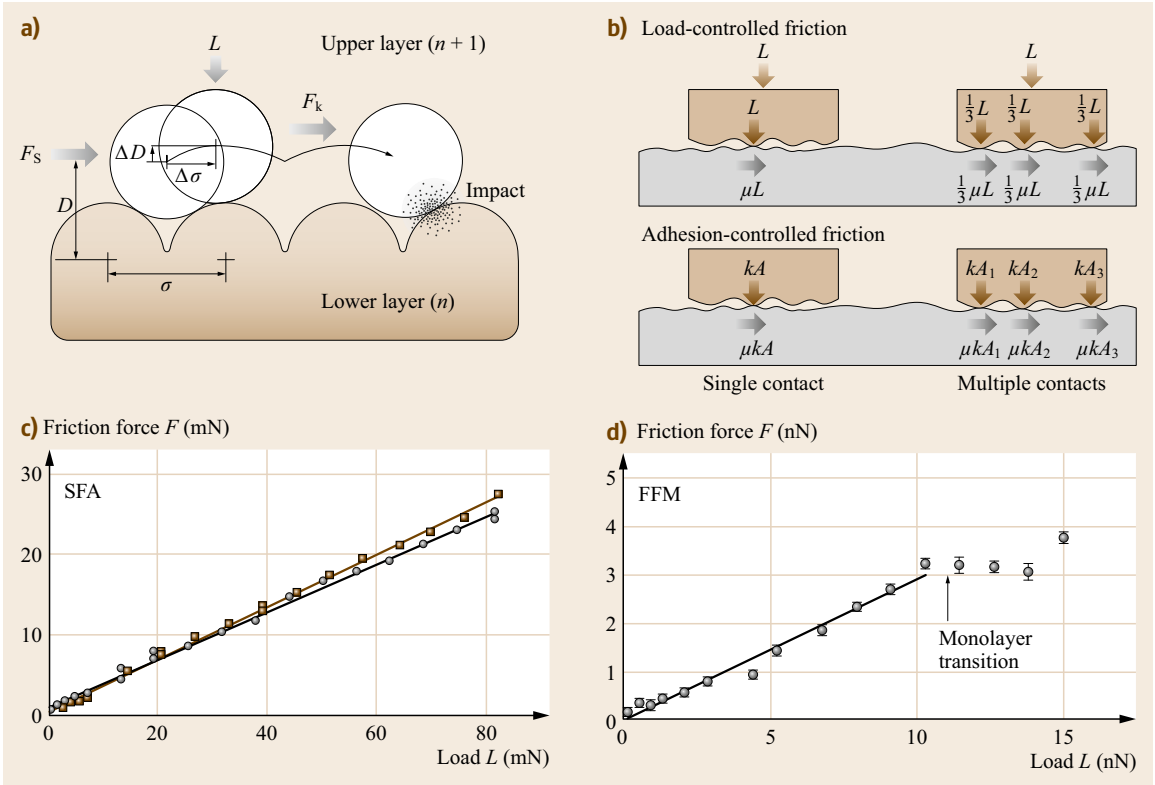


Fig. 28.13 **(a)** Schematic illustration of how one molecularly smooth surface moves over another when a lateral force F is applied (the *cobblestone model*). As the upper surface moves laterally by some fraction of the lattice dimension $\Delta\sigma$, it must also move up by some fraction of an atomic or molecular dimension ΔD before it can slide across the lower surface. On impact, some fraction ε of the kinetic energy is *transmitted* to the lower surface, the rest being *reflected* back to the colliding molecule (upper surface) (after [28.155], with permission). **(b)** Difference in the local distribution of the total applied external load or normal adhesive force between load-controlled nonadhering surfaces and adhesion-controlled surfaces. In the former case, the total friction force F is given either by $F = \mu L$ for one contact point (*left side*) or by $F = 1/3\mu L + 1/3\mu L + 1/3\mu L = \mu L$ for three contact points (*right side*). Thus the load-controlled friction is always proportional to the applied load, independently of the number of contacts and of their geometry. In the case of adhering surfaces, the effective *internal* load is given by kA, where A is the real local contact area, which is proportional to the number of intermolecular bonds being made and broken across each single contact point. The total friction force is now given by $F = \mu kA$ for one contact point (*left side*), and $F = \mu(kA_1 + kA_2 + kA_3) = \mu kA_{\text{tot}}$ for three contact points (*right side*). Thus, for adhesion-controlled friction, the friction is proportional to the real contact area, at least when no additional external load is applied to the system (after [28.151], with permission, American Chemical Society). **(c,d)** Friction force between benzyltrichlorosilane monolayers chemically bound to glass or Si, measured in ethanol ($\gamma < 1\,\text{mJ/m}^2$). **(c)** SFA measurements where both glass surfaces were covered with a monolayer. *Circles* and *squares* show two different experiments: one with $R = 2.6\,\text{cm}$, $v = 0.15\,\mu\text{m/s}$, giving $\mu = 0.33 \pm 0.01$; the other with $R = 1.6\,\text{cm}$, $v = 0.5\,\mu\text{m/s}$, giving $\mu = 0.30 \pm 0.01$. **(d)** Friction force microscopy (FFM) measurements of a monolayer-functionalized Si tip ($R = 11\,\text{nm}$) sliding on a monolayer-covered glass surface at $v = 0.15\,\mu\text{m/s}$, giving $\mu = 0.30 \pm 0.01$. Note the different scales in **(c,d)** (after [28.150], with permission, American Chemical Society)

where γ is the surface energy, A the contact area, and D_0 the surface separation at equilibrium. During steady-state sliding (kinetic friction), not all of this energy will be lost or absorbed by the lattice every time the surface molecules move by one lattice spacing: some fraction will be reflected during each impact of the cartwheel molecules [28.155]. Assuming that a fraction ε of the above surface energy is lost every time the surfaces move across the characteristic length $\Delta\sigma$ (Fig. 28.13a), we obtain after equating (28.28)

and (28.29)

$$S_c = \frac{F}{A} = \frac{4\gamma\varepsilon\Delta D}{D_0\Delta\sigma} \,. \tag{28.30}$$

The above model suggests that all interfaces, whether dry or lubricated, dilate just before they shear or slip. This dilation, which provides the crucial extra space for the molecules to slide across each other, is known to occur in macroscopic systems [28.157] and for nanoscopic systems it has been computed [28.128] and also measured experimentally [28.158].

This model may be extended to lubricated sliding, where a thin liquid film is present between the surfaces. With an increase in the number of liquid layers between the surfaces, D_0 increases while ΔD decreases, hence the friction force decreases. This is precisely what is observed, but with more than one liquid layer between two surfaces the situation becomes too complex to study analytically. Furthermore, even in systems as simple as linear alkanes, interdigitation and interdiffusion have been found to contribute strongly to the properties of the system [28.159]. Sophisticated modeling based on computer simulations is now required, as discussed in the following section.

Relation Between Boundary Friction and Adhesion Energy Hysteresis

While the above equations suggest that there is a direct correlation between friction and adhesion, this is not the case. The correlation is really between friction and adhesion hysteresis, described in Sect. 28.5.4. In the case of friction, this subtle point is hidden in the factor ε, which is a measure of the amount of energy absorbed (dissipated, transferred, or lost) by the lower surface when it is impacted by a molecule from the upper surface. If $\varepsilon = 0$, all the energy is reflected, and there will be no kinetic friction force or any adhesion hysteresis, but the absolute magnitude of the adhesion force or energy will remain finite and unchanged. This is illustrated in Fig. 28.17.

The following simple model shows how adhesion hysteresis and friction may be quantitatively related. Let $\Delta\gamma = \gamma_R - \gamma_A$ be the adhesion energy hysteresis per unit area, as measured during a typical loading-unloading cycle. Now consider the same two surfaces sliding past each other and assume that frictional energy dissipation occurs through the same mechanism as adhesion energy dissipation, and that both occur over the same characteristic molecular length scale σ. Thus, when the two surfaces (of contact area $A = \pi r^2$) move a distance σ, equating the frictional energy $(F \times \sigma)$ to the dissipated adhesion energy $(A \times \Delta\gamma)$, we obtain

$$\text{Friction force: } F = \frac{A\,\Delta\gamma}{\sigma} = \frac{\pi r^2}{\sigma}(\gamma_R - \gamma_A) \,, \tag{28.31}$$

$$\text{or critical shear stress: } S_c = \frac{F}{A} = \frac{\Delta\gamma}{\sigma} \,, \tag{28.32}$$

which is the desired expression and has been found to give order-of-magnitude agreement between measured friction forces and adhesion energy hysteresis [28.133]. If we equate (28.32) with (28.30), since $4\Delta D/(D_0\Delta\sigma) \approx 1/\sigma$, we obtain the intuitive relation

$$\varepsilon = \frac{\Delta\gamma}{\gamma} \,. \tag{28.33}$$

External Load Contribution to Interfacial Friction

When there is no interfacial adhesion, S_c is zero. Thus, the only attractive force that needs to be overcome for sliding to occur is the externally applied load or pressure, as shown in Fig. 28.13b.

For a general semiquantitative analysis, again consider the cobblestone model used to derive (28.30), but now include an additional contribution to the surface-energy change of (28.29) due to the work done against the external load or pressure, $L\Delta D = P_{ext}A\Delta D$ (this is equivalent to the work done against gravity in the case of a cart being pushed over cobblestones). Thus

$$S_c = \frac{F}{A} = \frac{4\gamma\varepsilon\Delta D}{D_0\Delta\sigma} + \frac{P_{ext}\varepsilon\Delta D}{\Delta\sigma} \,, \tag{28.34}$$

which gives the more general relation

$$S_c = \frac{F}{A} = C_1 + C_2 + P_{ext} \,, \tag{28.35}$$

where $P_{ext} = L/A$ and C_1 and C_2 are characteristics of the surfaces and sliding conditions. The constant $C_1 = 4\gamma\varepsilon\Delta D/(D_0\Delta\sigma)$ depends on the mutual adhesion of the two surfaces, while both C_1 and $C_2 = \varepsilon\Delta D/\Delta\sigma$ depend on the topography or atomic bumpiness of the surface groups (Fig. 28.13a). The smoother the surface groups the smaller the ratio $\Delta D/\Delta\sigma$ and hence the lower the value of C_2. In addition, both C_1 and C_2 depend on ε (the fraction of energy dissipated per collision), which depends on the relative masses of the shearing molecules, the sliding velocity, the temperature, and the characteristic molecular relaxation processes of the surfaces. This is by far the most difficult parameter to compute, and yet it is the most important since it represents the energy-transfer mechanism in any friction process, and since ε can vary

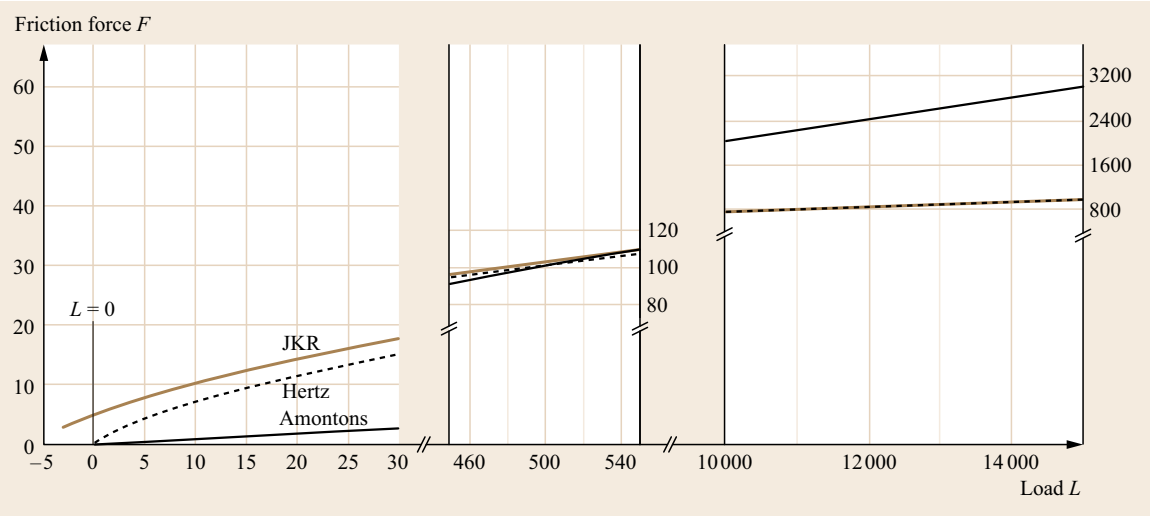


Fig. 28.14 Friction as a function of load for smooth surfaces. At low loads, the friction is dominated by the C_1A term of (28.36). The adhesion contribution (JKR curve) is most prominent near zero load where the Hertzian and Amontons' contributions to the friction are minimal. As the load increases, the adhesion contribution becomes smaller as the JKR and Hertz curves converge. In this range of loads, the linear C_2L contribution surpasses the area contribution to the friction. At much higher loads the explicit load dependence of the friction dominates the interactions, and the observed behavior approaches Amontons' law. It is interesting to note that for smooth surfaces the pressure over the contact area does not increase as rapidly as the load. This is because as the load is increased, the surfaces deform to increase the surface area and thus moderate the contact pressure (after [28.160])

between 1 and 0, it determines whether a particular friction force will be large or close to zero.

Finally, the above equation may also be expressed in terms of the friction force F

$$F = S_c A = C_1 A + C_2 L\,. \tag{28.36}$$

Equations similar to (28.35) and (28.36) were previously derived by *Derjaguin* [28.161, 162] and by *Briscoe* and *Evans* [28.163], where the constants C_1 and C_2 were interpreted somewhat differently than in this model. In the absence of any attractive interfacial force, we have $C_1 \approx 0$, and the second term in (28.35) and (28.36) should dominate (Fig. 28.14). Such situations typically arise when surfaces repel each other across the lubricating liquid film. In such cases, the total frictional force should be low and should increase linearly with the external load according to

$$F = C_2 L\,. \tag{28.37}$$

An example of such lubricated sliding occurs when two mica surfaces slide in water or in salt solution (Fig. 28.18a) where the short-range hydration forces between the surfaces are repulsive. Thus, for sliding in 0.5 M KCl it was found that $C_2 = 0.015$ [28.147]. Another case where repulsive surfaces eliminate the adhesive contribution to friction is for polymer chains attached to surfaces at one end and swollen by a good solvent [28.104]. For this class of systems, $C_2 < 0.001$ for a finite range of polymer layer compressions (normal loads L). The low friction between the surfaces in this regime is attributed to the entropic repulsion between the opposing brush layers with a minimum of entanglement between the two layers. However, with higher normal loads, the brush layers become compressed and begin to entangle, which results in higher friction [28.164]. It is important to note that (28.37) has exactly the same form as Amontons' law (28.25).

Figure 28.13c,d shows the kinetic friction force measured with both SFA and FFM (friction force microscopy, using AFM) on a system where both surfaces were covered with a chemically bound benzyltrichlorosilane monolayer [28.150]. When immersed in ethanol, the adhesion in this system is low, and very different contact areas and loads give a linear dependence of F on L with the same friction coefficients, and $F \to 0$ as $L \to 0$. In the FFM measurements (Fig. 28.13d), the plateau in the data at higher loads suggests a transition in the monolayers, similar to previous observations on other monolayer systems. The pressure in the contact region in the SFA is much lower than in the FFM, and no transitions in the friction forces or in the thickness of the confined monolayers were observed in the SFA experiments (and no damage to the monolayers or the

underlying substrates was observed during the experiments, indicating that the friction was wearless).

A thermodynamic analog of the Coulomb or cobblestone models (Sect. 28.7.2) based on the contact value theorem [28.1, 147, 160] can explain why $F \propto L$ also holds at the microscopic or molecular level. In this analysis we consider the surface molecular groups as being momentarily compressed and decompressed as the surfaces move along. Under irreversible conditions, which always occur when a cycle is completed in a finite amount of time, the energy lost in the compression-decompression cycle is dissipated as heat. For two nonadhering surfaces, the stabilizing pressure P_i acting locally between any two elemental contact points i of the surfaces may be expressed by the contact value theorem [28.1]

$$P_i = \rho_i k_B T = \frac{k_B T}{V_i} , \tag{28.38}$$

where $\rho_i = V_i^{-1}$ is the local number density (per unit volume) or activity of the interacting entities, be they molecules, atoms, ions or the electron clouds of atoms. This equation is essentially the osmotic or entropic pressure of a gas of confined molecules. As one surface moves across the other, local regions become compressed and decompressed by a volume ΔV_i. The work done per cycle can be written as $\varepsilon P_i \Delta V_i$, where ε ($\varepsilon \leq 1$) is the fraction of energy per cycle lost as heat, as defined earlier. The energy balance shows that, for each compression-decompression cycle, the dissipated energy is related to the friction force by

$$F_i x_i = \varepsilon P_i \Delta V_i , \tag{28.39}$$

where x_i is the lateral distance moved per cycle, which can be the distance between asperities or the distance between surface lattice sites. The pressure at each contact junction can be expressed in terms of the local normal load L_i and local area of contact A_i as $P_i = L_i/A_i$. The volume change over a cycle can thus be expressed as $\Delta V_i = A_i z_i$, where z_i is the vertical distance of confinement. Inserting these into (28.39), we get

$$F_i = \varepsilon L_i \left(\frac{z_i}{x_i}\right) , \tag{28.40}$$

which is independent of the local contact area A_i. The total friction force is thus

$$F = \sum F_i = \sum \varepsilon L_i \left(\frac{z_i}{x_i}\right) = \varepsilon \left\langle \frac{z_i}{x_i} \right\rangle \sum L_i = \mu L , \tag{28.41}$$

where it is assumed that on average the local values of L_i and P_i are independent of the local slope z_i/x_i. Therefore, the friction coefficient μ is a function only of the average surface topography and the sliding velocity, but is independent of the local (real) or macroscopic (apparent) contact areas. While this analysis explains nonadhering surfaces, there is still an additional explicit contact area contribution for the case of adhering surfaces, as in (28.36). The distinction between the two cases arises because the initial assumption of the contact value theorem (28.38) is incomplete for adhering systems.

Molecular-scale roughness, real contact areas, and changes therein are difficult to quantify in nanoscale friction experiments, which involve deformable substrates in continuous motion under considerable pressure. In contrast, the nominal contact area can be either directly measured (e.g., in the SFA) or estimated from contact mechanics models, although concerns have been raised about the applicability of such models on the nanometer and subnanometer scale [28.165] where atomic-level roughness causes local changes in the pressure. Recent computer simulations [28.166–168] have suggested that within the nominal contact area, the size of the real, molecular contact area and its increase with increasing load are adhesion dependent, but in a different manner than that observed for nominal or macroscopic contacts.

28.7.3 Examples of Experimentally Observed Friction of Dry Surfaces

Numerous model systems, including a variety of both unlubricated and solid- and liquid-lubricated surfaces, have been studied with a surface forces apparatus (SFA) modified for friction experiments (Sect. 28.2.3). The apparatus allows for control of load (normal force) and sliding speed, and simultaneous measurement of surface separation, surface shape, true (molecular) area of contact between smooth surfaces, and friction forces.

Figure 28.15 shows the contact area A and friction force F, both plotted against the applied load L in an experiment in which two molecularly smooth surfaces of mica in adhesive contact were slid past each other in an atmosphere of dry nitrogen gas. This is an example of the low-load adhesion-controlled limit, which is excellently described by (28.27). In a number of different experiments, S_c was measured to be $2.5 \times 10^7\,\mathrm{N\,m^{-2}}$ and to be independent of the sliding velocity [28.32, 169]. Note that there is a friction force even at negative loads, where the surfaces are still sliding in adhesive contact.

The high friction force of unlubricated sliding can often be reduced by treating the solid surface with a boundary layer of some other solid material that exhibits lower friction, such as a surfactant monolayer,

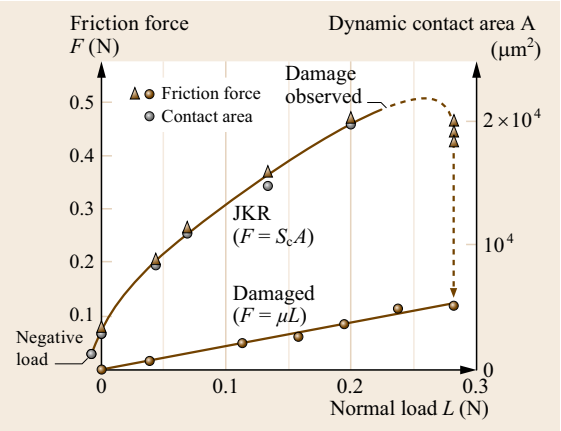


Fig. 28.15 Friction force F and contact area A versus load L for two mica surfaces sliding in adhesive contact in dry air. The contact area is well described by the JKR theory, (28.21), even during sliding, and the friction force is found to be directly proportional to this area, (28.27). The *vertical dashed line* and *arrow* show the transition from interfacial to normal friction with the onset of wear (*lower curve*). The sliding velocity is $0.2\,\mu m\,s^{-1}$ (after [28.32], with permission, American Society of Mechanical Engineers)

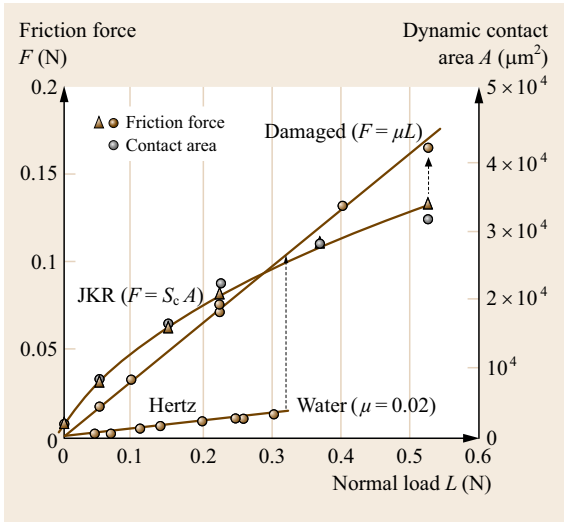


Fig. 28.16 Sliding of mica surfaces, each coated with a 2.5 nm-thick monolayer of calcium stearate surfactant, in the absence of damage (obeying JKR-type boundary friction) and in the presence of damage (obeying Amontons-type normal friction). Note that both for this system and for the bare mica in Figs. 28.15 and 28.18a, the friction force obeys Amontons' law with a friction coefficient of $\mu \approx 0.3$ after damage occurs. At much higher applied loads, the undamaged surfaces also follow Amontons-type sliding, but for a different reason: the dependence on adhesion becomes smaller. *Lower line*: interfacial sliding with a monolayer of water between the mica surfaces (load-controlled friction, Fig. 28.18a), shown for comparison (after [28.169], with permission from Elsevier Science)

or by ensuring that during sliding a thin liquid film remains between the surfaces (as will be discussed in Sect. 28.8). The effectiveness of a solid boundary lubricant layer on reducing the forces of friction is illustrated in Fig. 28.16. Comparing this with the friction of the unlubricated/untreated surfaces (Fig. 28.15) shows that the critical shear stress has been reduced by a factor of about ten: from 2.5×10^7 to $3.5\times10^6\,N\,m^{-2}$. At much higher applied loads or pressures, the friction force is proportional to the load, rather than the area of contact [28.170], as expected from (28.26).

The adhesion and friction of fluorocarbon surfaces (surfactant-coated boundary lubricant layers) have been compared to those of hydrocarbon surfaces [28.171]. It was concluded that well-ordered fluorocarbon surfaces have high friction, in spite of their lower adhesion energy. The low friction coefficient of Teflon (polytetrafluoroethylene, PTFE) must, therefore, be due to some effect other than low adhesion. On a related issue, it has been found that C_{60} surfaces also exhibit low adhesion but high friction [28.172]. In both cases the high friction appears to arise from the bulky surface groups – fluorocarbon compared to hydrocarbon groups in the former, large fullerene spheres in the latter. Apparently, the fact that C_{60} molecules rotate *in their lattice* does not make them a good lubricant: the molecules of the opposing surface must still climb over them in order to slide, and this requires energy that is independent of whether the surface molecules are fixed or freely rotating.

Figure 28.17 illustrates the relationship between adhesion hysteresis and friction for surfactant-coated surfaces under different conditions. This effect, however, is much more general and has been shown to hold for other surfaces as well [28.30, 134, 155, 173].

Direct comparisons between absolute adhesion energies and friction forces show little correlation. In some cases, higher adhesion energies for the same system under different conditions correspond to lower friction forces. For example, for hydrophilic silica surfaces it was found that with increasing relative humidity the adhesion energy increases, but the adhesion energy hysteresis measured in a loading–unloading cycle decreases, as does the friction force [28.30].

A liquid lubricant film (Sect. 28.8.3) is usually much more effective at lowering the friction of two surfaces than a solid boundary lubricant layer. However, to use a liquid lubricant successfully, it must *wet* the surfaces, that is, it should have a high affinity for the

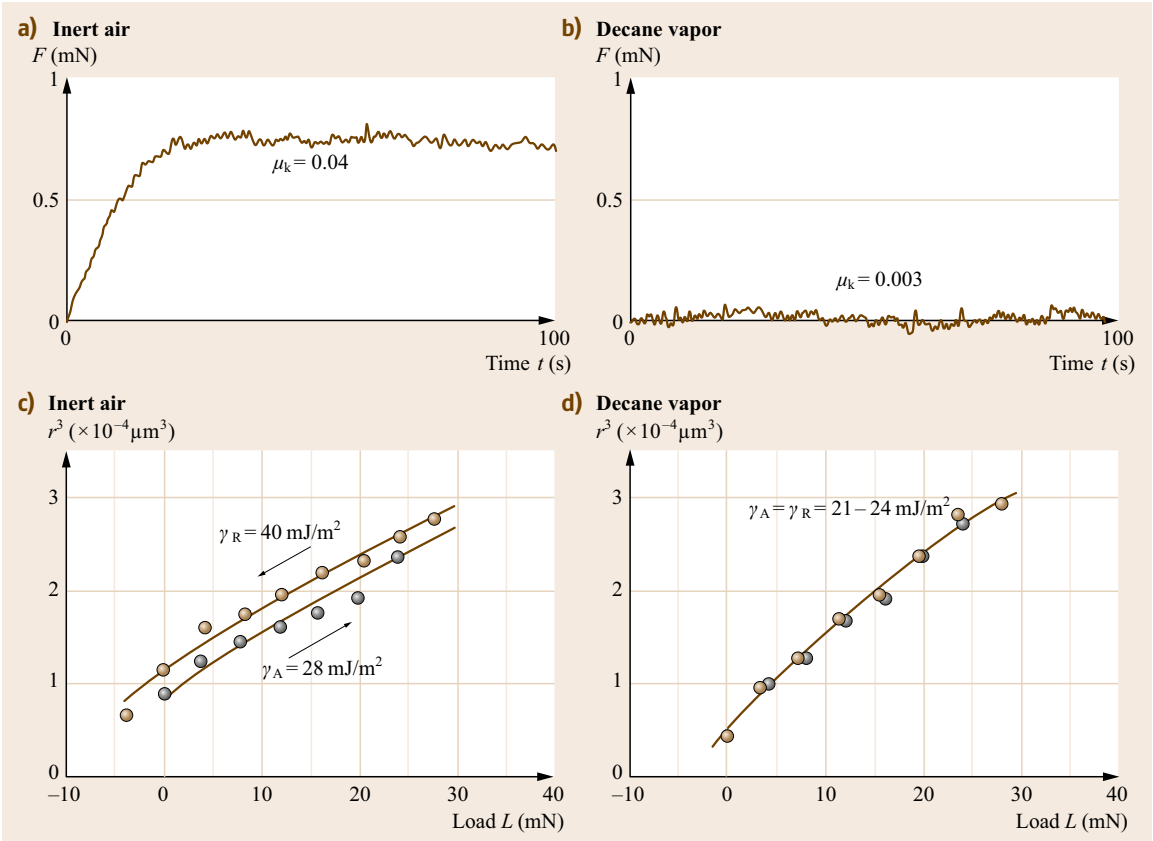


Fig. 28.17a–d *Top*: friction traces for two fluid-like calcium alkylbenzene sulfonate monolayer-coated surfaces at 25 °C showing that the friction force is much higher between dry monolayers **(a)** than between monolayers whose fluidity has been enhanced by hydrocarbon penetration from vapor **(b)**. *Bottom*: Contact radius versus load (r^3 versus L) data measured for the same two surfaces as above and fitted to the JKR equation (28.21), shown by the *solid curves*. For dry monolayers **(c)** the adhesion energy on unloading ($\gamma_R = 40\,\mathrm{mJ\,m^{-2}}$) is greater than that on loading ($\gamma_A = 28\,\mathrm{mJ\,m^{-2}}$), which is indicative of an adhesion energy hysteresis of $\Delta\gamma = \gamma_R - \gamma_A = 12\,\mathrm{mJ\,m^{-2}}$. For monolayers exposed to saturated decane vapor **(d)** their adhesion hysteresis is zero ($\gamma_A = \gamma_R$), and both the loading and unloading data are well fitted by the thermodynamic value of the surface energy of fluid hydrocarbon chains, $\gamma = 24\,\mathrm{mJ\,m^{-2}}$ (after [28.133], with permission, American Chemical Society)

surfaces, so that not all the liquid molecules become squeezed out when the surfaces come close together, even under a large compressive load. Another important requirement is that the liquid film remains a liquid under tribological conditions, i. e., that it does not epitaxially solidify between the surfaces.

Effective lubrication usually requires that the lubricant be injected between the surfaces, but in some cases the liquid can be made to condense from the vapor. This is illustrated in Fig. 28.18a for two untreated mica surfaces sliding with a thin layer of water between them. A monomolecular film of water (of thickness 0.25 nm per surface) has reduced S_c from its value for dry surfaces (Fig. 28.15) by a factor of more than 30, which may be compared with the factor of ten attained with the boundary lubricant layer (of thickness 2.5 nm per surface) in Fig. 28.16. Water appears to have unusual lubricating properties and usually gives wearless friction with no stick–slip [28.174].

The effectiveness of a water film only 0.25 nm thick to lower the friction force by more than an order of magnitude is attributed to the hydrophilicity of the mica surface (mica is wetted by water) and to the existence of a strongly repulsive short-range hydration force between such surfaces in aqueous solutions, which effectively removes the adhesion-controlled contribution to the friction force [28.148]. It is also interesting that a 0.25 nm-thick water film between two mica surfaces is sufficient to bring the coefficient of friction down to 0.01–0.02, a value that corresponds to the unusually

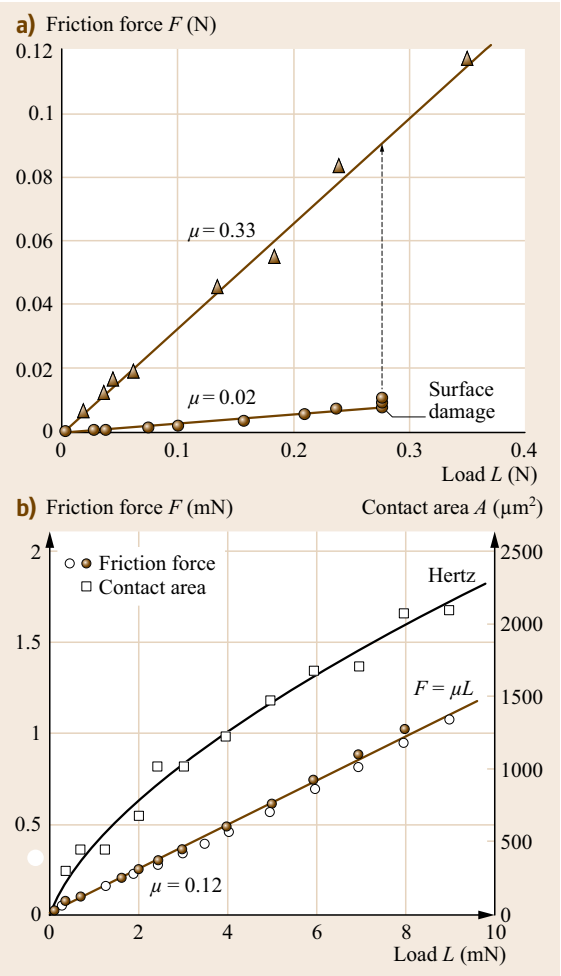


Fig. 28.18a,b Load-controlled friction. **(a)** Two mica surfaces sliding past each other while immersed in a 0.01 M KCl salt solution (nonadhesive conditions). The water film is molecularly thin, 0.25–0.5 nm thick, and the interfacial friction force is very low: $S_c \approx 5\times10^5\,\mathrm{N\,m^{-2}}$, $\mu \approx 0.02$ (before damage occurs). After the surfaces have become damaged, the friction coefficient is about 0.3 (after [28.169], with permission from Elsevier Science). **(b)** Steady-state friction force and contact area measured on a confined squalane film between two undamaged mica surfaces at $v = 0.6\,\mu\mathrm{m/s}$ in the smooth sliding regime (no stick–slip). *Open circles* show F obtained on loading (increasing L); *solid circles* show unloading. Both datasets are *straight lines* passing through the origin, as shown by the *brown line* ($\mu = 0.12$). The *black curve* is a fit of the Hertz equation (Sect. 28.5.2 and [28.1]) to the A versus L data (*open squares*) using $K = 10^{10}\,\mathrm{N/m^2}$, $R = 2\,\mathrm{cm}$. The thickness D varies monotonically from $D = 2.5$ to 1.7 nm as the load increases from $L = 0$ to 10 mN (adapted from [28.149], American Physical Society)

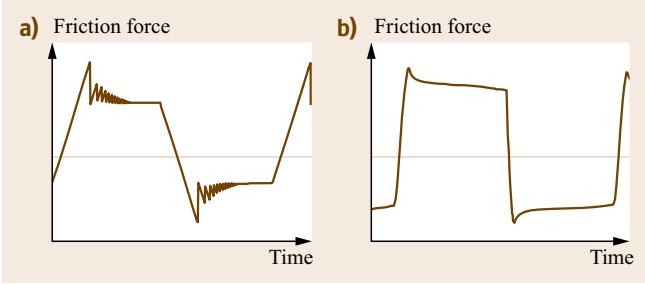


Fig. 28.19a,b Typical friction traces showing how the friction force varies with the sliding time for two symmetric, glassy polymer films under dry conditions. Qualitative features that are common to both polystyrene and polyvinyl benzyl chloride: **(a)** Decaying stick–slip motion is observed until smooth sliding is attained if the motion continues for a sufficiently long distance. **(b)** Smooth sliding observed at sufficiently high speeds. Similar observations have been made (after [28.175]) in measurements on polymethyl methacrylate (after [28.134], with permission, American Association for the Advancement of Science)

low friction of ice. Clearly, a single monolayer of water can be a very good lubricant – much better than most other monomolecular liquid films – for reasons that will be discussed in Sect. 28.9. A linear dependence of F on L has also been observed for mica surfaces separated by certain hydrocarbon liquids [28.141, 149]. Figure 28.18b shows the kinetic friction forces measured at a high velocity across thin films of squalane, a branched hydrocarbon liquid ($C_{30}H_{62}$), which is a model for lubricating oils. Very low adhesive forces are measured between mica surfaces across this liquid [28.149] and the film thickness decreases monotonically with load. The friction force at a given load was found to be velocity dependent, whereas the contact area was not [28.149].

Dry polymer layers (Fig. 28.19) typically show a high initial static friction (stiction) as sliding commences from rest in adhesive contact. The development of the friction force after a change in sliding direction, a gradual transition from stick–slip to smooth sliding, is shown in Fig. 28.19. As shown in Fig. 28.11b,c, the adhesion hysteresis for polystyrene surfaces can be increased by irradiation to induce scission of chains, and it has been found that the steady-state friction force (kinetic friction) shows a similar increase with irradiation time [28.134].

28.7.4 Transition from Interfacial to Normal Friction with Wear

Frictional damage can have many causes, such as adhesive tearing at high loads or overheating at high sliding speeds. Once damage occurs, there is a transition

from interfacial to normal or load-controlled friction as the surfaces become forced apart by the torn-out asperities (wear particles). For low loads, the friction changes from obeying $F = S_c A$ to obeying Amontons' law, $F = \mu L$, as shown in Figs. 28.15 and 28.16, and sliding now proceeds smoothly with the surfaces separated by a 10–100 nm forest of wear debris (in this case, mica flakes). The wear particles keep the surfaces apart over an area that is much greater than their size, so that even one submicroscopic particle or asperity can cause a significant reduction in the area of contact and, therefore, in the friction [28.169]. For this type of frictional sliding, one can no longer talk of the molecular contact area of the two surfaces, although the macroscopic or apparent area is still a useful parameter.

One remarkable feature of the transition from interfacial to normal friction of brittle surfaces is that, while the strength of interfacial friction, as reflected in the values of S_c, is very dependent on the type of surface and on the liquid film between the surfaces, this is not the case once the transition to normal friction has occurred. At the onset of damage, the material properties of the underlying substrates control the friction. In Figs. 28.15, 28.16, and 28.18a the friction for the damaged surfaces is that of any damaged mica-mica system, $\mu \approx 0.3$, *independent of the initial surface coatings or liquid films between the surfaces*. A similar friction coefficient was found for damaged silica surfaces [28.30].

Clearly, the mechanism and factors that determine normal friction are quite different from those that govern interfacial friction (Sects. 28.7.1 and 28.7.2). This effect is not general and may only apply to brittle materials. For example, the friction of ductile surfaces is totally different and involves the continuous plastic deformation of contacting surface asperities during sliding, rather than the rolling of two surfaces on hard wear particles [28.138]. Furthermore, in the case of ductile surfaces, water and other surface-active components do have an effect on the friction coefficients under normal sliding conditions.

28.8 Liquid Lubricated Surfaces

28.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime

Experimentally, it is usually difficult to unambiguously establish which type of sliding mode is occurring, but an empirical criterion, based on the Stribeck curve (Fig. 28.12), is often used as an indicator. This curve shows how the friction force or the coefficient of friction is expected to vary with sliding speed, depending on which type of friction regime is operating. For thick liquid lubricant films whose behavior can be described by bulk (continuum) properties, the friction forces are essentially the hydrodynamic or viscous drag forces. For example, for two plane parallel surfaces of area A separated by a distance D and moving laterally relative to each other with velocity v, if the intervening liquid is Newtonian, i. e., if its viscosity η is independent of the shear rate, the frictional force experienced by the surfaces is given by

$$F = \frac{\eta A v}{D} , \tag{28.42}$$

where the shear rate $\dot{\gamma}$ is defined by

$$\dot{\gamma} = \frac{v}{D} . \tag{28.43}$$

At higher shear rates, two additional effects often come into play. First, certain properties of liquids may change at high $\dot{\gamma}$ values. In particular, the effective viscosity may become non-Newtonian, one form given by

$$\eta \propto \dot{\gamma}^n , \tag{28.44}$$

where $n = 0$ (i. e., $\eta_{\text{eff}} =$ constant) for Newtonian fluids, $n > 0$ for shear-thickening (dilatant) fluids, and $n < 0$ for shear-thinning (pseudoplastic) fluids (the latter become less viscous, i. e., flows more easily, with increasing shear rate). An additional effect on η can arise from the higher local stresses (pressures) experienced by the liquid film as $\dot{\gamma}$ increases. Since the viscosity is generally also sensitive to the pressure (usually increasing with P), this effect also acts to increase η_{eff} and thus the friction force.

A second effect that occurs at high shear rates is surface deformation, arising from the large hydrodynamic forces acting on the sliding surfaces. For example, Fig. 28.20 shows how two surfaces deform elastically when the sliding speed increases to a high value. These deformations alter the hydrodynamic friction forces, and this type of friction is often referred to as elastohydrodynamic lubrication (EHD) or (EHL), as mentioned in Table 28.4.

How thin can a liquid film be before its dynamic, e.g., viscous flow, behavior ceases to be described by bulk properties and continuum models? Concerning the static properties, we have already seen in Sect. 28.4.3

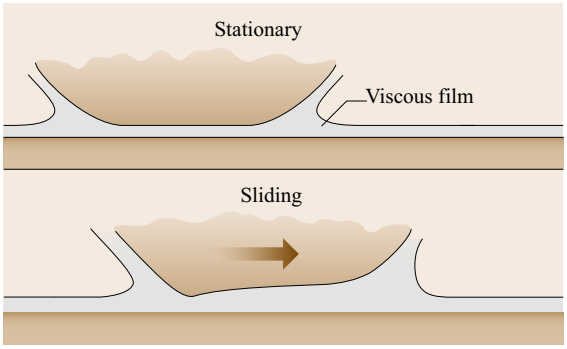


Fig. 28.20 *Top*: Stationary surfaces (one more deformable and one rigid) separated by a thick liquid film. *Bottom*: Elastohydrodynamic deformation of the upper surface during sliding (after [28.5], with permission)

that films composed of simple liquids display continuum behavior down to thicknesses of 4–10 molecular diameters. Similar effects have been found to apply to the dynamic properties, such as the viscosity, of simple liquids in thin films. Previous studies [28.6, 7, 176, 177] show that, for simple liquids including linear chain molecules such as alkanes, the viscosity in thin films is the same, within 10%, as the bulk even for films as thin as 10 molecular diameters (or segment widths). This implies that the shear plane is effectively located within one molecular diameter of the solid–liquid interface, and these conclusions were found to remain valid even at the highest shear rates studied (of $\approx 2\times10^5\,\mathrm{s}^{-1}$). These results appear to be independent of the existence of electrostatic double-layer or hydration forces. For the case of the simple liquid toluene confined between surfaces with adsorbed layers of C_{60} molecules, this type of viscosity measurement has shown that the traditional no-slip assumption for flow at a solid interface does not always hold [28.178]. The C_{60} layer at the mica–toluene interface results in a full-slip boundary, which dramatically lowers the viscous drag or effective viscosity for regular Couette or Poiseuille flow.

With polymeric liquids (polymer melts) such as polydimethylsiloxanes (PDMS) and polybutadienes (PBD), or with polystyrene (PS) adsorbed onto surfaces from solution, the far-field viscosity is again equal to the bulk value, but with the nonslip plane (hydrodynamic layer thickness) being located at $D = 1–2R_g$ away from each surface [28.6, 35], or at $D = L$ or less for polymer brush layers of thickness L per surface [28.7, 179]. In contrast, the same technique was used to show that, for nonadsorbing polymers in solution, there is actually a depletion layer of nearly pure solvent that exists at the surfaces that affects the confined solution flow properties [28.180]. These effects are observed from near contact to surface separations in excess of 200 nm.

Further experiments with surfaces closer than a few molecular diameters ($D < 2–4$ nm for simple liquids, or $D < 2–4R_g$ for polymer fluids) indicate that large deviations occur for thinner films, described below. One important conclusion from these studies is, therefore, that the dynamic properties of simple liquids, including water, near an isolated surface are similar to those of the bulk *liquid already within the first layer of molecules adjacent to the surface*, only changes when another surface approaches the first due to the constraining effects on the liquid molecules.

28.8.2 Friction of Intermediate Thickness Films

For liquid films in the thickness range between 4 and 10 molecular diameters, the properties can be significantly different from those of bulk films. Still, the fluids do not undergo a phase transition into a solid or liquid-crystalline phase. Films of simple liquids have been found to become non-Newtonian in the 2.5–5 nm regime (about 10 molecular diameters, Fig. 28.21) [28.139] while polymer melts are known to become non-Newtonian at much larger film thicknesses, depending on their molecular weight [28.35].

Based on studies of interaction forces and friction of small quasispherical liquid molecules such as cyclohexane between molecularly smooth mica surfaces [28.34, 144], it has been concluded that surface epitaxial effects can cause the liquid film to solidify already at six molecular diameters, resulting in a sudden (discontinuous) jump to high friction at low shear rates. Such dynamic first-order transitions, however, may depend on the shear rate.

A generalized friction map (Fig. 28.21c,d) has been proposed [28.181, 182] that illustrates the changes in η_{eff} from bulk Newtonian behavior ($n = 0$, $\eta_{\mathrm{eff}} = \eta_{\mathrm{bulk}}$) through the transition regime where n reaches a minimum of -1 with decreasing shear rate to the solid-like creep regime at very low $\dot{\gamma}$ where n returns to 0. A number of results from experimental, theoretical, and computer simulation work have shown values of n from $-1/2$ to -1 for this transition regime for a variety of systems and assumptions [28.139, 140, 183–185].

The intermediate regime appears to extend over a narrow range of film thickness, from about 4–10 molecular diameters or polymer radii of gyration. Thinner films begin to adopt boundary or interfacial friction properties (described below, see also Table 28.5). Note that the intermediate regime is actually a very narrow one when defined in terms of film thickness, for ex-

Fig. 28.21a–d Typical rheological behavior of liquid films in the mixed lubrication regime. **(a)** Increase in effective viscosity of dodecane film between two mica surfaces with decreasing film thickness. At distances larger than 4–5 nm, the effective viscosity η_{eff} approaches the bulk value η_{bulk} and does not depend on the shear rate $\dot{\gamma}$ (after [28.139], American Association for the Advancement of Science). **(b)** Non-Newtonian variation of η_{eff} with shear rate of a 2.7 nm-thick dodecane film at a net normal pressure of 0.12 MPa and at 28 °C. The effective viscosity decays as a power law, as in (28.44). In this example, $n = 0$ at the lowest $\dot{\gamma}$ and changes to $n = -2/3$ and -1 at higher $\dot{\gamma}$. For films of bulk thickness, dodecane is a low-viscosity Newtonian fluid ($n = 0$). **(c)** Proposed general friction map of effective viscosity η_{eff} (arbitrary units) as a function of effective shear rate $\dot{\gamma}$ (arbitrary units) on logarithmic scales. Three main classes of behavior emerge: (i) Thick films: EHD sliding. At $L = 0$, approximating bulk conditions, η_{eff} is independent of shear rate except when shear thinning might occur at sufficiently large $\dot{\gamma}$. (ii) Boundary layer films, intermediate regime. A Newtonian regime is again observed ($\eta_{\text{eff}} =$ constant, $n = 0$ in (28.44)) at low loads and low shear rates, but η_{eff} is much higher than the bulk value. As the shear rate $\dot{\gamma}$ increases beyond $\dot{\gamma}_{\text{min}}$, the effective viscosity starts to drop with a power-law dependence on the shear rate **(b)** with n in the range $-1/2$ to -1 most commonly observed. As the shear rate $\dot{\gamma}$ increases still more, beyond $\dot{\gamma}_{\text{max}}$, a second Newtonian plateau is encountered. (iii) Boundary layer films, high load. The η_{eff} continues to grow with load and to be Newtonian provided that the shear rate is sufficiently low. Transition to sliding at high velocity is discontinuous ($n < -1$) and usually of the stick–slip variety. **(d)** Proposed friction map of friction force as a function of sliding velocity in various tribological regimes. With increasing load, Newtonian flow in the EHD regimes crosses into the boundary regime of lubrication. Note that even EHD lubrication changes, at the highest velocities, to limiting shear stress response. At the highest loads (L) and smallest film thickness (D), the friction force goes through a maximum (the static friction F_s) followed by a regime where the friction coefficient (μ) is roughly constant with increasing velocity (meaning that the kinetic friction, F_k, is roughly constant). Non-Newtonian shear thinning is observed at somewhat smaller load and larger film thickness; the friction force passes through a maximum at the point where De = 1. De – the Deborah number – is the point at which the applied shear rate exceeds the natural relaxation time of the boundary layer film. The velocity axis from 10^{-10} to 10^{10} (arbitrary units) indicates a large span. (Panels **(b–d)** after [28.181, 182], with permission from Elsevier Science) ►

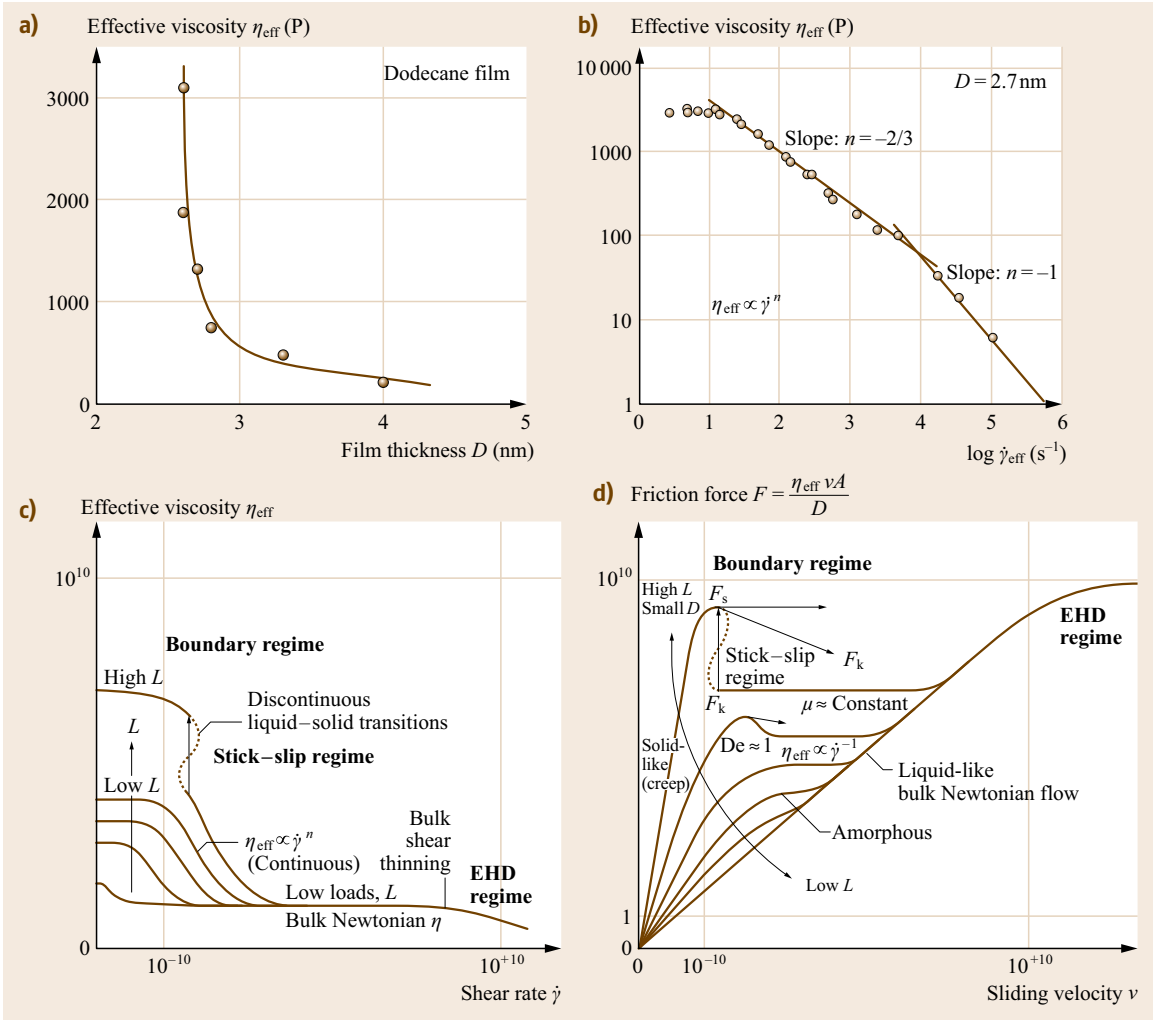
ample varying from about $D = 2$–4 nm for hexadecane films [28.139].

A fluid's effective viscosity η_{eff} in the intermediate regime is usually higher than in the bulk, but η_{eff} usually decreases with increasing sliding velocity, v (known as shear thinning). When two surfaces slide in the intermediate regime, the motion tends to thicken the film (dilatancy). This sends the system into the bulk EHD regime where, as indicated by (28.42), the friction force now increases with velocity. This initial decrease, followed by an increase, in the frictional forces of many lubricant systems is the basis for the empirical Stribeck

Table 28.5 Effect of molecular shape and short-range forces on tribological properties for molecularly thin liquid films between two shearing mica surfaces at 20 °C

Liquid	Short-range force	Type of friction	Friction coefficient	Bulk liquid viscosity (cP)
Organic (water-free) Cyclohexane	Oscillatory	Quantized stick–slip	$\gg 1$	0.6
OMCTS[a]	Oscillatory	Quantized stick–slip	$\gg 1$	2.3
Octane	Oscillatory	Quantized stick–slip	1.5	0.5
Tetradecane	Oscillatory ↔ smooth	stick–slip ↔ smooth	1.0	2.3
Octadecane (branched)	Oscillatory ↔ smooth	stick–slip ↔ smooth	0.3	5.5
PDMS[a] ($M = 3700\,\text{g mol}^{-1}$, melt)	Oscillatory ↔ smooth	Smooth	0.4	50
PBD[a] ($M = 3500\,\text{g mol}^{-1}$, branched)	Smooth	Smooth	0.03	800
Water				
Water (KCl solution)	Smooth	Smooth	0.01–0.03	1.0

[a] OMCTS: Octamethylcyclotetrasiloxane, [b] PDMS: Polydimethylsiloxane, [c] PBD: Polybutadiene



curve of Fig. 28.12. In the transition from bulk to boundary behavior there is first a quantitative change in the material properties (viscosity and elasticity), which can be continuous, to discontinuous qualitative changes that result in yield stresses and nonliquid-like behavior.

28.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology

When a liquid is confined between two surfaces or within any narrow space whose dimensions are less than 4–10 molecular diameters, both the static (equilibrium) and dynamic properties of the liquid, such as its compressibility and viscosity, can no longer be described even qualitatively in terms of the bulk properties. The molecules confined within such molecularly thin films become ordered into layers (out-of-plane ordering), and within each layer they can also have lateral order (in-plane ordering). Such films may be thought of as behaving more like a liquid crystal or a solid than a liquid, and the structuring of molecules in thin films and the oscillatory forces it gives rise to (Sect. 28.4.3) are now reasonably well understood, both experimentally and theoretically, at least for simple liquids.

Work has also recently been done on the dynamic, e.g., viscous or shear, forces associated with molecularly thin films. Both experiments [28.28, 34, 129, 141, 144, 145, 187] and theory [28.127, 128, 183, 188] indicate that, even when two surfaces are in steady-state sliding, they still prefer to remain in one of their stable potential-energy minima, i. e., a sheared film of liquid can retain its basic layered structure. Thus even during motion the film does not become totally liquid-like. Indeed, if there is some in-plane ordering within a film, it will exhibit a yield point before it begins to flow. Such films can therefore sustain a finite shear stress,

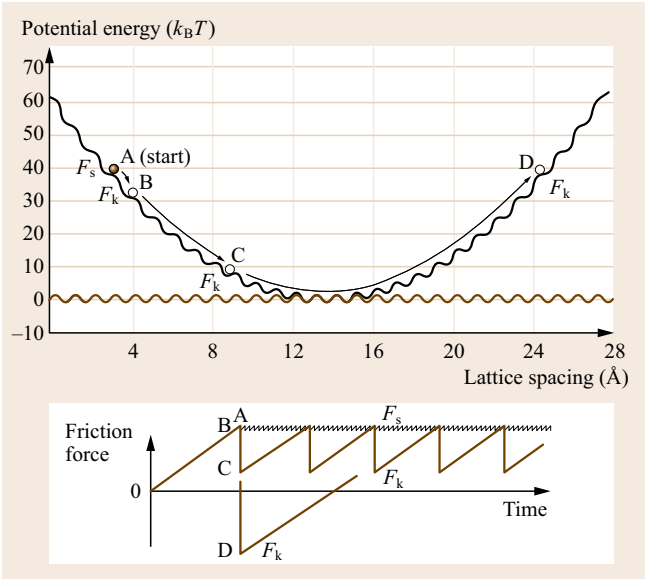


Fig. 28.22 Simple schematic illustration of the most common molecular mechanism leading from smooth to stick–slip sliding in terms of the efficiency of the energy transfer from mechanical to kinetic to phonons. The potential energy of the corrugated surface lattice is shown by the *horizontal sine wave*. Let the depth of each minimum be ε, which is typically $> k_BT$. At equilibrium, a molecule will *sit* at one of these minima. When the molecule is connected to a horizontal spring, a smooth parabolic curve must be added to the horizontal curve. If this spring is now pushed or pulled laterally at a constant velocity v, the sine curve will move like a wave along the parabola carrying the molecule up with it (towards point A). When the point of inflection at A is reached the molecule will drop and acquire a kinetic energy greater than ε even before it reaches the next lattice site. This energy can be *released* at the next lattice site (i. e., on the first collision), in which case the processes will now be repeated – each time the molecule reaches point A it will fall to point B. This type of motion will give rise to periodic changes in temperature at the interface, as predicted by computer simulations (after [28.186]). The stick–slip here will have a magnitude of the lattice dimension and the measured macro- and microscopic friction forces will be smooth and independent of v. If the energy dissipation (or *transfer*) mechanism is less than 100% efficient on each collision, the molecule will move further before it stops. In this case the stick–slip amplitude can be large (point C), and the kinetic friction F_k can even be negative in the case of an overshoot (point D) (after [28.151], with permission, American Chemical Society)

in addition to a finite normal stress. The value of the yield stress depends on the number of layers comprising the film and represents another quantized property of molecularly thin films [28.127].

The dynamic properties of a liquid film undergoing shear are very complex. Depending on whether the film is more liquid-like or solid-like, the motion will be smooth or of the stick–slip type illustrated schematically in Fig. 28.22. During sliding, transitions can occur between n layers and $(n-1)$ or $(n+1)$ layers. The details of the motion depend critically on the externally applied load, the temperature, the sliding velocity, the twist angle between the two surface lattices, and the sliding direction relative to the lattices.

Smooth and Stick–Slip Sliding

Recent advances in friction-measuring techniques have enabled the interfacial friction of molecularly thin films to be measured with great accuracy. Some of these advances have involved the surface forces apparatus technique [28.28, 32–35, 140, 141, 144, 145, 149, 150, 158, 159, 169, 174, 187, 189]. In addition, computer simulations [28.57, 127, 128, 151, 188, 190] have become sufficiently sophisticated to enable fairly complex tribological systems to be studied. All these advances are necessary if one is to probe such subtle effects as smooth or stick–slip friction, transient and memory effects, and ultralow friction mechanisms at the molecular level.

stick–slip occurs between a static state with high friction and a low-friction kinetic state, and a transition from this sliding regime to smooth sliding can be induced by an increase in velocity. Experimental data on various systems showing this behavior are shown below. Recent studies on adhesive systems have revealed the possibility of other dynamic responses such as inverted stick–slip between two kinetic states of higher and lower friction and with a transition from smooth sliding to stick–slip with increasing velocity [28.191]. Similar friction responses have also been seen in computer simulations [28.192].

With the added insights provided by computer simulations, a number of distinct molecular processes have been identified during smooth and stick–slip sliding in model systems for the more familiar static-to-kinetic stick–slip and transition from stick–slip to smooth sliding. These are shown schematically in Fig. 28.23 for the case of spherical liquid molecules between two solid crystalline surfaces. The following regimes may be identified:

- *Surfaces at rest* (Fig. 28.23a): Even with no externally applied load, solvent–surface epitaxial interactions can cause the liquid molecules in the film to attain a solid-like state. Thus at rest the surfaces are stuck to each other through the film.
- *Sticking regime (frozen, solid-like film)* (Fig. 28.23b): A progressively increasing lateral shear stress is applied. The solid-like film responds elastically with a small lateral displace-

ment and a small increase or dilatancy in film thickness (of less than a lattice spacing or molecular dimension σ). In this regime the film retains its frozen, solid-like state: all the strains are elastic and reversible, and the surfaces remain effectively stuck to each other. However, slow creep may occur over long time periods.

- *Slipping and sliding regimes (molten, liquid-like film)* (Fig. 28.23c–c″): When the applied shear stress or force has reached a certain critical value, the static friction force F_s, the film suddenly melts (known as shear melting) or rearranges to allow for wall slip or slip within the film to occur at which point the two surfaces begin to slip rapidly past each other. If the applied stress is kept at a high value, the upper surface will continue to slide indefinitely.
- *Refreezing regime (resolidification of film)* (Fig. 28.23d): In many practical cases, the rapid slip of the upper surface relieves some of the applied force, which eventually falls below another critical value, the kinetic friction force F_k, at which point the film resolidifies and the whole stick–slip cycle is repeated. On the other hand, if the slip rate is smaller than the rate at which the external stress is applied, the surfaces will continue to slide smoothly in the kinetic state and there will be no more stick–slip.

Experiments on the friction of two molecularly smooth mica surfaces separated by three molecular layers of the liquid octamethylcyclotetrasiloxane (OMCTS, Fig. 28.24) show how the friction increases to higher values in a quantized way when the number of layers falls from $n = 3$ to $n = 2$ and then to $n = 1$. Computer simulations for simple spherical molecules [28.128] further indicate that during slip the film thickness is roughly 15% higher than at rest (i. e., the film density falls), and that the order parameter within the film drops from 0.85 to about 0.25. Such dilatancy has been investigated both experimentally [28.158] and in further computer simulations [28.190]. The changes in thickness and in the order parameter are consistent with a disorganized state for the whole film during the slip [28.188], as illustrated schematically in Fig. 28.23c. At this stage, we can only speculate on other possible configurations of molecules in the slipping and sliding regimes. This probably depends on the shapes of the molecules (e.g., whether they are spherical or linear or branched), on the atomic structure of the surfaces, on the sliding velocity, etc. [28.193]. Figure 28.23c–c″ shows three possible sliding modes wherein the shearing film either totally melts, or where the molecules retain their layered structure and where slip occurs between two or more layers.

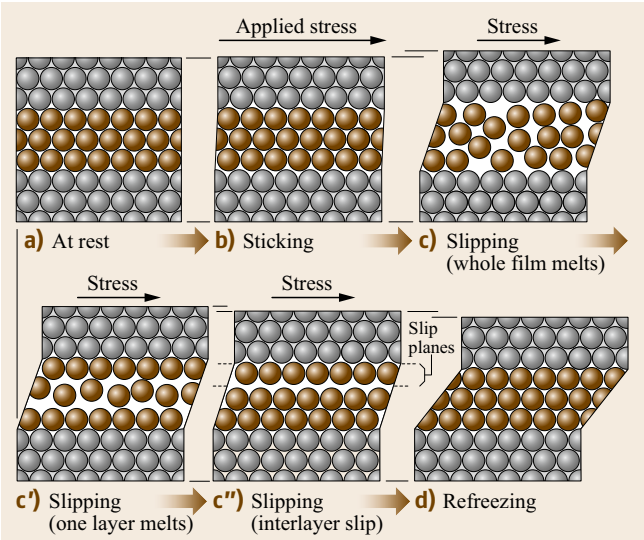


Fig. 28.23a–d Idealized schematic illustration of molecular rearrangements occurring in a molecularly thin film of spherical or simple chain molecules between two solid surfaces during shear. **(a)** At rest, and **(b)** when sticking. Depending on the system, a number of different molecular configurations within the film are possible during slipping and sliding, shown here as stages **(c)** – total disorder as the whole film melts; **(c′)** – partial disorder; and **(c″)** – order persists even during sliding with slip occurring at a single slip plane either within the film or at the walls. A dilation is predicted in the direction normal to the surfaces. **(d)** Refreezing (after [28.142], with permission)

Other sliding modes, for example involving the movement of dislocations or disclinations, are also possible, and it is unlikely that one single mechanism applies in all cases.

Both friction and adhesion hysteresis vary nonlinearly with temperature, often peaking at some particular temperature T_0. The temperature dependence of these forces can, therefore, be represented on a friction phase diagram such as the one shown in Fig. 28.25. Experiments have shown that T_0, and the whole bell-shaped curve, are shifted along the temperature axis (as well as in the vertical direction) in a systematic way when the load, sliding velocity, etc., are varied. These shifts also appear to be highly correlated with one another.

Such effects are also commonly observed in other energy-dissipating phenomena such as polymer viscoelasticity [28.194]. A possible molecular process underlying the energy dissipation of chain molecules during boundary-layer sliding is illustrated in Fig. 28.26, which shows the three main dynamic phase states of boundary monolayers.

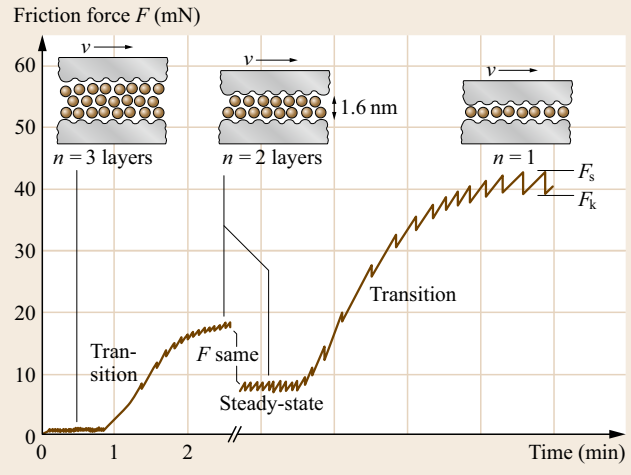


Fig. 28.24 Measured change in friction during interlayer transitions of the silicone liquid OMCTS, an inert liquid whose quasispherical molecules have a diameter of 0.8 nm. In this system, the shear stress $S_c = F/A$ was found to be constant as long as the number of layers, n, remained constant. Qualitatively similar results have been obtained with other quasispherical molecules such as cyclohexane (after [28.187]). The shear stresses are only weakly dependent on the sliding velocity v. However, for sliding velocities above some critical value, v_c, the stick–slip disappears and sliding proceeds smoothly at the kinetic value (after [28.141], with permission)

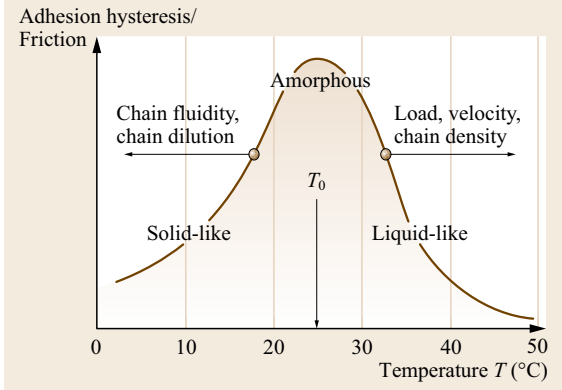


Fig. 28.25 Schematic friction phase diagram representing the trends observed in the boundary friction of a variety of different surfactant monolayers. The characteristic *bell-shaped curve* also correlates with the adhesion energy hysteresis of the monolayers. The *arrows* indicate the direction in which the whole curve is dragged when the load, velocity, etc. is increased (after [28.155], with permission)

Abrupt versus Continuous Transitions Between Smooth and Stick–Slip Sliding

An understanding of stick–slip is of great practical importance in tribology [28.195], since these spikes are the major cause of damage and wear of moving parts. stick–slip motion is a very common phenomenon and is also the cause of sound generation (the sound of a violin string, a squeaking door, or the chatter of machinery), sensory perception (taste texture and feel), earthquakes, granular flow, nonuniform fluid flow such as the spurting flow of polymeric liquids, etc. In the previous section, the stick–slip motion arising from freezing-melting transitions in thin interfacial films was described. There are other mechanisms that can give rise to stick–slip friction, which will now be considered.

Rough Surfaces or Surface Topology Model

Rapid slips can occur whenever an asperity on one surface goes over the top of an asperity on the other surface. The extent of the slip will depend on asperity heights and slopes, on the speed of sliding, and on the elastic compliance of the surfaces and the moving stage. As in all cases of stick–slip motion, the driving velocity v may be constant, but the resulting motion at the surfaces v_0 will display large slips. This type of stick–slip [28.195] will not be of much concern here since it is essentially a noise-type fluctuation, resulting from surface imperfections rather than from the intrinsic interaction between two surfaces.

Distance-Dependent or Creep Model

Another theory of stick–slip, observed in solid-on-solid sliding, is one that involves a characteristic distance, but also a characteristic time τ_s, this being the characteristic time required for two asperities to increase their adhesion strength after coming into contact. Originally proposed by *Rabinowicz* [28.195, 196], this model suggests that two rough macroscopic surfaces adhere through their microscopic asperities of a characteristic length. During shearing, each surface must first creep this distance (the size of the contacting junctions) after which the surfaces continue to slide, but with a lower (kinetic) friction force than the original (static) value. The reason for the decrease in the friction force is that even though, on average, new asperity junctions should form as rapidly as the old ones break, the time-dependent adhesion and friction of the new ones will be lower than the old ones.

The friction force, therefore, remains high during the creep stage of the slip. However, once the surfaces have moved the characteristic distance, the friction rapidly drops to the kinetic value. Any system where the kinetic friction is less than the static force (or one that has a negative slope over some part of its curve of F_0 versus v_0) will exhibit regular stick–slip sliding motion for certain values of K, m, and driving velocity v.

This type of friction has been observed in a variety of dry (unlubricated) systems such as paper-on-paper [28.197] and steel-on-steel [28.196, 198]. This model is also used extensively in geologic systems to analyze rock-on-rock sliding [28.199]. While originally described for adhering macroscopic asperity junctions, the distance-dependent model may also apply to molecularly smooth surfaces. For example, for polymer lubricant films, the characteristic length would now be the chain-chain entanglement length, which could be much larger in a confined geometry than in the bulk.

Velocity-Dependent Friction Model

In contrast to the two friction models mentioned above, which apply mainly to unlubricated, solid-on-solid contacts, the stick–slip of surfaces with thin liquid films between them is better described by other mechanisms. The velocity-dependent friction model is the most studied mechanism of stick–slip and, until recently, was considered to be the only cause of intrinsic stick–slip. If a friction force decreases with increasing sliding velocity, as occurs with boundary films exhibiting shear thinning, the force (F_s) needed to initiate motion will be higher than the force (F_k) needed to maintain motion.

A decreasing intrinsic friction force F_0 with sliding velocity v_0 results in the sliding surface or stage moving in a periodic fashion, where during each cycle rapid acceleration is followed by rapid deceleration. As long as the drive continues to move at a fixed velocity v, the surfaces will continue to move in a periodic fashion punctuated by abrupt stops and starts whose frequency and amplitude depend not only on the function $F_0(v_0)$, but also on the stiffness K and mass m of the moving stage, and on the starting conditions at $t = 0$.

More precisely, the motion of the sliding surface or stage can be determined by solving the differential equation

$$m\ddot{x} = (F_0 - F) = F_0 - (x_0 - x)K$$
$$\text{or} \quad m\ddot{x} + (x_0 - x)K - F_0 = 0\,, \tag{28.45}$$

where $F_0 = F_0(x_0,\ v_0,\ t)$ is the intrinsic or real friction force at the shearing surfaces, F is the force on the spring (the externally applied or measured force), and $(F_0 - F)$ is the force on the stage. To solve (28.45) fully, one must also know the initial (starting) conditions at $t = 0$, and the driving or steady-state conditions at finite t. Commonly, the driving condition is $x = 0$ for $t < 0$ and $x = vt$ for $t > 0$, where $v =$ constant. In other systems, the appropriate driving condition may be $F =$ constant.

Various, mainly phenomenological, forms for $F_0 = F_0(x_0, v_0, t)$ have been proposed to explain various kinds of stick–slip phenomena. These models generally assume a particular functional form for the friction as a function of velocity only, $F_0 = F_0(v_0)$, and they may also contain a number of mechanically coupled elements comprising the stage [28.200, 201]. One version is a two-state model characterized by two friction forces, F_s and F_k, which is a simplified version of the phase transitions model described below. More complicated versions can have a rich F–v spectrum [28.202]. Unless the experimental data are very detailed and extensive, these models cannot generally distinguish between different types of mechanisms. Neither do they address the basic question of the origin of the friction force, since this is assumed to begin with.

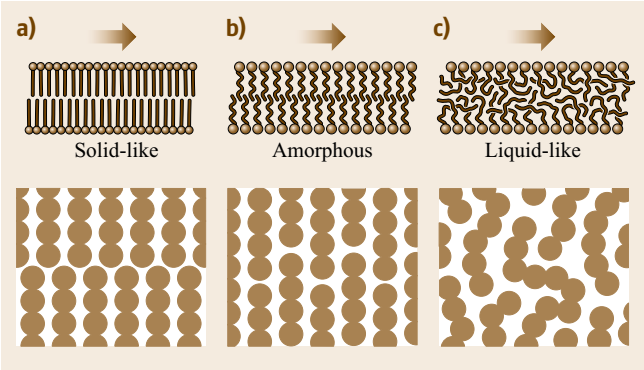


Fig. 28.26a–c Different dynamic phase states of boundary monolayers during adhesive contact and/or frictional sliding. Solid-like **(a)**, amorphous **(b)** and liquid-like monolayers **(c)** exhibit low adhesion hysteresis and friction. Increasing the temperature generally shifts a system from the left to the right. Changing the load, sliding velocity, or other experimental conditions can also change the dynamic phase state of surface layers, as shown in Fig. 28.25 (after [28.26, 155], with permission)

Experimental data has been used to calculate the friction force as a function of velocity within an individual stick–slip cycle [28.203]. For a macroscopic granular material confined between solid surfaces, the data show a velocity-weakening friction force during the first half of the slip. However, the data also show a hysteresis loop in the friction-velocity plot, with a different behavior in the deceleration half of the slip phase. Similar results were observed for a 1–2 nm liquid lubricant film between mica surfaces [28.193]. These results indicate that a purely velocity-dependent friction law is insufficient to describe such systems, and an additional element such as the state of the confined material must be considered.

Phase Transitions Model

In molecular dynamics computer simulations it has been found that thin interfacial films undergo first-

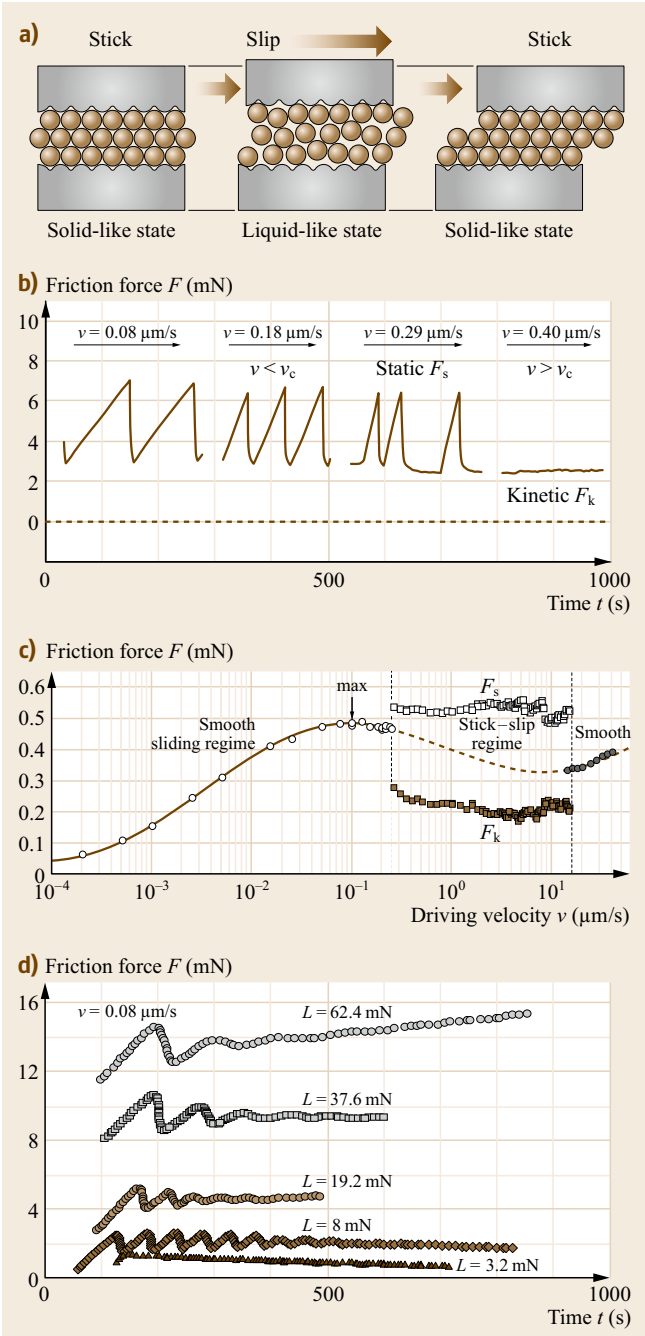


Fig. 28.27 **(a)** *Phase transitions* model of stick–slip where a thin liquid film alternately freezes and melts as it shears, shown here for 22 spherical molecules confined between two solid crystalline surfaces. In contrast to the velocity-dependent friction model, the intrinsic friction force is assumed to change abruptly (at the transitions), rather than smoothly or continuously. The resulting stick–slip is also different, for example the peaks are sharper and the stick–slip disappears above some critical velocity v_c. Note that while the slip displacement is here shown to be only two lattice spacings, in most practical situations it is much larger, and that freezing and melting transitions at surfaces or in thin films may not be the same as freezing or melting transitions between the bulk solid and liquid phases. **(b)** Exact reproduction of a chart-recorder trace of the friction force for hexadecane between two untreated mica surfaces at increasing sliding velocity v, plotted as a function of time. In general, with increasing sliding speed, the stick–slip spikes increase in frequency and decrease in magnitude. As the critical sliding velocity v_c is approached, the spikes become erratic, eventually disappearing altogether at $v = v_c$. At higher sliding velocities the sliding continues smoothly in the kinetic state. Such friction traces are fairly typical for simple liquid lubricants and dry boundary lubricant systems (Fig. 28.28a) and may be referred to as the *conventional* type of static-kinetic friction (in contrast to panel **(c)**). Experimental conditions: contact area $A = 4\times10^{-9}$ m^2, load $L = 10$ mN, film thickness $D = 0.4$–0.8 nm, $v = 0.08$–$0.4\,\mu\mathrm{m\,s^{-1}}$, $v_c \approx 0.3\,\mu\mathrm{m\,s^{-1}}$, atmosphere = dry N_2 gas, $T = 18\,^\circ$C. (**(a,b)** after [28.204] with permission, American Chemical Society). **(c)** Transition from smooth sliding to *inverted* stick–slip and to a second smooth-sliding regime with increasing driving velocity during shear of two adsorbed surfactant monolayers in aqueous solution at a load of $L = 4.5$ mN and $T = 20\,^\circ$C. The smooth sliding (*open circles*) to inverted stick–slip (*squares*) transition occurs at $v_c \approx 0.3\,\mu$m/s. Prior to the transition, the kinetic stress levels off after a logarithmic dependence on velocity. The quasismooth regime persists up to the transition at v_c. At high driving velocities (*filled circles*), a new transition to a smooth sliding regime is observed between 14 and 17 μm/s (after [28.191] with permission). **(d)** Friction response of a thin squalane (a branched hydrocarbon) film at different loads and a constant sliding velocity $v = 0.08\,\mu\mathrm{m\,s^{-1}}$, slightly above the critical velocity for this system at low loads. Initially, with increasing load, the stick–slip amplitude and the mean friction force decrease with sliding time or sliding distance. However, at high loads or pressures, the mean friction force increases with time, and the stick–slip takes on a more symmetrical, sinusoidal shape. At all loads investigated, the stick–slip component gradually decayed as the friction proceeded towards smooth sliding (after [28.149] with permission, American Physical Society) ◀

order phase transitions between solid-like and liquid-like states during sliding [28.128, 205], as illustrated in Fig. 28.27. It has been suggested that this is responsible for the observed stick–slip behavior of simple isotropic liquids between two solid crystalline surfaces. With this interpretation, stick–slip is seen to arise because of the abrupt change in the flow properties of a film at a transition [28.142, 143, 183], rather than the gradual or continuous change, as occurs in the previous example. Other computer simulations indicate that it is the stick–slip that induces a disorder (*shear melting*) in the film, not the other way around [28.206].

An interpretation of the well-known phenomenon of decreasing coefficient of friction with increasing sliding velocity has been proposed based on computer simulation [28.128]. This postulates that it is not the friction that changes with sliding speed v, but rather the time various parts of the system spend in the sticking and sliding modes. In other words, at any instant during sliding, the friction at any local region is always F_s or F_k, corresponding to the static or kinetic values. The measured frictional force, however, is the sum of all these discrete values averaged over the whole contact area. Since as v increases each local region spends more time in the sliding regime (F_k) and less in the sticking regime (F_s), the overall friction coefficient falls. One may note that this interpretation reverses the traditional way that stick–slip has been explained. Rather than invoking a decreasing friction with velocity to explain stick–slip, it is now the more fundamental stick–slip phenomenon that is producing the apparent decrease in the friction force with increasing sliding velocity. This approach has been studied analytically [28.207], with a comprehensive rate- and state-dependent friction force law. This model includes an analytic description of the freezing-melting transitions of a film, resulting in a friction force that is a function of sliding velocity in a natural way. This model predicts a full range of stick–slip behavior observed experimentally.

An example of the rate- and state-dependent model is observed when shearing thin films of OMCTS between mica surfaces [28.208, 209]. In this case, the static friction between the surfaces is dependent on the time that the surfaces are at rest with respect to each other, while the intrinsic kinetic friction $F_{k,0}$ is relatively constant over the range of velocities. At slow driving velocities, the system responds with stick–slip sliding with the surfaces reaching maximum static friction before each slip event, and the amplitude of the stick–slip, $F_s - F_k$, is relatively constant. As the driving velocity increases, the static friction decreases as the time at relative rest becomes shorter with respect to the characteristic time of the lubricant film. As the static friction decreases with increasing drive velocity, it eventually equals the intrinsic kinetic friction $F_{k,0}$, which defines the critical velocity v_c, above which the surfaces slide smoothly without the jerky stick–slip motion.

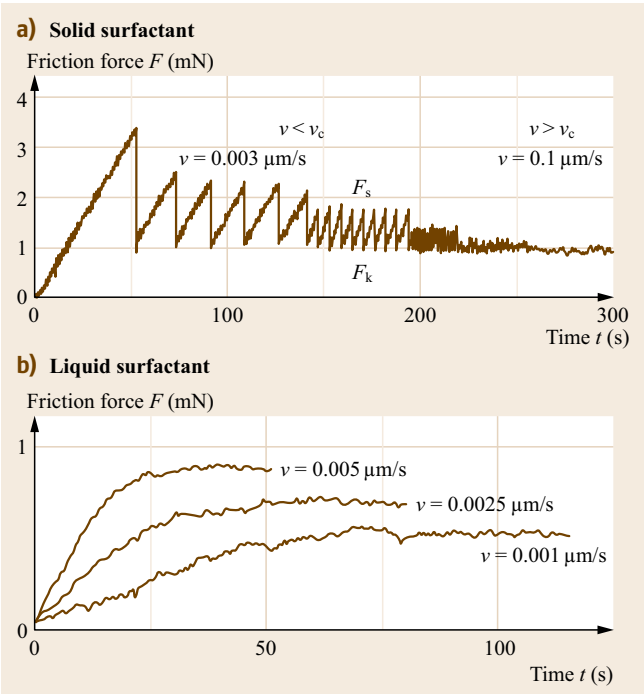


Fig. 28.28 **(a)** Exact reproduction of chart-recorder trace for the friction of closely packed surfactant monolayers (l-α-dimirystoylphosphatidyl-ethanolamine, DMPE) on mica (dry boundary friction) showing qualitatively similar behavior to that obtained with a liquid hexadecane film (Fig. 28.27b). In this case, $L = 0$, $v_c \approx 0.1\,\mu m\,s^{-1}$, atmosphere = dry N_2 gas, $T = 25\,°C$. **(b)** Sliding typical of liquid-like monolayers, here shown for calcium alkylbenzene sulfonate in dry N_2 gas at $T = 25\,°C$ and $L = 0$ (after [28.133], with permission, American Chemical Society)

The above classifications of stick–slip are not exclusive, and molecular mechanisms of real systems may exhibit aspects of different models simultaneously. They do, however, provide a convenient classification of existing models and indicate which experimental parameters should be varied to test the different models.

Critical Velocity for Stick–Slip

For any given set of conditions, stick–slip disappears above some critical sliding velocity v_c (Fig. 28.28), above which the motion continues smoothly in the liquid-like or kinetic state [28.133, 149, 191, 193, 204]. The critical velocity is well described by two simple equations. Both are based on the phase transition model, and both include some parameter associated with the inertia of the measuring instrument. The first equation is

based on both experiments and simple theoretical modeling [28.204]

$$v_c = \frac{(F_s - F_k)}{5k\tau_0} , \qquad (28.46)$$

where τ_0 is the characteristic nucleation time or freezing time of the film. For example, inserting the following typically measured values for $a \approx 1$ nm thick hexadecane film between mica $(F_s - F_k) \approx 5$ mN, spring constant $k \approx 500\,\mathrm{N\,m^{-1}}$, and nucleation time [28.204] $\tau_0 \approx 5$ s, we obtain $v_c \approx 0.4\,\mu\mathrm{m\,s^{-1}}$, which is close to typically measured values (Fig. 28.27b).

The second equation is based on computer simulations [28.205]

$$v_c \approx 0.1\sqrt{\frac{F_s\sigma}{m}} , \qquad (28.47)$$

where σ is a molecular dimension and m is the mass of the stage. Again, inserting typical experimental values into this equation, that is, $m \approx 20$ g, $\sigma \approx 0.5$ nm, and $(F_s - F_k) \approx 5$ mN as before, we obtain $v_c \approx 0.3\,\mu\mathrm{m\,s^{-1}}$, which is also close to measured values.

Stick–slip also disappears above some critical temperature T_c, which is not the same as the melting temperature of the bulk fluid [28.149]. Certain correlations have been found between v_c and T_c and between various other tribological parameters that appear to be consistent with the principle of time-temperature superposition (Sect. 28.8.3), similar to that occurring in viscoelastic polymer fluids [28.194, 210, 211].

Work on the coupling between the mechanical resonances of the sliding system and molecular-scale relaxations [28.189, 190] has resulted in a better understanding of a phenomenon previously noted in various engineering applications: the vibration of one of the sliding surfaces perpendicularly to the sliding direction can lead to a significant reduction of the friction. At certain oscillation amplitudes and a frequency higher than the molecular-scale relaxation frequency, stick–slip friction can be eliminated and replaced by an ultralow kinetic-friction state.

28.9 Effects of Nanoscale Texture on Friction

The above scenario is already quite complicated, and yet this is the situation for the simplest type of experimental system. The factors that appear to determine the critical velocity v_c depend on the type of liquid between the surfaces, as well as on the surface lattice structure.

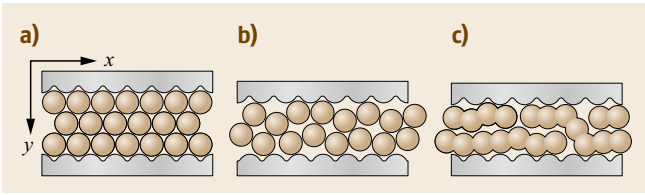


Fig. 28.29a–c Schematic view of an interfacial film composed of spherical molecules under a compressive pressure between two solid crystalline surfaces. **(a)** If the two surface lattices are free to move in the x–y–z directions, so as to attain the lowest energy state, they could equilibrate at values of x, y, and z, which induce the trapped molecules to become *epitaxially* ordered into a *solid-like* film. **(b)** Similar view of trapped molecules between two solid surfaces that are not free to adjust their positions, for example as occurs in capillary pores or in brittle cracks. **(c)** Similar to **(a)**, but with chain molecules replacing the spherical molecules in the gap. These may not be able to order as easily as do spherical molecules even if x, y, and z can adjust, resulting in a situation that is more akin to **(b)** (after [28.204] with permission, American Chemical Society)

28.9.1 Role of the Shape of Confined Molecules

Small spherical molecules such as cyclohexane and OMCTS have been found to have very high v_c, which indicates that these molecules can rearrange relatively quickly in thin films. Chain molecules and especially branched-chain molecules have been found to have much lower v_c, which is to be expected, and such liquids tend to slide smoothly or with erratic stick–slip [28.193], rather than in a stick–slip fashion (Table 28.5). With highly asymmetric molecules, such as multiply branched isoparaffins and polymer melts, no regular spikes or stick–slip behavior occurs at any speed, since these molecules can never order themselves sufficiently to solidify. Examples of such liquids are some perfluoropolyethers and PDMS.

Computer simulations [28.55, 57, 151] of the structure, interaction forces, and tribological behavior of chain molecules between two shearing surfaces indicate that both linear and singly or doubly branched-chain molecules order between two flat surfaces by aligning into discrete layers parallel to the surfaces. However, in the case of the weakly branched molecules, the expected oscillatory forces do not appear because of a complex cancellation of entropic and enthalpic contributions to the interaction free energy, which results in

a monotonically smooth interaction, exhibiting a weak energy minimum rather than the oscillatory force profile that is characteristic of linear molecules. During sliding, however, these molecules can be induced to further align, which can result in a transition from smooth to stick–slip sliding.

Table 28.5 shows the trends observed with some organic and polymeric liquids between smooth mica surfaces. Also listed are the bulk viscosities of the liquids. From the data of Table 28.5 it appears that there is a direct correlation between the shapes of molecules and their coefficients of friction or effectiveness as lubricants (at least at low shear rates). Small spherical or chain molecules have high friction with stick–slip because they can pack into ordered solid-like layers. In contrast, longer chained and irregularly shaped molecules remain in an entangled, disordered, fluid-like state even in very thin films, and these give low friction and smoother sliding. It is probably for this reason that irregularly shaped branched chain molecules are usually better lubricants. It is interesting to note that the friction coefficient generally decreases as the bulk viscosity of the liquids increases. This unexpected trend occurs because the factors that are conducive to low friction are generally conducive to high viscosity. Thus molecules with side groups such as branched alkanes and polymer melts usually have higher bulk viscosities than their linear homologues for obvious reasons. However, in thin films the linear molecules have higher shear stresses, because of their ability to become ordered. The only exception to the above correlations is water, which has been found to exhibit both low viscosity and low friction (Fig. 28.18a, and Sect. 28.7.3). In addition, the presence of water can drastically lower the friction and eliminate the stick–slip of hydrocarbon liquids when the sliding surfaces are hydrophilic.

If an *effective* viscosity, η_{eff}, were to be calculated for the liquids of Table 28.5, the values would be many orders of magnitude higher than those of the bulk liquids. This can be demonstrated by the following simple calculation based on the usual equation for Couette flow (28.42)

$$\eta_{\text{eff}} = \frac{F_{\text{k}} D}{A v} , \qquad (28.48)$$

where F_{k} is the kinetic friction force, D is the film thickness, A the contact area, and v the sliding velocity. Using typical values for experiments with hexadecane [28.204] – $F_{\text{k}} = 5\,\text{mN}$, $D = 1\,\text{nm}$, $A = 3 \times 10^{-9}\,\text{m}^2$, and $v = 1\,\mu\text{m}\,\text{s}^{-1}$ – one gets $\eta_{\text{eff}} \approx 2000\,\text{N}\,\text{s}\,\text{m}^{-2}$, or 20 000 P, which is ≈ 106 times higher than the bulk viscosity, η_{bulk}, of the liquid. It is instructive to consider that this very high effective viscosity nevertheless still produces a low friction force or friction coefficient μ of about 0.25. It is interesting to speculate that, if a 1 nm film were to exhibit bulk viscous behavior, the friction coefficient under the same sliding conditions would be as low as 0.000 001. While such a low value has never been reported for any tribological system, one may consider it a theoretical lower limit that could conceivably be attained under certain experimental conditions.

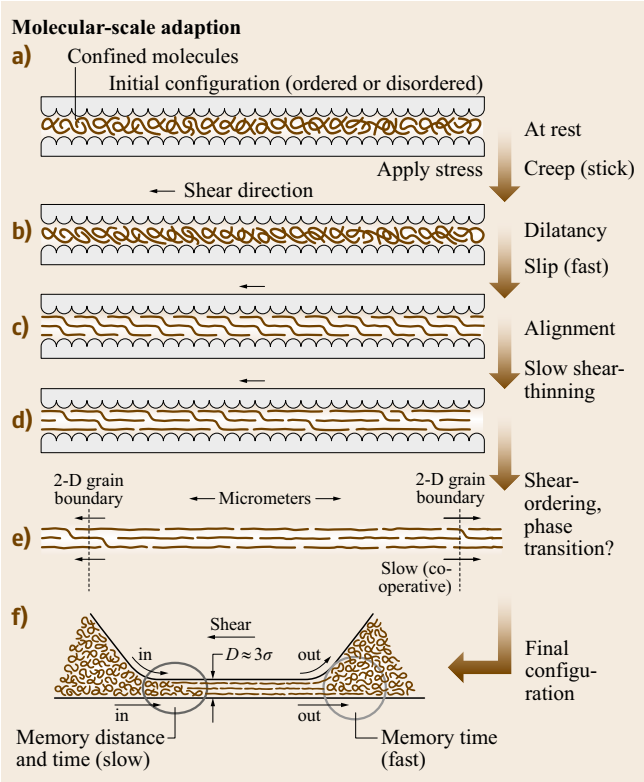


Fig. 28.30a–f Schematic representation of the film under shear. **(a)** The lubricant molecules are just confined, but not oriented in any particular direction. Because of the need to shear, the film dilates **(b)**. The molecules disentangle **(c)** and get oriented in a certain direction related to the shear direction **(d)**. **(e)** Slowly evolving domains grow inside the contact region. These macroscopic domains are responsible for the long relaxation times. **(f)** At the steady state, a continuous gradient of confinement time and molecular order is established in the contact region, which is different for molecules adsorbed on the upper and lower surfaces. Molecules entering into the contact are not oriented or ordered. The required sliding distance to modify their state defines a characteristic distance. Molecules leaving the contact region need some (short) characteristic time to regain their bulk, unconfined configuration (after [28.212], with permission, American Chemical Society)

28.9.2 Effects of Surface Structure

Various studies [28.33, 139, 140, 148–150] have shown that confinement and load generally increase the effective viscosity and/or relaxation times of molecules, suggestive of an increased glassiness or solid-like behavior (Figs. 28.29 and 28.30). This is in marked contrast to studies of liquids in small confining capillaries where the opposite effects have been observed [28.213, 214]. The reason for this is probably because the two modes of confinement are different. In the former case (confinement of molecules between two structured solid surfaces), there is generally little opposition to any lateral or vertical displacement of the two surface lattices relative to each other. This means that the two lattices can shift in the $x-y-z$ planes (Fig. 28.29a) to accommodate the trapped molecules in the most crystallographically commensurate or epitaxial way, which would favor an ordered, solid-like state. In contrast, the walls of capillaries are rigid and cannot easily move or adjust to accommodate the confined molecules (Fig. 28.29b), which will therefore be forced into a more disordered, liquid-like state (unless the capillary wall geometry and lattice are exactly commensurate with the liquid molecules, as occurs in certain zeolites [28.214]).

Experiments have demonstrated the effects of surface lattice mismatch on the friction between surfaces [28.129, 130, 215]. Similar to the effects of lattice mismatch on adhesion (Fig. 28.10), the static friction of a confined liquid film is maximum when the lattices of the confining surfaces are aligned. For OMCTS confined between mica surfaces [28.130] the static friction was found to vary by more than a factor of four, while for bare mica surfaces the variation was by a factor of 3.5 [28.215]. In contrast to the sharp variations in adhesion energy over small twist angles, the variations in friction as a function of twist angle were much broader both in magnitude and angular spread. Similar variations in friction as a function of twist or misfit angles have also been observed in computer simulations [28.216].

Robbins et al. [28.217] computed the friction forces of two clean crystalline surfaces as a function of the angle between their surface lattices. They found that, for all nonzero angles (finite twist angles), the friction forces fell to zero due to incommensurability effects. They further found that submonolayer amounts of organic or other impurities trapped between two incommensurate surfaces can generate a finite friction force. They therefore concluded that any finite friction force measured between incommensurate surfaces is probably due to such *third-body* effects.

The reason why surface texture (lattice structure, roughness, granularity, topography, etc.) has a larger effect on the lateral (shear or friction) forces between two surfaces than on their normal (adhesion) forces is because friction is proportional to the adhesion hysteresis (Sect. 28.7.2), which can be low even when the adhesion force is high. It is also important to recognize that a system might be defined by more than one length scale. Some systems have well-defined dimensions or size (e.g., a perfect lattice, monodisperse nanoparticles), while others have different lateral and vertical dimensions and macroscopic curvature [28.122]. Furthermore, the morphology or texture of many systems, such as asperities that are randomly distributed over a surface, affects adhesion and tribological properties [28.118, 122, 151, 218, 219].

With rough surfaces, i. e., those that have *random* protrusions rather than being periodically structured, we expect a smearing out of the correlated intermolecular interactions that are involved in film freezing and melting (and in phase transitions in general). This should effectively eliminate the highly regular stick–slip and may also affect the location of the slipping planes [28.57, 151, 175]. The stick–slip friction of real surfaces, which are generally rough, may therefore be quite different from those of perfectly smooth surfaces composed of the same material. We should note, however, that even between rough surfaces, most of the contacts occur between the tips of microscopic asperities, which may be smooth over their microscopic contact area [28.220].

References

28.1 J.N. Israelachvili: *Intermolecular and Surface Forces*, 3rd edn. (Academic, London 2011)

28.2 P.M. Claesson, T. Ederth, V. Bergeron, M.W. Rutland: Techniques for measuring surface forces, Adv. Colloid Interface Sci. **67**, 119–183 (1996)

28.3 V.S.J. Craig: An historical review of surface force measurement techniques, Colloid Surf. A Physicochem. Eng. Asp. **129**, 75–93 (1997)

28.4 J.N. Israelachvili, G.E. Adams: Measurements of forces between two mica surfaces in aqueous electrolyte solutions in the range 0–100 nm, J. Chem. Soc. – Faraday Trans. 1 **74**, 975–1001 (1978)

28.5 J.N. Israelachvili: Surface forces and microrheology of molecularly thin liquid films. In: *Handbook of Micro/Nanotribology*, ed. by B. Bhushan (CRC, Boca Raton 1995) pp. 467–319

28.6 J.N. Israelachvili: Measurements of the viscosity of thin fluid films between two surfaces with and without adsorbed polymers, Colloid Polymer Sci **264**, 1060–1065 (1986)

28.7 A. Dhinojwala, S. Granick: Surface forces in the tapping mode: Solvent permeability and hydrodynamic thickness of adsorbed polymer brushes, Macromolecules **30**, 1079–1085 (1997)

28.8 B.V. Derjaguin: Untersuchungen über die Reibung und Adhäsion, IV. Theorie des Anhaftens kleiner Teilchen, Kolloid-Zeitschrift **69**, 155–164 (1934)

28.9 D.E. Leckband, J.N. Israelachvili, F.J. Schmitt, W. Knoll: Long-range attraction and molecular-rearrangements in receptor-ligand interactions, Science **255**, 1419–1421 (1992)

28.10 K.L. Johnson, K. Kendall, A.D. Roberts: Surface energy and the contact of elastic solids, Proc. R. Soc. A **324**, 301–313 (1971)

28.11 J.N. Israelachvili, P.M. McGuiggan: Adhesion and short-range forces between surfaces. Part 1: New apparatus for surface force measurements, J. Mater. Res. **5**, 2223–2231 (1990)

28.12 J.N. Israelachvili: Thin-film studies using multiple-beam interferometry, J. Colloid Interface Sci. **44**, 259–272 (1973)

28.13 M. Heuberger, G. Luengo, J. Israelachvili: Topographic information from multiple beam interferometry in the surface forces apparatus, Langmuir **13**, 3839–3848 (1997)

28.14 R.M. Pashley: DLVO and hydration forces between mica surfaces in Li^+, Na^+, K^+, and Cs^+ electrolyte solutions – A correlation of double-layer and hydration forces with surface cation-exchange properties, J. Colloid Interface Sci. **83**, 531–546 (1981)

28.15 R.M. Pashley: Hydration forces between mica surfaces in electrolyte solutions, Adv. Colloid Interface Sci. **16**, 57–62 (1982)

28.16 R.G. Horn, D.T. Smith, W. Haller: Surface forces and viscosity of water measured between silica sheets, Chem. Phys. Lett. **162**, 404–408 (1989)

28.17 R.G. Horn, D.R. Clarke, M.T. Clarkson: Direct measurement of surface forces between sapphire crystals in aqueous solutions, J. Mater. Res. **3**, 413–416 (1988)

28.18 W.W. Merrill, A.V. Pocius, B.V. Thakker, M. Tirrell: Direct measurement of molecular-level adhesion forces between biaxially oriented solid polymer films, Langmuir **7**, 1975–1980 (1991)

28.19 J. Klein: Forces between mica surfaces bearing adsorbed macromolecules in liquid-media, J. Chem. Soc. – Faraday Trans. 1 **79**, 99–118 (1983)

28.20 S.S. Patel, M. Tirrell: Measurement of forces between surfaces in polymer fluids, Annu. Rev. Phys. Chem. **40**, 597–635 (1989)

28.21 H. Watanabe, M. Tirrell: Measurement of forces in symmetric and asymmetric interactions between diblock copolymer layers adsorbed on mica, Macromolecules **26**, 6455–6466 (1993)

28.22 M. Ruths, D. Johannsmann, J. Ruhe, W. Knoll: Repulsive forces and relaxation on compression of entangled, polydisperse polystyrene brushes, Macromolecules **33**, 3860–3870 (2000)

28.23 C.A. Helm, J.N. Israelachvili, P.M. McGuiggan: Molecular mechanisms and forces involved in the adhesion and fusion of amphiphilic bilayers, Science **246**, 919–922 (1989)

28.24 Y.L. Chen, C.A. Helm, J.N. Israelachvili: Molecular mechanisms associated with adhesion and contact-angle hysteresis of monolayer surfaces, J. Phys. Chem. **95**, 10736–10747 (1991)

28.25 D.W. Lee, X. Banquy, K. Kristiansen, Y. Kaufman, J.M. Boggs, J.N. Israelachvili: Lipid domains control myelin basic protein adsorption and membrane interactions between model myelin lipid bilayers, Proc. Natl. Acad. Sci. USA **111**, E768–E775 (2014)

28.26 D.W. Lee, K. Kristiansen, S.H. Donaldson, N. Cadirov, X. Banquy, J.N. Israelachvili: Real-time intermembrane force measurements and imaging of lipid domain morphology during hemifusion, Nat. Commun. **6**, 8 (2015)

28.27 C.J. Coakley, D. Tabor: Direct measurement of van der Waals forces between solids in air, J. Phys. D **11**, L77–L82 (1978)

28.28 S.J. Hirz, A.M. Homola, G. Hadziioannou, C.W. Frank: Effect of substrate on shearing properties of ultrathin polymer-films, Langmuir **8**, 328–333 (1992)

28.29 S. Steinberg, W. Ducker, G. Vigil, C. Hyukjin, C. Frank, M.Z. Tseng, D.R. Clarke, J.N. Israelachvili: Van der Waals epitaxial-growth of alpha-alumina nanocrystals on mica, Science **260**, 656–659 (1993)

28.30 G. Vigil, Z.H. Xu, S. Steinberg, J. Israelachvili: Interactions of silica surfaces, J. Colloid Interface Sci. **165**, 367–385 (1994)

28.31 M. Ruths, M. Heuberger, V. Scheumann, J.J. Hu, W. Knoll: Confinement-induced film thickness transitions in liquid crystals between two alkanethiol monolayers on gold, Langmuir **17**, 6213–6219 (2001)

28.32 A.M. Homola, J.N. Israelachvili, M.L. Gee, P.M. McGuiggan: Measurements of and relation between the adhesion and friction of 2 surfaces separated by molecularly thin liquid-films, J. Tribol. – Trans. ASME **111**, 675–682 (1989)

28.33 J. Van Alsten, S. Granick: Shear rheology in a confined geometry-polysiloxane melts, Macromolecules **23**, 4856–4862 (1990)

28.34 J. Klein, E. Kumacheva: Simple liquids confined to molecularly thin layers. I. Confinement-induced liquid-to-solid phase transitions, J. Chem. Phys. **108**, 6996–7009 (1998)

28.35 G. Luengo, F.J. Schmitt, R. Hill, J. Israelachvili: Thin film rheology and tribology of confined polymer melts: Contrasts with bulk properties, Macromolecules **30**, 2482–2494 (1997)

28.36 D.W. Lee, X. Banquy, J.N. Israelachvili: Stick–slip friction and wear of articular joints, Proc. Natl. Acad. Sci. USA **110**, E567–E574 (2013)

28.37 E. Kumacheva: Interfacial friction measurement in surface force apparatus, Prog. Surf. Sci. **58**, 75–120 (1998)

28.38 H.B.G. Casimir, D. Polder: The influence of retardation on the London-van der Waals forces, Phys. Rev. **73**, 360–372 (1948)

28.39 H.C. Hamaker: The London-van der Waals attraction between spherical particles, Physica **4**, 1058–1072 (1937)

28.40 E.M. Lifshitz: The theory of molecular attractive forces between solids, Sov. Phys. JETP **2**, 73–83 (1956)

28.41 I.E. Dzyaloshinskii, E.M. Lifshitz, L.P. Pitaevskii: The general theory of van der Waals forces, Adv. Phys. **10**, 165–209 (1961)

28.42 H.W. Fox, W.A. Zisman: The spreading of liquids on low-energy surfaces. 3. Hydrocarbon surfaces, J. Colloid Sci. **7**, 428–442 (1952)

28.43 A. Meurk, P.F. Luckham, L. Bergstrom: Direct measurement of repulsive and attractive van der Waals forces between inorganic materials, Langmuir **13**, 3896–3899 (1997)

28.44 E.J.W. Verwey, J.T.G. Overbeek: *Theory of the Stability of lyophobic Colloids* (Elsevier, Amsterdam 1948)

28.45 J.E. Sader, S.L. Carnie, D.Y.C. Chan: Accurate analytic formulas for the double-layer interaction between spheres, J. Colloid Interface Sci. **171**, 46–54 (1995)

28.46 B. Derjaguin, L. Landau: Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes, Acta Physicochim. URSS **14**, 633–662 (1941)

28.47 D. Leckband, J. Israelachvili: Intermolecular forces in biology, Q. Rev. Biophys. **34**, 105–267 (2001)

28.48 M.L. Gee, J.N. Israelachvili: Interactions of surfactant monolayers across hydrocarbon liquids, J. Chem. Soc. – Faraday Trans. **86**, 4049–4058 (1990)

28.49 H.K. Christenson, D.W.R. Gruen, R.G. Horn, J.N. Israelachvili: Structuring in liquid alkanes between solid-surfaces-force measurements and mean-field theory, J. Chem. Phys. **87**, 1834–1841 (1987)

28.50 J.N. Israelachvili, S.J. Kott, M.L. Gee, T.A. Witten: Forces between mica surfaces across hydrocarbon liquids-effects of branching and polydispersity, Macromolecules **22**, 4247–4253 (1989)

28.51 R.G. Horn, J.N. Israelachvili: Direct measurement of structural forces between two surfaces in a non-polar liquid, J. Chem. Phys. **75**, 1400–1411 (1981)

28.52 I.K. Snook, W. Vanmegen: Solvation forces in simple dense fluids 1, J. Chem. Phys. **72**, 2907–2913 (1980)

28.53 R. Kjellander, S. Marcelja: Perturbation of hydrogen-bonding in water near polar surfaces, Chem. Phys. Lett. **120**, 393–396 (1985)

28.54 D. Henderson, M. Lozadacassou: A simple theory for the force between spheres immersed in a fluid, J. Colloid Interface Sci. **114**, 180–183 (1986)

28.55 J.P. Gao, W.D. Luedtke, U. Landman: Layering transitions and dynamics of confined liquid films, Phys. Rev. Lett. **79**, 705–708 (1997)

28.56 L.J.D. Frink, F. van Swol: Solvation forces between rough surfaces, J. Chem. Phys. **108**, 5588–5598 (1998)

28.57 J.P. Gao, W.D. Luedtke, U. Landman: Structures, solvation forces and shear of molecular films in a rough nano-confinement, Tribol. Lett. **9**, 3–13 (2000)

28.58 H.E. Stanley, J. Teixeira: Interpretation of the unusual behavior of H_2O and D_2O at low-temperatures-tests of a percolation model, J. Chem. Phys. **73**, 3404–3422 (1980)

28.59 H. van Olphen: *An Introduction to Clay Colloid Chemistry*, 2nd edn. (Wiley, New York 1977)

28.60 A.M. Schrader, S.H. Donaldson, J. Song, C.Y. Cheng, D.W. Lee, S. Han, J.N. Israelachvili: Correlating steric hydration forces with water dynamics through surface force and diffusion NMR measurements in a lipid-DMSO-H_2O system, Proc. Natl. Acad. Sci. USA **112**, 10708–10713 (2015)

28.61 B.R. Shrestha, X. Banquy: Hydration forces at solid and fluid biointerfaces, Biointerphases **11**, 11 (2016)

28.62 B.E. Viani, P.F. Low, C.B. Roth: Direct measurement of the relation between interlayer force and interlayer distance in the swelling of montmorillonite, J. Colloid Interface Sci. **96**, 229–244 (1983)

28.63 R.G. Horn: Surface forces and their action in ceramic materials, J. Am. Ceram. Soc. **73**, 1117–1135 (1990)

28.64 R.R. Lessard, Z. Sa: Bubble coalescence and gas transfer in aqueous electrolytic solutions, Ind. Eng. Chem. Fundam. **10**, 260–269 (1971)

28.65 J. Israelachvili, R. Pashley: The hydrophobic interaction is long-range, decaying exponentially with distance, Nature **300**, 341–342 (1982)

28.66 R.M. Pashley, P.M. McGuiggan, B.W. Ninham, D.F. Evans: Attractive forces between uncharged hydrophobic surfaces-direct measurements in aqueous-solution, Science **229**, 1088–1089 (1985)

28.67 P.M. Claesson, C.E. Blom, P.C. Herder, B.W. Ninham: Interactions between water-stable hydrophobic Langmuir–Blodgett monolayers on mica, J. Colloid Interface Sci. **114**, 234–242 (1986)

28.68 H.K. Christenson, J.F. Fang, B.W. Ninham, J.L. Parker: Effect of divalent electrolyte on the hydrophobic attraction, J. Phys. Chem. **94**, 8004–8006 (1990)

28.69 Y.H. Tsao, D.F. Evans, H. Wennerstrom: Long-range attractive force between hydrophobic surfaces observed by atomic-force microscopy, Science **262**, 547–550 (1993)

28.70 Y.I. Rabinovich, R.H. Yoon: Use of atomic-force microscope for the measurements of hydrophobic forces between silanated silica plate and glass sphere, Langmuir **10**, 1903–1909 (1994)

28.71 V.S.J. Craig, B.W. Ninham, R.M. Pashley: Study of the long-range hydrophobic attraction in concentrated salt solutions and its implications for electrostatic models, Langmuir **14**, 3326–3332 (1998)

28.72 H.K. Christenson, P.M. Claesson: Direct measurements of the force between hydrophobic surfaces

in water, Adv. Colloid Interface Sci. **91**, 391–436 (2001)
28.73 Q. Lin, E.E. Meyer, M. Tadmor, J.N. Israelachvili, T.L. Kuhl: Measurement of the long- and short-range hydrophobic attraction between surfactant-coated surfaces, Langmuir **21**, 251–255 (2005)
28.74 E.E. Meyer, Q. Lin, J.N. Israelachvili: Effects of dissolved gas on the hydrophobic attraction between surfactant-coated surfaces, Langmuir **21**, 256–259 (2005)
28.75 E.E. Meyer, Q. Lin, T. Hassenkam, E. Oroudjev, J.N. Israelachvili: Origin of the long-range attraction between surfactant-coated surfaces, Proc. Natl. Acad. Sci. USA **102**, 6839–6842 (2005)
28.76 S.H. Donaldson, A. Royne, K. Kristiansen, M.V. Rapp, S. Das, M.A. Gebbie, D.W. Lee, P. Stock, M. Valtiner, J. Israelachvili: Developing a general interaction potential for hydrophobic and hydrophilic interactions, Langmuir **31**, 2051–2064 (2015)
28.77 H.K. Christenson, P.M. Claesson, J. Berg, P.C. Herder: Forces between fluorocarbon surfactant monolayers-salt effects on the hydrophobic interaction, J. Phys. Chem. **93**, 1472–1478 (1989)
28.78 J.L. Parker, P.M. Claesson, P. Attard: Bubbles, cavities and the long-ranged attraction between hydrophobic surfaces, J. Phys. Chem. **98**, 8468–8480 (1994)
28.79 P. Attard: Nanobubbles and the hydrophobic attraction, Adv. Colloid Interface Sci. **104**, 75–91 (2003)
28.80 S.H. Donaldson, C.T. Lee, B.F. Chmelka, J.N. Israelachvili: General hydrophobic interaction potential for surfactant/lipid bilayers from direct force measurements between light-modulated bilayers, Proc. Natl. Acad. Sci. USA **108**, 15699–15704 (2011)
28.81 S. Marcelja, D.J. Mitchell, B.W. Ninham, M.J. Sculley: Role of solvent structure in solution theory, J. Chem. Soc. – Faraday Trans. II **73**, 630–648 (1977)
28.82 B. Jonsson, H. Wennerstrom: Image-charge forces in phospholipid-bilayer systems, J. Chem. Soc. – Faraday Trans. II **79**, 19–35 (1983)
28.83 A. Luzar, D. Bratko, L. Blum: Monte-Carlo simulation of hydrophobic interaction, J. Chem. Phys. **86**, 2955–2959 (1987)
28.84 K. Leung, A. Luzar: Dynamics of capillary evaporation. II. Free energy barriers, J. Chem. Phys. **113**, 5845–5852 (2000)
28.85 M. Mao, J.H. Zhang, R.H. Yoon, W.A. Ducker: Is there a thin film of air at the interface between water and smooth hydrophobic solids?, Langmuir **20**, 1843–1849 (2004)
28.86 M. Mao, J.H. Zhang, R.H. Yoon, W.A. Ducker: Erratum to 'Is there a thin film of air at the interface between water and smooth hydrophobic solids?', Langmuir **20**, 4310 (2004)
28.87 Y.X. Zhu, S. Granick: Rate-dependent slip of Newtonian liquid at smooth surfaces, Phys. Rev. Lett. **87**, 096105 (2001)
28.88 H.J. Taunton, C. Toprakcioglu, L.J. Fetters, J. Klein: Interactions between surfaces bearing end-adsorbed chains in a good solvent, Macromolecules **23**, 571–580 (1990)
28.89 J. Klein, P. Luckham: Forces between two adsorbed polyethylene oxide layers immersed in a good aqueous solvent, Nature **300**, 429–431 (1982)
28.90 J. Klein, P.F. Luckham: Long-range attractive forces between two mica surfaces in an aqueous polymer-solution, Nature **308**, 836–837 (1984)
28.91 P.F. Luckham, J. Klein: Forces between mica surfaces bearing adsorbed homopolymers in good solvents-the effect of bridging and dangling tails, J. Chem. Soc. – Faraday Trans. **86**, 1363–1368 (1990)
28.92 P.G. de Gennes: Polymers at an interface. 2. Interaction between two plates carrying adsorbed polymer layers, Macromolecules **15**, 492–500 (1982)
28.93 J. Scheutjens, G.J. Fleer: Interaction between two adsorbed polymer layers, Macromolecules **18**, 1882–1900 (1985)
28.94 E.A. Evans: Force between surfaces that confine a polymer-solution: Derivation from self-consistent field-theories, Macromolecules **22**, 2277–2286 (1989)
28.95 S. Asakura, F. Oosawa: Interaction between particles suspended in solutions of macromolecules, J. Polymer Sci. **33**, 183–192 (1958)
28.96 J.F. Joanny, L. Leibler, P.G. De Gennes: Effects of polymer-solutions on colloid stability, J. Polymer Sci. Pt. B **17**, 1073–1084 (1979)
28.97 R.I. Feigin, D.H. Napper: Stabilization of colloids by free polymer, J. Colloid Interface Sci. **74**, 567–571 (1980)
28.98 P.G. de Gennes: Polymer solutions near an interface. Adsorption and depletion layers, Macromolecules **14**, 1637–1644 (1981)
28.99 P.G. de Gennes: Polymers at an interface-a simplified view, Adv. Colloid Interface Sci. **27**, 189–209 (1987)
28.100 S. Alexander: Adsorption of chain molecules with a polar head a-scaling description, J. Phys. France **38**, 983–987 (1977)
28.101 P.G. de Gennes: Conformations of polymers attached to an interface, Macromolecules **13**, 1069–1075 (1980)
28.102 S.T. Milner, T.A. Witten, M.E. Cates: Theory of the grafted polymer brush, Macromolecules **21**, 2610–2619 (1988)
28.103 E.B. Zhulina, O.V. Borisov, V.A. Priamitsyn: Theory of steric stabilization of colloid dispersions by grafted polymers, J. Colloid Interface Sci. **137**, 495–511 (1990)
28.104 J. Klein, E. Kumacheva, D. Mahalu, D. Perahia, L.J. Fetters: Reduction of frictional forces between solid-surfaces bearing polymer brushes, Nature **370**, 634–636 (1994)
28.105 L.R. Fisher, J.N. Israelachvili: Direct measurement of the effect of meniscus forces on adhesion – Study of the applicability of macroscopic thermodynamics to microscopic liquid interfaces, Colloids Surf. **3**, 303–319 (1981)

28.106 H.K. Christenson: Adhesion between surfaces in undersaturated vapors – A reexamination of the influence of meniscus curvature and surface forces, J. Colloid Interface Sci. **121**, 170–178 (1988)

28.107 H. Hertz: Über die Berührung fester elastischer Körper, J. Reine Angew. Math. **92**, 156–171 (1881)

28.108 B.V. Derjaguin, V.M. Muller, Y.P. Toporov: Effect of contact deformations on adhesion of particles, J. Colloid Interface Sci. **53**, 314–326 (1975)

28.109 V.M. Muller, V.S. Yushchenko, B.V. Derjaguin: On the influence of molecular forces on the deformation of an elastic sphere and its sticking to a rigid plane, J. Colloid Interface Sci. **77**, 91–101 (1980)

28.110 V.M. Muller, B.V. Derjaguin, Y.P. Toporov: On 2 methods of calculation of the force of sticking of an elastic sphere to a rigid plane, Colloids Surf. **7**, 251–259 (1983)

28.111 D. Tabor: Surface forces and surface interactions, J. Colloid Interface Sci. **58**, 2–13 (1977)

28.112 D. Maugis: Adhesion of spheres-the JKR-DMT transition using a Dugdale model, J. Colloid Interface Sci. **150**, 243–269 (1992)

28.113 R.G. Horn, J.N. Israelachvili, F. Pribac: Measurement of the deformation and adhesion of solids in contact, J. Colloid Interface Sci. **115**, 480–492 (1987)

28.114 V. Mangipudi, M. Tirrell, A.V. Pocius: Direct measurement of molecular-level adhesion between poly(ethylene-terephthalate) and polyethylene films-determination of surface and interfacial energies, J. Adhes. Sci. Technol. **8**, 1251–1270 (1994)

28.115 I. Sridhar, K.L. Johnson, N.A. Fleck: Adhesion mechanics of the surface force apparatus, J. Phys. D **30**, 1710–1719 (1997)

28.116 I. Sridhar, Z.W. Zheng, K.L. Johnson: A detailed analysis of adhesion mechanics between a compliant elastic coating and a spherical probe, J. Phys. D **37**, 2886–2895 (2004)

28.117 J.A. Greenwood, J.B.P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. A **295**, 300–319 (1966)

28.118 B.N.J. Persson: Elastoplastic contact between randomly rough surfaces, Phys. Rev. Lett. **87**, 1–4 (2001)

28.119 F.P. Bowden, D. Tabor: *An Introduction to Tribology* (Anchor/Doubleday, Garden City 1973)

28.120 D. Maugis, H.M. Pollock: Surface forces, deformation and adherence at metal microcontacts, Acta Metall **32**, 1323–1334 (1984)

28.121 K.N.G. Fuller, D. Tabor: Effect of surface-roughness on adhesion of elastic solids, Proc. R. Soc. A **345**, 327–342 (1975)

28.122 B.N.J. Persson, E. Tosatti: The effect of surface roughness on the adhesion of elastic solids, J. Chem. Phys. **115**, 5597–5610 (2001)

28.123 H.C. Kim, T.P. Russell: Contact of elastic solids with rough surfaces, J. Polymer Sci. Pt. B **39**, 1848–1854 (2001)

28.124 P.M. McGuiggan, J.N. Israelachvili: Adhesion and short-range forces between surfaces. 2. Effects of surface lattice mismatch, J. Mater. Res. **5**, 2232–2243 (1990)

28.125 P.M. McGuiggan, J.N. Israelachvili: Measurements of the effect of angular lattice mismatch on the adhesion energy between two mica surfaces in water. In: *Characterization of the Structure and Chemistry of Defects in Materials*, ed. by B.C. Larson, M. Ruhle, D.N. Seidman (Materials Research Society, Pittsburgh 1989) pp. 349–360

28.126 C.L. Rhykerd, M. Schoen, D.J. Diestler, J.H. Cushman: Epitaxy in simple classical fluids in micropores and near-solid surfaces, Nature **330**, 461–463 (1987)

28.127 M. Schoen, C.L. Rhykerd, D.J. Diestler, J.H. Cushman: Shear forces in molecularly thin-films, Science **245**, 1223–1225 (1989)

28.128 P.A. Thompson, M.O. Robbins: Origin of stick–slip motion in boundary lubrication, Science **250**, 792–794 (1990)

28.129 M. Ruths, S. Granick: Influence of alignment of crystalline confining surfaces on static forces and shear in a liquid crystal, 4′-*n*-pentyl-4-cyanobiphenyl, Langmuir **16**, 8368–8376 (2000)

28.130 A.D. Berman: *Dynamics of Molecules at Surfaces*, Ph.D. Thesis (Univ. California, Santa Barbara 1996)

28.131 E. Barthel, S. Roux: Velocity-dependent adherence: An analytical approach for the JKR and DMT models, Langmuir **16**, 8134–8138 (2000)

28.132 M. Ruths, S. Granick: Rate-dependent adhesion between polymer and surfactant monolayers on elastic substrates, Langmuir **14**, 1804–1814 (1998)

28.133 H. Yoshizawa, Y.L. Chen, J. Israelachvili: Fundamental mechanisms of interfacial friction. 1. Relation between adhesion and friction, J. Phys. Chem. **97**, 4128–4140 (1993)

28.134 N. Maeda, N.H. Chen, M. Tirrell, J.N. Israelachvili: Adhesion and friction mechanisms of polymer-on-polymer surfaces, Science **297**, 379–382 (2002)

28.135 C.A. Miller, P. Neogi: *Interfacial Phenomena: Equilibrium and Dynamic Effects* (Marcel Dekker, New York 1985)

28.136 J. Israelachvili, A. Berman: Irreversibility, energy-dissipation, and time effects in intermolecular and surface interactions, Isr. J. Chem. **35**, 85–91 (1995)

28.137 M. Deruelle, M. Tirrell, Y. Marciano, H. Hervet, L. Léger: Adhesion energy between polymer networks and solid surfaces modified by polymer attachment, Faraday Discuss. **98**, 55–65 (1995)

28.138 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids* (Clarendon, London 1971)

28.139 S. Granick: Motions and relaxations of confined liquids, Science **253**, 1374–1379 (1991)

28.140 H.W. Hu, S. Granick: Viscoelastic dynamics of confined polymer melts, Science **258**, 1339–1342 (1992)

28.141 M.L. Gee, P.M. McGuiggan, J.N. Israelachvili, A.M. Homola: Liquid to solid-like transitions of molecularly thin-films under shear, J. Chem. Phys. **93**, 1895–1906 (1990)

28.142 J. Israelachvili, M. Gee, P. McGuiggan, P. Thompson, M. Robbins: Melting freezing transitions in molecularly thin liquid films during shed. In: *Proc. Fall Meet. Mater. Res. Soc. Boston*, ed. by J.M. Drake, J. Klafter, R. Kopelman (MRS, Pitts-

burgh 1990) pp. 3–6
28.143 J. Israelachvili, P. McGuiggan, M. Gee, A. Homola, M. Robbins, P. Thompson: Liquid dynamics in molecularly thin-films, J. Phys. Condensed Matter **2**, SA89–SA98 (1990)
28.144 J. Klein, E. Kumacheva: Confinement-induced phase-transitions in simple liquids, Science **269**, 816–819 (1995)
28.145 A.L. Demirel, S. Granick: Glasslike transition of a confined simple fluid, Phys. Rev. Lett. **77**, 2261–2264 (1996)
28.146 D. Dowson: *History of Tribology*, 2nd edn. (Professional Engineering, London 1998)
28.147 A. Berman, C. Drummond, J. Israelachvili: Amontons' law at the molecular level, Tribol. Lett. **4**, 95–101 (1998)
28.148 M. Ruths: Boundary friction of aromatic self-assembled monolayers: Comparison of systems with one or both sliding surfaces covered with a thiol monolayer, Langmuir **19**, 6788–6795 (2003)
28.149 D. Gourdon, J.N. Israelachvili: Transitions between smooth and complex stick–slip sliding of surfaces, Phys. Rev. E **68**, 10 (2003)
28.150 M. Ruths, N.A. Alcantar, J.N. Israelachvili: Boundary friction of aromatic silane self-assembled monolayers measured with the surface forces apparatus and friction force microscopy, J. Phys. Chem. B **107**, 11149–11157 (2003)
28.151 J.P. Gao, W.D. Luedtke, D. Gourdon, M. Ruths, J.N. Israelachvili, U. Landman: Frictional forces and Amontons' law: From the molecular to the macroscopic scale, J. Phys. Chem. B **108**, 3410–3425 (2004)
28.152 D. Tabor: The role of surface and intermolecular forces in thin film lubrication, Tribol. Ser. **7**, 651–682 (1981)
28.153 M.J. Sutcliffe, S.R. Taylor, A. Cameron: Molecular asperity theory of boundary friction, Wear **51**, 181–192 (1978)
28.154 G.M. McClelland: Friction at weakly interacting interfaces. In: *Adhesion and Friction*, ed. by M. Grunze, H.J. Kreuzer (Springer, Heidelberg 1989)
28.155 J.N. Israelachvili, Y.L. Chen, H. Yoshizawa: Relationship between adhesion and friction forces, J. Adhes. Sci. Technol. **8**, 1231–1249 (1994)
28.156 D.H. Buckley: Metal-to-metal interface and its effect on adhesion and friction, J. Colloid Interface Sci. **58**, 36–53 (1977)
28.157 R.G. Cain, N.W. Page, S. Biggs: Microscopic and macroscopic aspects of stick–slip motion in granular shear, Phys. Rev. E **64**, 8 (2001)
28.158 A. Dhinojwala, S.C. Bae, S. Granick: Shear-induced dilation of confined liquid films, Tribol. Lett. **9**, 55–62 (2000)
28.159 L.M. Qian, G. Luengo, E. Perez: Thermally activated lubrication with alkanes: The effect of chain length, Europhys. Lett. **61**, 268–274 (2003)
28.160 A. Berman, J. Israelachvili: Control and minimization of friction via surface modification. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 317–329
28.161 B.V. Derjaguin: Molekulartheorie der äußeren Reibung, Z. Phys. **88**, 661–675 (1934)
28.162 B.V. Derjaguin: Mechanical-properties of the boundary lubrication layer, Wear **128**, 19–27 (1988)
28.163 B.J. Briscoe, D.C.B. Evans: The shear properties of Langmuir–Blodgett layers, Proc. R. Soc. A **380**, 389–407 (1982)
28.164 P.F. Luckham, S. Manimaaran: Investigating adsorbed polymer layer behaviour using dynamic surface force apparatuses – A review, Adv. Colloid Interface Sci. **73**, 1–46 (1997)
28.165 B.Q. Luan, M.O. Robbins: The breakdown of continuum models for mechanical contacts, Nature **435**, 929–932 (2005)
28.166 Y.F. Mo, K.T. Turner, I. Szlufarska: Friction laws at the nanoscale, Nature **457**, 1116–1119 (2009)
28.167 Y.F. Mo, I. Szlufarska: Roughness picture of friction in dry nanoscale contacts, Phys. Rev. B **81**, 17 (2010)
28.168 S.F. Cheng, B.Q. Luan, M.O. Robbins: Contact and friction of nanoasperities: Effects of adsorbed monolayers, Phys. Rev. E **81**, 17 (2010)
28.169 A.M. Homola, J.N. Israelachvili, P.M. McGuiggan, M.L. Gee: Fundamental experimental studies in tribology – The transition from interfacial friction of undamaged molecularly smooth surfaces to normal friction with wear, Wear **136**, 65–83 (1990)
28.170 B.J. Briscoe, D.C.B. Evans, D. Tabor: Influence of contact pressure and saponification on sliding behavior of stearic-acid monolayers, J. Colloid Interface Sci. **61**, 9–13 (1977)
28.171 S. Yamada, J. Israelachvili: Friction and adhesion hysteresis of fluorocarbon surfactant monolayer-coated surfaces measured with the surface forces apparatus, J. Phys. Chem. B **102**, 234–244 (1998)
28.172 G. Luengo, S.E. Campbell, V.I. Srdanov, F. Wudl, J.N. Israelachvili: Direct measurement of the adhesion and friction of smooth C_{60} surfaces, Chem. Mater. **9**, 1166–1171 (1997)
28.173 J. Israelachvili, Y.-L. Chen, H. Yoshizawa: Relationship between adhesion and friction forces. In: *Fundamentals of Adhesion and Interfaces*, ed. by D.S. Rimai, L.P. DeMejo, K.L. Mittal (VSP, Utrecht 1995) pp. 261–279
28.174 U. Raviv, S. Perkin, P. Laurat, J. Klein: Fluidity of water confined down to subnanometer films, Langmuir **20**, 5322–5332 (2004)
28.175 P. Berthoud, T. Baumberger, C. G'Sell, J.M. Hiver: Physical analysis of the state- and rate-dependent friction law: Static friction, Phys. Rev. B **59**, 14313–14327 (1999)
28.176 D.Y.C. Chan, R.G. Horn: The drainage of thin liquid-films between solid-surfaces, J. Chem. Phys. **83**, 5311–5324 (1985)
28.177 J.N. Israelachvili, S.J. Kott: Shear properties and structure of simple liquids in molecularly thin-films-the transition from bulk (continuum) to molecular behavior with decreasing film thickness, J. Colloid Interface Sci. **129**, 461–467 (1989)
28.178 S.E. Campbell, G. Luengo, V.I. Srdanov, F. Wudl, J.N. Israelachvili: Very low viscosity at the solid-liquid interface induced by adsorbed C-60 mono-

layers, Nature **382**, 520–522 (1996)
28.179 J. Klein, Y. Kamiyama, H. Yoshizawa, J.N. Israelachvili, G.H. Fredrickson, P. Pincus, L.J. Fetters: Lubrication forces between surfaces bearing polymer brushes, Macromolecules **26**, 5552–5560 (1993)
28.180 T.L. Kuhl, A.D. Berman, S.W. Hui, J.N. Israelachvili: Part 1. Direct measurement of depletion attraction and thin film viscosity between lipid bilayers in aqueous polyethylene glycol solutions, Macromolecules **31**, 8250–8257 (1998)
28.181 G. Luengo, J. Israelachvili, A. Dhinojwala, S. Granick: Generalized effectes in confined fluids: New friction map for boundary lubrication, Wear **200**, 328–335 (1996)
28.182 G. Luengo, J. Israelachvili, A. Dhinojwala, S. Granick: Erratum to 'Generalized effectes in confined fluids: New friction map for boundary lubrication', Wear **205**, 246 (1997)
28.183 P.A. Thompson, G.S. Grest, M.O. Robbins: Phase-transitions and universal dynamics in confined films, Phys. Rev. Lett. **68**, 3448–3451 (1992)
28.184 A. Subbotin, A. Semenov, E. Manias, G. Hadzioannou, G. Tenbrinke: Rheology of confined polymer melts under shear-flow-strong adsorption limit, Macromolecules **28**, 1511–1515 (1995)
28.185 A. Subbotin, A. Semenov, G. Hadziioannou, G. Tenbrinke: Rheology of confined polymer melts under shear-flow-weak adsorption limit, Macromolecules **28**, 3901–3903 (1995)
28.186 H.W. Xie, K.Y. Song, D.J. Mann, W.L. Hase: Temperature gradients and frictional energy dissipation in the sliding of hydroxylated α-alumina surfaces, Phys. Chem. Chem. Phys. **4**, 5377–5385 (2002)
28.187 J.N. Israelachvili, P.M. McGuiggan, A.M. Homola: Dynamic properties of molecularly thin liquid-films, Science **240**, 189–191 (1988)
28.188 M. Schoen, S. Hess, D.J. Diestler: Rheological properties of confined thin-films, Phys. Rev. E **52**, 2587–2602 (1995)
28.189 M. Heuberger, C. Drummond, J. Israelachvili: Coupling of normal and transverse motions during frictional sliding, J. Phys. Chem. B **102**, 5038–5041 (1998)
28.190 V. Zaloj, M. Urbakh, J. Klafter: Modifying friction by manipulating normal response to lateral motion, Phys. Rev. Lett. **82**, 4823–4826 (1999)
28.191 C. Drummond, J. Elezgaray, P. Richetti: Behavior of adhesive boundary lubricated surfaces under shear: A new dynamic transition, Europhys. Lett. **58**, 503–509 (2002)
28.192 A.E. Filippov, J. Klafter, M. Urbakh: Inverted stick–slip friction: What is the mechanism?, J. Chem. Phys. **116**, 6871–6874 (2002)
28.193 C. Drummond, J. Israelachvili: Dynamic phase transitions in confined lubricant fluids under shear, Phys. Rev. E **63**, 11 (2001)
28.194 J.D. Ferry: *Viscoelastic Properties of Polymers*, 3rd edn. (Wiley, New York 1980)
28.195 E. Rabinowicz: *Friction and Wear of Materials* (Wiley, New York 1995)
28.196 E. Rabinowicz: The intrinsic variables affecting the stick–slip process, Proc. Phys. Soc. Lond. **71**, 668–675 (1958)
28.197 F. Heslot, T. Baumberger, B. Perrin, B. Caroli, C. Caroli: Creep, stick–slip, and dry-friction dynamics: Experiments and a heuristic model, Phys. Rev. E **49**, 4973–4988 (1994)
28.198 F. Heymann, E. Rabinowicz, B. Rightmire: Friction aparatus for very low-speed sliding studies, Rev. Sci. Instrum. **26**, 56–58 (1955)
28.199 J.H. Dieterich: Time-dependent friction and mechanics of stick–slip, Pure Appl. Geophys. **116**, 790–806 (1978)
28.200 G.A. Tomlinson: A molecular theory of friction, Philos. Mag. **7**, 905–939 (1929)
28.201 J.M. Carlson, J.S. Langer: Mechanical model of an earthquake fault, Phys. Rev. A **40**, 6470–6484 (1989)
28.202 B.N.J. Persson: Theory of friction-the role of elasticity in boundary lubrication, Phys. Rev. B **50**, 4771–4786 (1994)
28.203 S. Nasuno, A. Kudrolli, J.P. Gollub: Friction in granular layers: Hysteresis and precursors, Phys. Rev. Lett. **79**, 949–952 (1997)
28.204 H. Yoshizawa, J. Israelachvili: Fundamental mechanisms of interfacial friction. 2. Stick–slip friction of spherical and chain molecules, J. Phys. Chem. **97**, 11300–11313 (1993)
28.205 M.O. Robbins, P.A. Thompson: Critical velocity of stick–slip motion, Science **253**, 916–916 (1991)
28.206 P. Bordarier, M. Schoen, A.H. Fuchs: Stick–slip phase transitions in confined solidlike films from an equilibrium perspective, Phys. Rev. E **57**, 1621–1635 (1998)
28.207 J.M. Carlson, A.A. Batista: Constitutive relation for the friction between lubricated surfaces, Phys. Rev. E **53**, 4153–4165 (1996)
28.208 A.D. Berman, W.A. Ducker, J.N. Israelachvili: Origin and characterization of different stick–slip friction mechanisms, Langmuir **12**, 4559–4563 (1996)
28.209 A.D. Berman, W.A. Ducker, J.N. Israelachvili: Experimental and theoretical investigations of stick–slip friction mechanisms. In: *Physics of Sliding Friction*, ed. by B.N.J. Persson, E. Tosatti (Kluwer, Dordrecht 1996) pp. 51–67
28.210 K.G. McLaren, D. Tabor: Visco-elastic properties and friction of solids-friction of polymers – Influence of speed and temperature, Nature **197**, 856–858 (1963)
28.211 K.A. Grosch: Relation between friction and viscoelastic properties of rubber, Nature **197**, 858–859 (1963)
28.212 C. Drummond, J. Israelachvili: Dynamic behavior of confined branched hydrocarbon lubricant fluids under shear, Macromolecules **33**, 4910–4920 (2000)
28.213 J. Warnock, D.D. Awschalom, M.W. Shafer: Orientational behavior of molecular liquids in restricted geometries, Phys. Rev. B **34**, 475–478 (1986)
28.214 D.D. Awschalom, J. Warnock: Supercooled liquids and solids in porous-glass, Phys. Rev. B **35**, 6779–

6785 (1987)
28.215 M. Hirano, K. Shinjo, R. Kaneko, Y. Murata: Anisotropy of frictional forces in muscovite mica, Phys. Rev. Lett. **67**, 2642–2645 (1991)
28.216 T. Gyalog, H. Thomas: Friction between atomically flat surfaces, Europhys. Lett. **37**, 195–200 (1997)
28.217 G. He, M.H. Muser, M.O. Robbins: Adsorbed layers and the origin of static friction, Science **284**, 1650–1652 (1999)
28.218 S. Hyun, L. Pei, J.F. Molinari, M.O. Robbins: Finite-element analysis of contact between elastic self-affine surfaces, Phys. Rev. E **70**, 12 (2004)
28.219 J. Israelachvili, N. Maeda, K.J. Rosenberg, M. Akbulut: Effects of sub-ångstrom (pico-scale) structure of surfaces on adhesion, friction, and bulk mechanical properties, J. Mater. Res. **20**, 1952–1972 (2005)
28.220 T.R. Thomas: *Rough Surfaces*, 2nd edn. (Imperial College, London 1999)

29. Atomic Scale Friction Phenomena

Enrico Gnecco, Rémy Pawlak, Marcin Kisiel, Thilo Glatzel, Ernst Meyer

Friction has long been the subject of research: the empirical da Vinci–Amontons friction laws have been common knowledge for centuries. Macroscopic experiments performed by the school of *Bowden* and *Tabor* revealed that macroscopic friction can be related to the collective action of small asperities. Over the last 25 years, experiments performed with the atomic force microscope have provided new insights into the physics of single asperities sliding over surfaces. This development, together with the results from complementary experiments using surface force apparatus and the quartz microbalance, have led to the new field of *nanotribology*. At the same time, increasing computing power has permitted the simulation of processes that occur during sliding contact involving several hundreds of atoms. It has become clear that atomic processes cannot be neglected when interpreting nanotribology experiments. Even on well-defined surfaces, experiments have revealed that atomic structure is directly linked to friction force. This chapter will describe friction force microscopy (FFM) experiments that reveal, more or less directly, atomic processes during sliding contact.

We will begin by introducing friction force microscopy, including the calibration of cantilever force sensors and special aspects of the ultrahigh vacuum environment. The empirical Prandtl–Tomlinson model often used to describe atomic stick–slip results is therefore presented in detail. We review experimental results regarding atomic friction, including thermal activation, velocity dependence, and temperature dependence. The geometry of the contact is crucial to the interpretation of experimental results, such as the calculation of the lateral contact stiffness. The onset of wear on the atomic scale has recently been studied experimentally and it is described here. The chapter ends with a discussion of recent experiments aimed at detecting the dissipative forces acting when a sharp tip is moved parallel and very close to a solid surface without being in contact with it, or when small entities such as single polymer chains, graphene nanoribbons, or large organic molecules are manipulated.

29.1 **Friction Force Microscopy in Selected Environments** 988
29.1.1 Force Calibration 988
29.1.2 Friction Force Microscopy in Ultra High Vacuum 990
29.1.3 Energy Dissipation in Noncontact Atomic Force Microscopy 992
29.1.4 Friction Force Microscopy in Water 993
29.2 **The Prandtl–Tomlinson Model** 993
29.2.1 One-Dimensional Prandtl–Tomlinson Model 993
29.2.2 Two-Dimensional Prandtl–Tomlinson Model 994
29.2.3 Friction Between Atomically Flat Surfaces 995
29.3 **Friction Experiments on the Atomic Scale** 996
29.4 **Thermal Effects on Atomic Friction** 998
29.4.1 The Prandtl–Tomlinson Model at Finite Temperature 998
29.4.2 Velocity and Temperature Dependence of Atomic-Scale Friction 999
29.5 **Friction on the Nanometer Scale** 999
29.5.1 Continuum Mechanics of Single Asperities 1000
29.5.2 Dependence of Friction on Load 1000
29.5.3 Estimation of the Contact Area 1001
29.6 **Wear on the Atomic Scale** 1002
29.7 **Noncontact Friction** 1003
29.8 **Single Molecule Friction** 1005
29.9 **Conclusion** 1008
References 1008

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_29

29.1 Friction Force Microscopy in Selected Environments

The *friction force microscope* (FFM, also called the *lateral force microscope*, LFM) is used to record the torsion deformation of an elastic cantilever driving a sharp tip on a solid surface and to investigate dissipative processes occurring in a sliding contact down to the atomic scale (Fig. 29.1). The relative motion of tip and surface is realized by a *scanner* made by piezoelectric elements, which moves the surface below the tip at a certain rate. The scanner can also be extended or retracted to vary the normal force F_N applied to the surface. This force is responsible for the deflection of the flexible cantilever supporting the tip. If the normal force F_N increases while scanning due to local topographic variations, the scanner is retracted by a feedback loop. Conversely, if F_N decreases, the surface is brought closer to the tip by extending the scanner. In this way, the surface topography can be determined line-by-line from the vertical displacement of the scanner. Accurate control of such movement is made possible by a light beam reflected from the rear of the lever into a photodetector. When the cantilever bends, the light spot on the detector moves up or down and causes the photocurrent to vary, when in turn triggers a corresponding change in the normal force F_N applied.

The relative sliding of tip and surface is usually also accompanied by *friction*. A lateral force F_L, which acts in the opposite direction to the scan velocity v hinders the motion of the tip. This force causes torsion in the cantilever, which can be observed along with the topography if the photodetector can detect not only the normal deflection but also the lateral movement of the lever while scanning. In practice this is achieved using four-quadrant photodetectors, as shown in Fig. 29.1. We should note that friction forces also cause lateral bending of the cantilever, but this effect is negligible if the thickness of the lever is much less than the width.

The FFM was first used by *Mate* et al. in 1987 to study the friction associated with atomic features [29.1]. In their experiment, Mate used a tungsten wire and a slightly different technique (nonfiber interferometry) to detect lateral forces. Optical beam deflection was introduced later by *Marti* et al. [29.2] and *Meyer* et al. [29.3]. Other methods of measuring the forces between tip and surface include capacitance detection [29.4], dual fiber interferometry [29.5], and piezolevers [29.6]. In the first method, two plates close to the cantilever reveal the capacitance while scanning. The second technique uses two optical fibers to detect the cantilever deflection along two orthogonal directions aligned 45° with respect to the surface normal. Finally, in the third method, cantilevers with two Wheatstone bridges at their bases reveal normal and lateral forces, which are respectively proportional to the sum and the difference of both bridge signals.

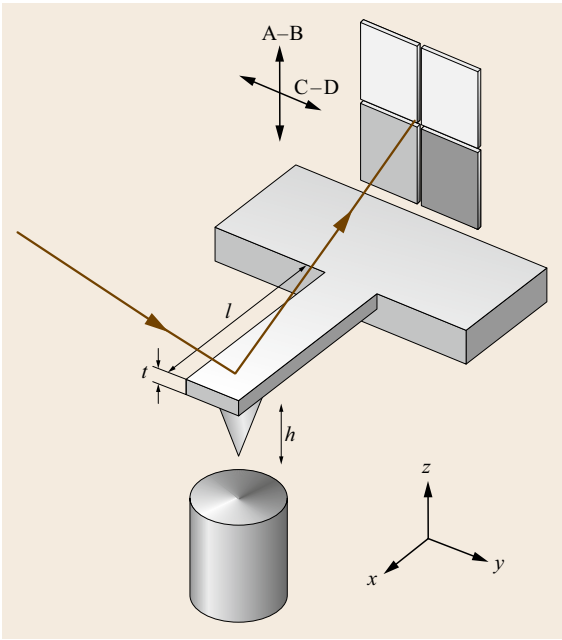


Fig. 29.1 Schematic diagram of a beam-deflection friction force microscope

29.1.1 Force Calibration

Force calibration is relatively simple if rectangular cantilevers are used. Because of possible discrepancies with the geometric values provided by manufacturers, one should use optical and electron microscopes to determine the width, thickness, and length of the cantilever (w, t, l), the tip height h, and the position of the tip with respect to the cantilever. The thickness of the cantilever can also be determined from the resonance frequency of the lever f_0 using the relation [29.7]

$$t = \frac{2\sqrt{12}\pi}{1.875^2}\sqrt{\frac{\rho}{E}}f_0 l^2 \,, \tag{29.1}$$

where ρ is the density of the cantilever and E is its Young's modulus. The normal spring constant (c_N) and the lateral spring constant (c_L) of the lever are given by

$$c_N = \frac{Ewt^3}{4l^3}, \quad c_L = \frac{Gwt^3}{3h^2 l} \,, \tag{29.2}$$

where G is the shear modulus. Figure 29.2 shows some SEM (scanning electron microscope) images of rectangular silicon cantilevers used for FFM. In the case of silicon $\rho = 2.33 \times 10^3\,\mathrm{kg/m^3}$, $E = 1.69 \times 10^{11}\,\mathrm{N/m^2}$, and

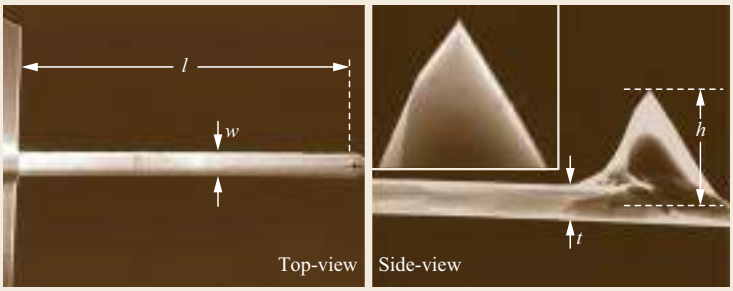


Fig. 29.2 SEM images of a rectangular cantilever. The relevant dimensions are $l = 445\,\mu\text{m}$, $w = 43\,\mu\text{m}$, $t = 4.5\,\mu\text{m}$, $h = 14.75\,\mu\text{m}$. Note that h is given by the sum of the tip height and half of the cantilever thickness

$G = 0.5 \times 10^{11}\,\text{N/m}^2$. Thus, for the cantilever shown in Fig. 29.2, $c_\text{N} = 1.9\,\text{N/m}$ and $c_\text{L} = 675\,\text{N/m}$.

The next step consists of measuring the sensitivity of the photodetector S_z (nm/V). For beam-deflection FFMs, the sensitivity S_z can be determined by force versus distance curves measured on hard surfaces (e.g., Al_2O_3), where elastic deformations are negligible and the vertical movement of the scanner equals the deflection of the cantilever. A typical dependence of the vertical signal recorded by the four-quadrant detector V_N in ambient conditions on the tip–surface distance is sketched in Fig. 29.3. When the tip is approached towards the surface, no signal is revealed until the tip jumps into contact at $z = z_1$. Further extension and successive retraction of the scanner result in elastic behavior until the tip jumps out of contact again at a distance $z_2 > z_1$ (when $V_\text{N} = V_\text{off}$). The slope of the elastic part of the curve gives the required sensitivity S_z.

The normal and lateral forces depend on the voltage V_N and the horizontal signals V_L as

$$F_\text{N} = c_\text{N} S_\text{z} V_\text{N}, \quad F_\text{L} = \frac{3}{2} c_\text{L} \frac{h}{l} S_\text{z} V_\text{L} \,. \tag{29.3}$$

It is assumed here that the light beam is positioned precisely above the probing tip, which may not always be the case in practice.

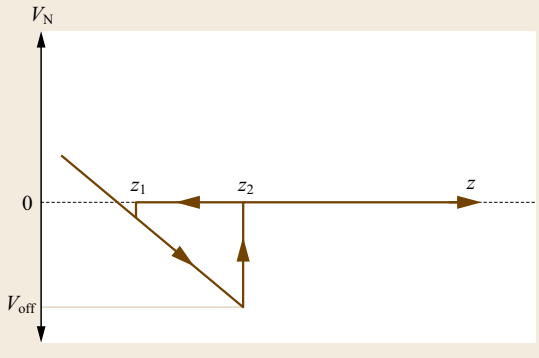


Fig. 29.3 Sketch of a typical force versus distance curve

The normal spring constant c_N can also be calibrated using other methods. *Cleveland* et al. [29.8] attached tungsten spheres to the tip, which changes the resonance frequency f_0 according to the formula

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{c_\text{N}}{M + m^*}} \,. \tag{29.4}$$

M is the mass of the added object and m^* is an effective mass of the cantilever, which depends on its geometry. The spring constant can be extrapolated from the frequency shifts corresponding to the different masses attached. As an alternative, *Hutter* et al. observed that the spring constant c_N can be related to the area of the power spectrum of the thermal fluctuations of the cantilever P [29.9]. The correct relation is $c_\text{N} = 4k_\text{B}T/(3P)$, where $k_\text{B} \approx 1.38 \times 10^{-23}\,\text{J/K}$ is Boltzmann's constant and T is the temperature [29.10]. Cantilevers with different shapes require finite element analysis, although analytical formulas can be derived in a few cases. For V-shaped cantilevers, *Neumeister* et al. derived the following approximation for the lateral spring constant c_L [29.11]

$$c_\text{L} = \frac{Et^2}{3(1+\upsilon)h^2} \times \left(\frac{1}{\tan\alpha} \ln \frac{w}{d \sin\alpha} + \frac{L\cos\alpha}{w} - \frac{3\sin 2\alpha}{8} \right)^{-1} . \tag{29.5}$$

The geometrical quantities L, w, α, d, t, and h are defined in Fig. 29.4. The expression for the normal constant is more complex and can be found in the cited reference.

Surfaces with well-defined profiles permit an alternative in situ calibration of lateral forces [29.12]. As an example, Fig. 29.5 shows a commercial grating formed by alternate faces with opposite inclinations with respect to the scan direction. When the tip slides on the inclined planes, the normal force F_N and the lateral force F_L with respect to the surface are different from the two components $F_\perp$ and $F_\parallel$, which are separated by the photodiode (Fig. 29.6a).

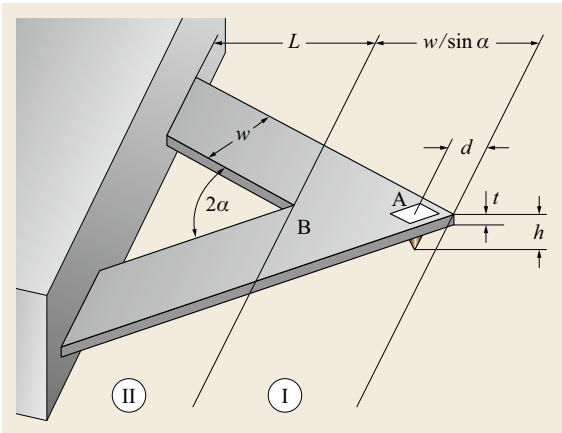


Fig. 29.4 Geometry of a V-shaped cantilever (after [29.11])

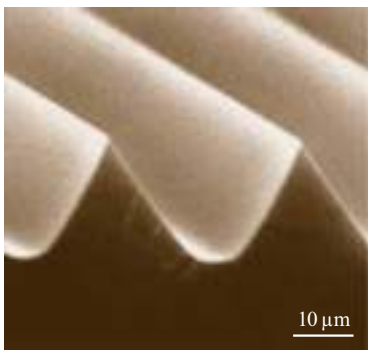


Fig. 29.5 Silicon grating formed by alternating faces angled at ±55° from the surface (Courtesy of Silicon-MDT Ltd., Moscow)

If Amontons' relation $F_L = \mu F_N$ holds, the component $F_\parallel$ can be expressed in terms of $F_\perp$

$$F_1 = \frac{\mu + \tan\theta}{1 - \mu\tan\theta} F_\perp \,. \tag{29.6}$$

The component $F_\perp$ is kept constant by the feedback loop. The sum of and the difference between the $F_\parallel$ values for the two planes (1) and (2) are given by

$$F_+ \equiv F_\parallel^{(1)} + Ff_\parallel^{(2)} = \frac{2\mu(1+\tan^2\theta)}{1-\mu^2\tan^2\theta} F_\perp \,,$$
$$F_- \equiv F_\parallel^{(1)} - F_\parallel^{(2)} = \frac{2(1+\mu^2)\tan\theta}{1-\mu^2\tan^2\theta} F_\perp \,. \tag{29.7}$$

The values of F_+ and F_- (in volts) can be measured by scanning the profile back and forth (Fig. 29.6b). If F_+ and F_- are recorded with different values of $F_\perp$, one can determine the conversion ratio between volts and nanonewtons as well as the coefficient of friction (COF) μ.

An accurate error analysis of lateral force calibration was provided by *Schwarz* et al., who discussed the importance of cantilever oscillations induced by the feedback loop and of geometrical positioning of the laser beam onto the cantilever [29.13]. Other sources of error (in-plane deformation and cantilever tilt) were discussed by *Sader* et al. [29.14, 15].

An adequate estimation of the radius of curvature of the tip R is also important. This quantity can be directly measured with a scanning electron microscope. If high-aspect structures such as step sites or whiskers are scanned, the resulting images are convolutions of those surfaces with the tip shape. A deconvolution algorithm that allows for the extraction of the probe tip's radius of curvature was proposed by *Villarrubia* [29.16].

29.1.2 Friction Force Microscopy in Ultra High Vacuum

Atomic friction studies require well-defined surfaces and – whenever possible – tips. For the surfaces, established methods of surface science performed in ultrahigh vacuum (UHV) can be employed. Ionic crystals such as NaCl have become standard materials for friction force microscopy on the atomic scale. Atomically clean and flat surfaces can be prepared by cleavage in UHV. The crystal has to be heated to $\approx 150\,^\circ$C for 1 h in order to remove charge after the cleavage process. Metal surfaces can be cleaned and flattened by cycles of sputtering with argon ions and annealing. Even surfaces prepared in air or liquids, such as self-assembled molecular monolayers, can be transferred into the vacuum and studied after careful heating procedures that remove water layers.

Tip preparation in UHV is more difficult. Most force sensors for friction studies have silicon nitride (Si_3N_4) or pure silicon tips. Tips can be cleaned and oxide layers removed by sputtering with argon ions. However, the sharpness of the tip is normally reduced by sputtering. As an alternative, tips can be etched in fluoric acid directly before transfer to the UHV. The significance of tip preparation is limited by the fact that the chemical and geometrical structure of the tip can undergo significant changes when sliding over the surface.

Using FFM in UHV conditions requires some additional effort. First of all, only materials with low vapor pressures can be used, which excludes most plastics and lubricants. Beam-deflection FFMs employ either a light source in the vacuum chamber or an optical fiber guiding the light into the chamber. The positioning of the light beam on the cantilever and the reflected beam on the position-sensitive detector is achieved by motorized mirrors or by moving the light source or detector. Furthermore, a motorized sample approach must be realized.

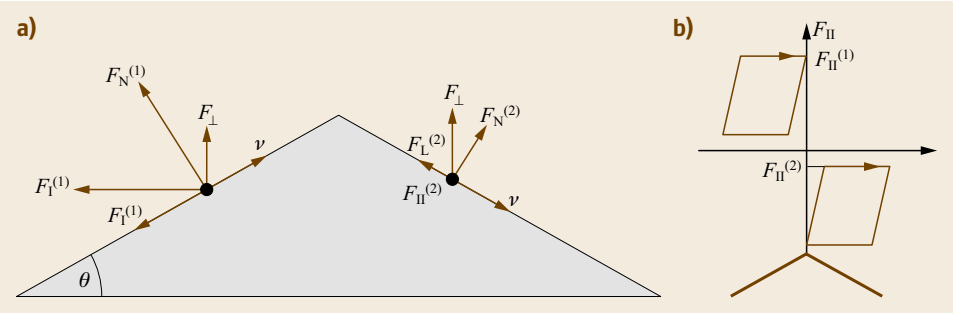


Fig. 29.6 **(a)** Forces acting on an FFM tip sliding on the grating shown in Fig. 29.5; **(b)** friction loops acquired on the two faces

The quality of the force sensor's electrical signal can seriously deteriorate when it is transferred out of the vacuum chamber. Low noise and high bandwidth can be preserved using a preamplifier in the vacuum. Again, the choice of materials for printing and devices is limited by the need for low vapor pressure. Stronger heating of the electrical circuitry in vacuum, therefore, may be needed.

A typical atomic force microscope (AFM) used in UHV is shown in Fig. 29.7. The housing (1) contains the light source and a set of lenses that focus the light onto the cantilever. Alternatively, the light can be guided via an optical fiber into the vacuum. By using light emitting diodes with low coherency it is possible to avoid interference effects often found in instruments that use a laser as the light source. A plane mirror fixed on the spherical rotor of a first stepping motor (2) can be rotated around vertical and horizontal axes in order to guide the light beam onto the rear of the cantilever, which is mounted on a removable carrier plate (3). The light is reflected off the cantilever toward a second motorized mirror (4) that guides the beam to the center of the quadrant photodiode (5), where the light is then converted into four photocurrents. Four preamplifiers in close vicinity to the photodiode allow low-noise measurements with a bandwidth of 3 MHz.

The two motors with spherical rotors, used to realign the light path after the cantilever has been exchanged, work as *inertial stepping motors*: the sphere rests on three piezoelectric legs that can be moved in small amounts tangentially to the sphere. Each step of the motor consists of the slow forward motion of two legs followed by an abrupt jump backwards. During the slow forward motion, the sphere follows the legs due to friction, whereas it cannot follow the sudden jump due to its inertia. A series of these tiny steps rotates the sphere macroscopically. The sample, which is also placed on an exchangeable carrier plate, is mounted at the end of a tube scanner (6), which can move the sample in three dimensions over several micrometers. The whole scanning head (7) is the slider of a third inertial stepping motor for coarse positioning of the

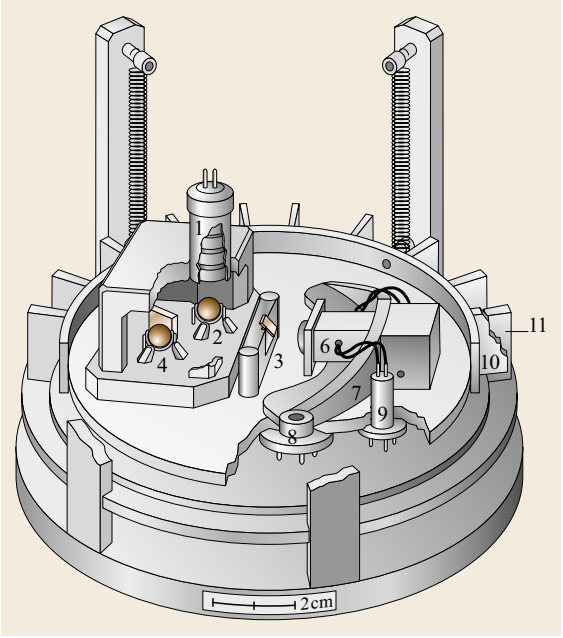


Fig. 29.7 Schematic view of the UHV-FFM realized at the University of Basel. *1*: light source, *2, 4*: mirrors, *3*: cantilever holders, *5*: photodetector, *6*: scanner, *7*: slider, *8*: driving piezo, *9*: fixed post, *10, 11*: eddy current damping (after [29.17])

sample. It rests with its flat and polished bottom on three supports. Two of them are symmetrically placed piezoelectric legs (8), whereas the third central support is passive. The slider (7) can be moved in two dimensions and rotated about a vertical axis by several millimeters (rotation is achieved by antiparallel operation of the two legs). The slider is held down by two magnets, close to the active supports, and its travel is limited by two fixed posts (9) that also serve as cable attachments. The whole platform is suspended by four springs. A ring of radial copper lamellae (10), floating between a ring of permanent magnets (11) on the base flange, acts to efficiently damp eddy currents.

29.1.3 Energy Dissipation in Noncontact Atomic Force Microscopy

Historically, the measurement of energy dissipation induced by tip–sample interaction has been the domain of FFM. In a typical stick–slip experiment, the energy dissipated in a single atomic slip event is of the order of 1 eV and the lateral resolution is limited by the contact area, which is usually formed by several atoms. These limitations can be overcome using noncontact (NC) AFM. In this case, the tip oscillates with constant amplitude A of typically 1−20 nm at the eigenfrequency f of the cantilever, which shifts by Δf due to interaction forces between tip and sample. In NC-AFM, power dissipation P_0 caused by internal friction of the freely oscillating cantilever is observed. This dissipation is proportional to the eigenfrequency f_0 and to the square of the amplitude A and is inversely proportional to the known Q value of the cantilever. When the tip–sample distance is reduced, the tip interacts with the sample and therefore additional damping of the oscillation is encountered. The extra dissipation P_{ts} caused by the tip–sample interaction can be calculated from the excitation signal A_{exc} [29.18]. The observed energy losses per oscillation cycle (in the order of 100 meV) [29.20] are comparable to the 1 eV energy loss in the contact slip process observing that in that a few atoms are involved in the contact. The origins of dissipation in NC-AFM are manifold: phonon excitation, Joule dissipation, and van der Waals friction, as schematized in Fig. 29.8.

A recent improvement in force measurements on the nanoscale using NC-AFM came from the introduction of q-plus sensors and vertically suspended ultrasensitive cantilevers as scanning probes (Fig. 29.9). Both configurations, namely *tuning fork AFM* and *pendulum AFM* hold two opposite positions in terms of force sensitivity and amplitude stability. The tuning fork AFM is in fact a combination of STM (scanning tunneling microscope) and AFM and is dedicated to performing high-resolution imaging. The measurements are typically conducted at a temperature $T = 5$ K and in UHV. STM characterizations are done in constant-current mode with a bias voltage applied to the tungsten tip whereas AFM is performed in noncontact mode using a phase lock loop (PLL). The spring constant k of the q-plus sensor is large, typically of about 2 kN/m. Although not extremely force sensitive the probe provides outstanding atomic-scale resolution due to small amplitudes of oscillation. Amplitude stability and high stiffness of the q-plus sensor also allow one to perform quantitative measurements of molecular friction. This possibility relies on the fact that the variation Δf of the natural contact resonance f_0 of the cantilever are related to the *normal* force F_N on the tip by

$$\frac{\Delta f}{f_0} = -\frac{1}{2k}\frac{dF_N}{dz} ,$$

(provided that the oscillation amplitude is small compared to the characteristic decay length of F_N) [29.21].

The pendulum UHV-AFM is a unique design dedicated to the measurement of extremely small forces. In this geometry the cantilever is hovering perpendicularly to the sample surface, shearing a nanometer vacuum gap between tip and sample. In this way, the

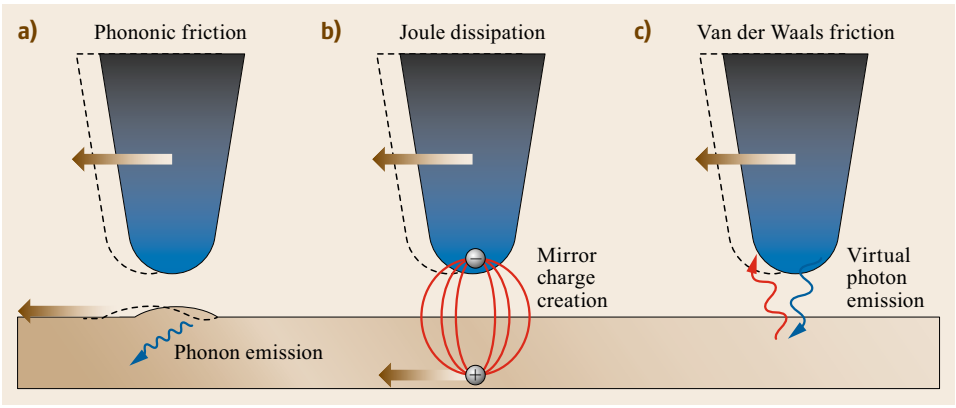


Fig. 29.8 **(a)** Phononic friction occurs when due to attractive tip–sample interactions the surface is deformed following the tip movement and, the strain energy is eventually dissipated into phonons. **(b)** Joule dissipation takes place when noncompensated charges beneath the surface move through a resistive medium. The tip oscillation causes the displacement current and Joule heating. **(c)** The van der Waals friction is initiated by the emission of a virtual photon towards the surface. The reflected photon is adsorbed back in the tip. The friction rises because the two photons are Doppler shifted in the moving reference frame of the oscillating tip (after [29.19])

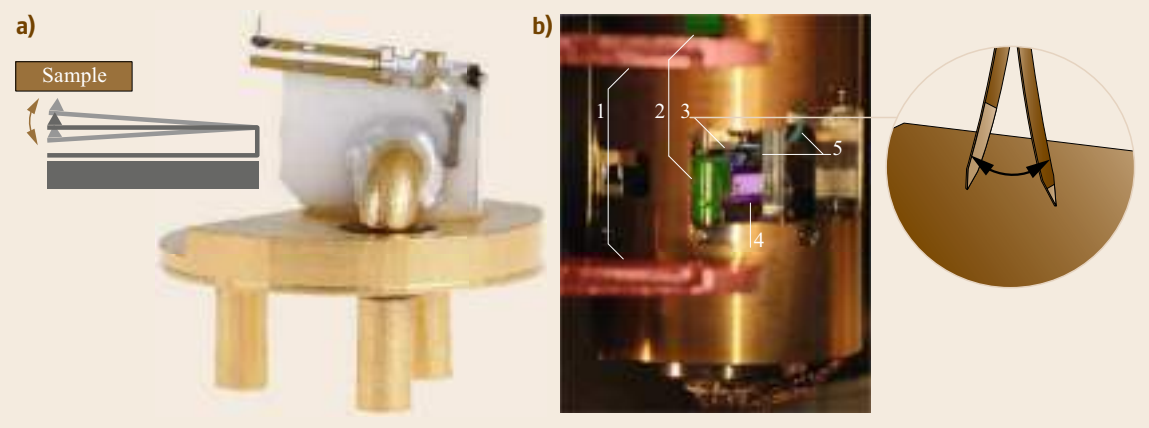


Fig. 29.9 **(a)** Schematic drawing of a tuning fork sensor in q-plus configuration. **(b)** Pendulum AFM configuration where 1 is the microscope, 2 are the mirrors of the beam deflection detection system, 3 is the cantilever suspended in pendulum geometry, 4 is the sample holder, and 5 is the STM line

so-called *noncontact friction* can be measured. In the pendulum AFM very soft cantilevers with spring constants k in the order of 10^{-5} to 10^{-3} N/m are used, which avoids snapping into contact with the sample. In contrast to contact friction measurements, where the slider is usually moved at constant velocity, in the pendulum AFM the tip oscillates above the sample surface with a constant frequency. The pendulum AFM operates at cryogenic temperatures and typical quality factors are in the order of $Q \approx 10^5$–10^6. Such high Q together with extremely small k values (both at the limit for silicon cantilevers) imply that the minimal detectable energy dissipation is a few orders of magnitude lower as compared to standard AFM configuration.

29.1.4 Friction Force Microscopy in Water

A robust alternative to UHV is measuring friction with the tip completely embedded in water or aqueous solution. Here, capillary condensation effects are completely removed and lattice resolution is easily obtained in standard FFM maps. This possibility is achieved using dedicated liquid cells, and is particularly interesting for studying dissolution and growth of mineral surfaces, but also the molecular arrangement of self-assembled monolayers when UHV-AFM is not available. A detailed molecular dynamics (MD) investigation of atomic-scale stick–slip on a (hydrophobic) graphite surface in water, considering collisions of single H_2O molecules with the probing tip, has recently been presented by *Vilhena* et al. [29.22]. The main conclusion is that, for small forces in the nN regime, friction in water is almost indistinguishable from friction in UHV except for a slightly enhanced background noise due to molecular collisions. Although this result cannot be generalized to hydrophilic materials, it is reasonable to expect that tip-water collisions do not play a pivotal role also in that case.

29.2 The Prandtl–Tomlinson Model

The contrast obtained in FFM images of crystal surfaces on the atomic scale is ultimately determined by the *stick–slip* motion of the probing tip. This process was described by *Prandtl* well before modern scanning probe techniques were introduced [29.23] and, for historical reasons, it is usually known as the Prandtl–Tomlinson (PT) mechanism. In combination with Kramer's theory for thermally activated processes, the PT model also explains the temperature and velocity dependence of atomic-scale friction.

29.2.1 One-Dimensional Prandtl–Tomlinson Model

In the PT model, the motion of the tip is determined by both the repulsive interaction with the atomic lattice of the surface and the elastic deformations of the driving support and the contact region. The shape of the tip–surface potential $U(\mathbf{r})$ depends on several factors, such as the chemical composition of the materials in contact and the atomic arrangement on the crystal

surface and the tip apex. For the sake of simplicity, we will start the analysis in the one-dimensional case considering a simple sinusoidal profile with a periodicity a and an amplitude U_0 increasing with the applied load. If the cantilever support moves with a constant velocity v along the x-direction, as usually occurs in FFM measurements, the total energy of the system is

$$U_{\mathrm{tot}}(x,t) = -U_0 \cos\frac{2\pi x}{a} + \frac{1}{2}k_{\mathrm{eff}}(vt-x)^2 \,, \quad (29.8)$$

where k_{eff} is an effective lateral stiffness (Sect. 29.5.3). Figure 29.10a shows the energy profile $U_{\mathrm{tot}}(x,t)$ at two different times t. When $t=0$, the tip is pinned in the absolute minimum of U_{tot}. This minimum becomes less and less pronounced with time, and, at a critical time $t=t^*$, this position becomes unstable and the tip suddenly slips into the next minimum of the potential profile. Correspondingly, the lateral force $F_{\mathrm{L}} = k(vt-x)$ takes the value [29.24]

$$F_{\mathrm{s}} = \frac{k_{\mathrm{eff}}a}{2\pi}\sqrt{\eta^2-1} \,, \quad (29.9)$$

where $\eta = 4\pi^2 U_0/(k_{\mathrm{eff}}a^2)$. In this way, (29.9) defines the *static friction* force acting on the tip, and stick–slip is observed only if $\eta > 1$, i. e., when the system is not too stiff or the tip–surface interaction is not too weak.

The lower curve in Fig. 29.10b shows the variation of the lateral force F_{L} when the cantilever is moved to the right. If, at a certain point, the direction of motion is suddenly inverted, the force follows the profile corresponding to the upper part of the curve. The area of the *friction loop* obtained by scanning back and forth gives the total energy dissipated in the process. Conversely, when $\eta < 1$, the tip slides continuously on the surface and the lateral force oscillates between negative and positive values. Elastic instabilities are suppressed, which leads to disappearance of lateral force hysteresis (so-called *superlubricity*: see also Sect. 29.2.3).

The average *kinetic friction* can be directly estimated from the energy drop ΔU accompanying the slip as $F_{\mathrm{k}} = \Delta U/a$. It increases linearly with the parameter η when the latter is high enough and it approaches the superlubric transition as $(\eta-1)^2$ [29.25].

Note that all previous considerations are valid in the so-called *quasistatic* limit, assuming that the tip, once unpinned from the surface, immediately rebinds to it. This approximation is acceptable if the damping coefficient γ describing dissipative processes due to tip-substrate coupling is high enough. If this is not the case *long* jumps of several lattice constants are possible and, if γ is very low, a chaotic regime of motion is established [29.25].

29.2.2 Two-Dimensional Prandtl–Tomlinson Model

In two dimensions, (29.8) is replaced by

$$U_{\mathrm{tot}}(\boldsymbol{r},t) = U(r) + \frac{k_{\mathrm{eff}}}{2}(\boldsymbol{v}t-\boldsymbol{r})^2 \,, \quad (29.10)$$

where $\boldsymbol{r} \equiv (x,y)$ gives the tip position and the velocity vector $\boldsymbol{v}$ is arbitrarily oriented on the surface (note that $\boldsymbol{v} \neq \mathrm{d}\boldsymbol{r}/\mathrm{d}t$). Figure 29.11a shows the total energy corresponding to a periodic potential of the form

$$U(x,y,t) = -\frac{U_0}{2}\left(\cos\frac{2\pi x}{a} + \cos\frac{2\pi y}{a}\right) + U_1 \cos\frac{2\pi x}{a}\cos\frac{2\pi y}{a} \,. \quad (29.11)$$

In general, the equilibrium condition becomes

$$\nabla U_{\mathrm{tot}}(\boldsymbol{r},t) = \nabla U(\boldsymbol{r}) + k_{\mathrm{eff}}(\boldsymbol{r}-\boldsymbol{v}t) = 0 \quad (29.12)$$

and the stability is described by the Hessian matrix

$$H = \begin{pmatrix} \frac{\partial^2 U}{\partial x^2} + k_{\mathrm{eff}} & \frac{\partial^2 U}{\partial x \partial y} \\ \frac{\partial U}{\partial y \partial x} & \frac{\partial^2 U}{\partial y^2} + k_{\mathrm{eff}} \end{pmatrix} . \quad (29.13)$$

When both eigenvalues $\lambda_{1,2}$ of the Hessian are positive, the position of the tip is stable. Figure 29.11b shows these regions for a potential of the form (29.11).

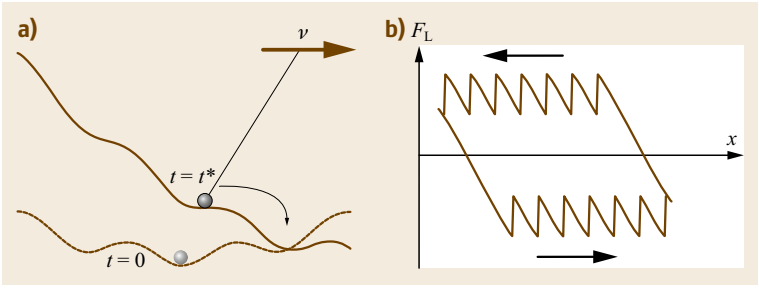


Fig. 29.10 **(a)** Energy profile experienced by the FFM tip (*black circle*) at $t=0$ (*dotted line*) and $t=t^*$ (*continuous line*). **(b)** Friction loop obtained by scanning back and forth in the one-dimensional (1-D) PT model. The effective spring constant k_{eff} is the slope of the sticking part of the loop (if $\eta \gg 1$)

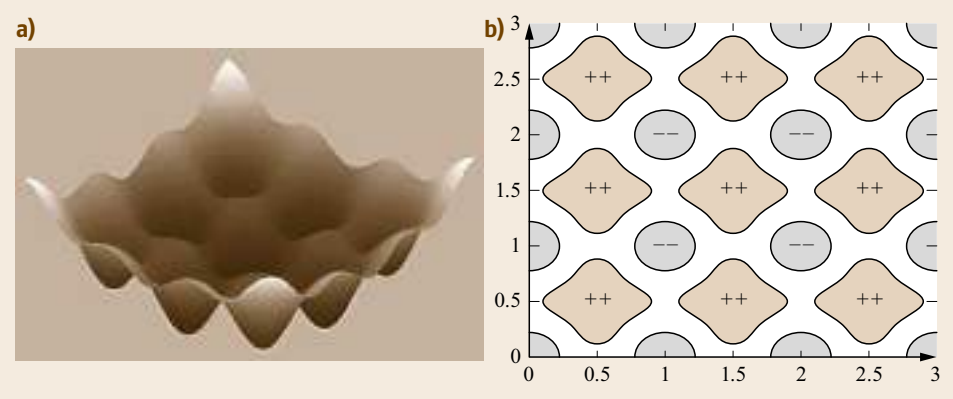


Fig. 29.11 **(a)** Energy landscape experienced by the FFM tip in two-dimensional (2-D) PT model. **(b)** Regions on the tip plane labeled according to the signs of the eigenvalues of the Hessian matrix (after [29.26])

The tip follows the cantilever adiabatically as long as it remains in the (++)-region. When the tip reaches the border of this region, it slips into a neighboring (++)-region and, as a result of several repetitions of this process, the tip is expected to follow a zigzag pathway. Analytical expressions for the static friction F_s and the kinetic friction F_k can be derived on simple square [29.27] and hexagonal [29.28] potentials.

29.2.3 Friction Between Atomically Flat Surfaces

Thus far we have implicitly assumed that the tip is ideally terminated by only one atom. It is also instructive to consider the case of a periodic surface sliding on another periodic surface. In the *Frenkel–Kontorova–Tomlinson* (FKT) *model*, the atoms of one surface are harmonically coupled with their nearest neighbors. In this context, the role of *commensurability* is essential. In 1-D, *Weiss* and *Elmer* predicted that friction should decrease with decreasing commensurability, the minimum friction being reached when the ratio between the two lattice constants is equal to the *golden mean* [29.30]

$$\bar{z} = 1 + \frac{1}{1 + \frac{1}{1+\ldots}} = \frac{\sqrt{5}-1}{2} \,. \tag{29.14}$$

In 2-D, *Gyalog* and *Thomas* studied the case of two square lattices with $a_1 = a_2$, and a misalignment given by an angle θ (Fig. 29.12a). When the sliding direction changes, friction also varies from a minimum value (corresponding to the sliding angle $\varphi = \theta/2$) to a maximum value (which is reached when $\varphi = \theta/2 + \pi/4$; Fig. 29.12b). The misfit angle θ is related to the commensurability. Since the misfit angles that give rise to commensurate structure form a dense subset, the dependence of friction on θ is expected to be discontinuous as indeed confirmed by additional numerical simulations.

In general, if the contact is incommensurate, the lateral forces of single atoms tend to cancel out if the load decreases or the stiffness increases and a transition from stick–slip to negligible friction is expected [29.31]. This situation may be called *superlubricity* or, as suggested by *Müser*, *structural lubricity* [29.32].

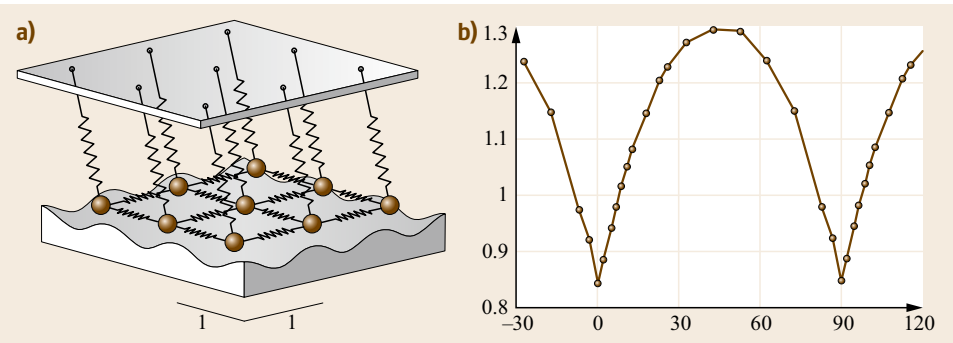


Fig. 29.12 **(a)** The FKT model in 2-D. **(b)** Friction as a function of the sliding angle φ in the 2-D FKT model (after [29.29])

29.3 Friction Experiments on the Atomic Scale

In the first lattice-resolved FFM measurements, *Mate* et al. scanned a graphite surface in ambient conditions with normal forces up to few tens of μN [29.1]. According to continuum mechanics, these values correspond to contact diameters of about 100 nm suggesting that a graphite flake may have detached from the surface and adhered to the tip. A rather linear load dependence of friction was also observed with a small friction coefficient $\mu = 0.01$. More precise and reproducible results are obtained in UHV, where the surface cleanness is better controlled than in ambient conditions. An atomic-scale friction map, acquired with the setup described in Sect. 29.1.2, is shown in Fig. 29.13. Here the lateral force F_L recorded on a sodium chloride surface periodically increases up to values $F_s \approx 0.4$ nN, when a slip event occurs and the tip promptly rebinds to the surface. The process is repeated several times, and the resulting FFM map perfectly reproduces the symmetry of the surface lattice. Using (29.9) the values $U_0 \approx 0.22$ eV and $k \approx 1$ N/m are estimated for the amplitude of the interaction potential and the effective lateral stiffness of the system. Note that the duration of the slip event, which is related to damping coefficient γ of the tip oscillations, remains beyond the time resolution of state-of-the-art FFMs [29.33]. However, this parameter may be indirectly estimated from statistical analysis of the slip length at different loading, as suggested by *Roth* et al. [29.34]. By applying subnanonewton loads to an NaCl surface, *Socoliuc* et al. were able to observe the transition from stick–slip to continuous sliding discussed in Sect. 29.2.1 [29.24] In other experiments, the transition was induced *dynamically* by exciting flexural or torsional resonance modes of the cantilever in contact with the surface [29.35, 36].

Note that no individual defects are seen in Fig. 29.13. This is expected if the contact between tip and surface is formed by several atoms. The availability of ultrasharp tips with curvature radius below 2 nm has recently allowed this limitation to be overcome. High-resolution FFM imaging of the moiré pattern formed by an ultrathin KBr film epitaxially grown on NaCl is demonstrated in Fig. 29.14a. This pattern results from the lattice mismatch between the two surfaces, with a substrate versus adsorbate lattice constant ratio of about 6 : 7. It is not observed if the load is close to the pull-off value. Hexagonal moiré patterns have also been reported by *Filleter* et al. using FFM on a graphene film on SiC(0001) [29.38]. Reconstruction domains on semiconducting surfaces [29.39], and friction variations across surface step edges of sodium chloride and germanium (Fig. 29.14b) have also been observed with *quasiatomic* resolution using FFM.

The anisotropy of friction and the zigzag walk of the tip predicted by the 2-D PT model (Sect. 29.2.2) was first observed by *Fujisawa* et al. [29.42] and *Kawakatsu* et al. [29.43] on mica and on MoS_2 using original 2-D FFM with two laser beams and two quadrant photodetectors. With another dedicated *Tribolever Dienwiebel* et al. were able to track friction forces with arbitrary scan orientations [29.44]. In this way, stick–slip and a significant energy dissipation were only observed at angles of rotation corresponding to a commensurate contact between a graphite surface and graphene flakes picked up by the tip. More recently, friction anisotropy has also been demonstrated by *Fessler* et al. on organic crystal surfaces [29.45]. The chemical structure of the used benzylammonium-terminated crystalline surface is presented in Fig. 29.15a. The FFM maps in UHV revealed a significant contrast between different molecular orientations yielding molecular rows which differ from their neighboring ones in dependence of the applied load (Fig. 29.15b). These maps could be repro-

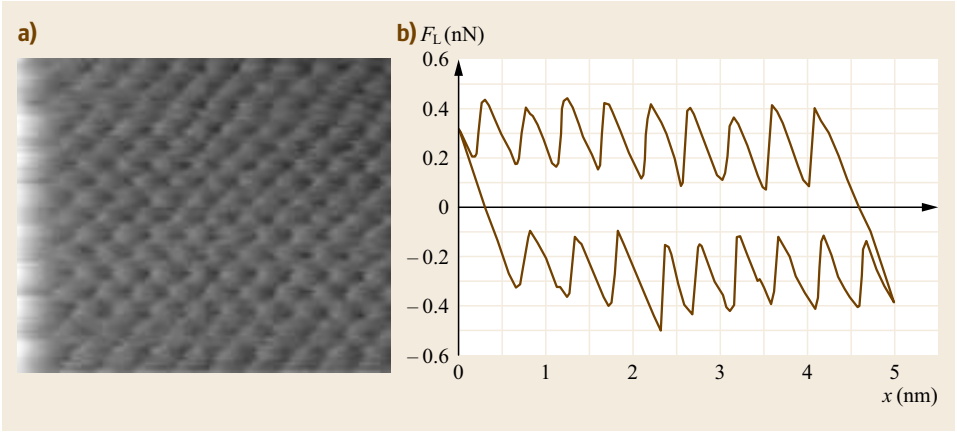


Fig. 29.13 **(a)** FFM map of a sodium chloride surface in UHV and **(b)** corresponding friction loop (Reprinted from [29.37]. Copyright 2000 by the American Physical Society)

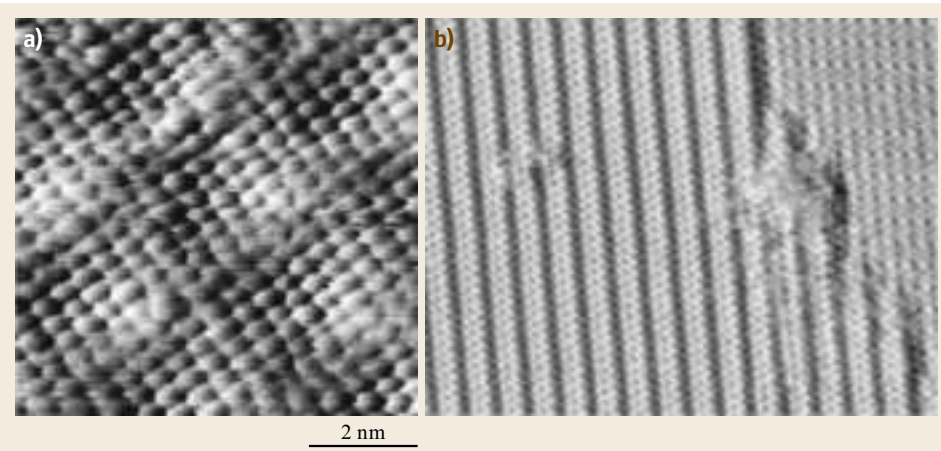


Fig. 29.14a,b FFM maps **(a)** of a moiré superstructure formed by potassium bromide on a sodium chloride surface (Reprinted from [29.40] with permission by the American Physical Society) and **(b)** of two terraces of a reconstructed germanium surface rotated by 90° with respect to each other (Reprinted from [29.41] with permission by the American Physical Society)

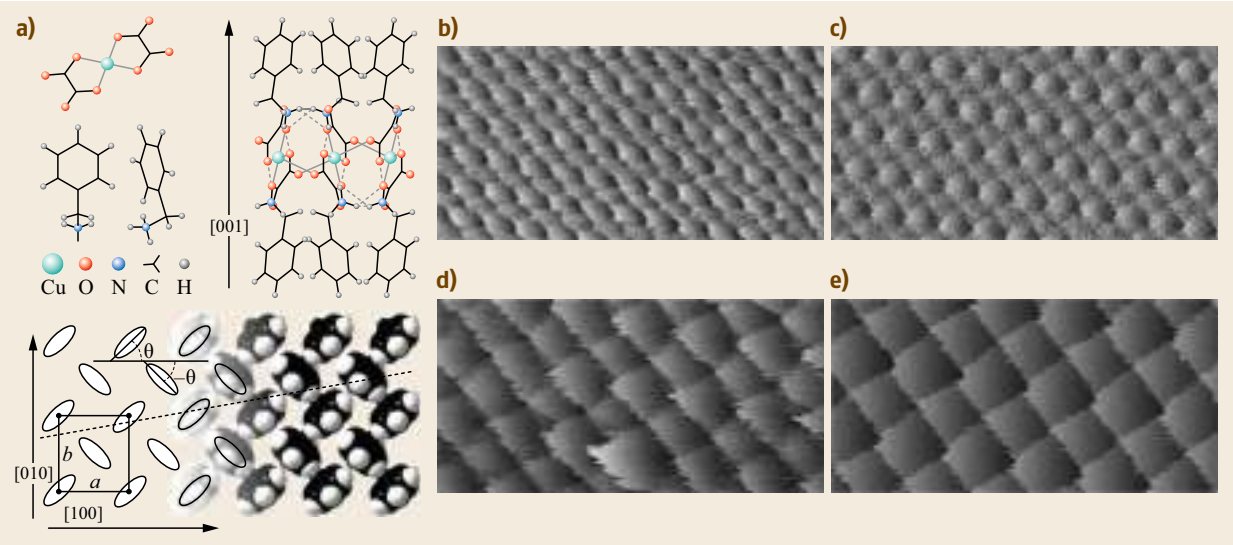


Fig. 29.15 **(a)** The $[Cu(C_2O_4)_2]^{2-}$ complex builds up an extended two-dimensional structure surrounded by the organic BA cations. The organic crystal is built up by multiple stacks of the illustrated layers and the [001] surface shows the different orientations of the BA molecules. The lattice constants are $a = 7.1$ and $b = 8.0$ Å. **(b)** and **(d)** show FFM measurements at different loads (0 and 1.1 nN, respectively). Simulations with the modified potential were done with $\eta = 1$, $c = 0.35$ **(c)** and with $\eta = 3$, $c = 0.1$ **(e)**. (Reprinted from [29.45] with permission of AIP Publishing)

duced introducing the empirical potential

$$U(x, y) = -U_0 \cos \frac{2\pi x}{a} \left(\cos \frac{2\pi y}{b} + c \right)$$

in the 2-D PT model, where a and b are the lattice constants and the parameter c gives rise to two different energy barriers $\Delta E^+ = U_0(1 + c)$ and $\Delta E^- = U_0(1 - c)$. Two examples of measurements and corresponding simulations are shown in Fig. 29.15b–e. The good agreement between them is demonstrated by the alternate contrast of a row with almost smooth sliding and the neighboring ones with considerably larger friction corresponding to the energy barriers ΔE^- and ΔE^+, respectively.

The potential of liquid environments for high-resolution FFM is also well established. Figure 29.16 shows

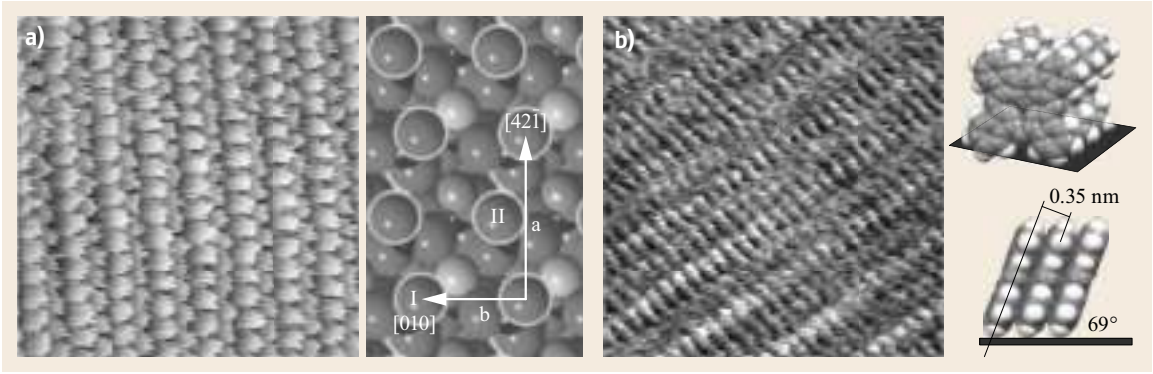


Fig. 29.16a,b FFM maps of **(a)** a dolomite cleavage surface (after [29.46]) and **(b)** of a Cu phthalocyanine ultrathin film grown on it (after [29.47])

a friction map of a dolomite cleavage surface and of a thin film of Cu phthalocyanine (PC) molecules previously deposited on it. On the dolomite surface two pinning sites per unit cell are resolved, corresponding to protruding carbonate groups. The contrast of each group varies with the scan direction, which can also be explained by the 2-D PT model. A comparison with the structure of PC crystals, as obtained by X-ray diffraction, clearly reveals that the molecules are stacked perpendicular to the substrate. This information could not be obtained in ambient conditions, where, due to capillary condensation, lattice resolution was not achieved.

29.4 Thermal Effects on Atomic Friction

Although the PT model gives a good interpretation of the basic mechanism of the atomic stick–slip discussed in Sect. 29.3, it cannot explain some important features observed in the friction maps. For example, the peaks in the sawtooth profile in Fig. 29.13b have different heights, which is in contrast to the result in Fig. 29.10b. Another effect is observed if the scan velocity v is varied: the mean friction force increases with the logarithm of v [29.37]. This effect cannot be understood within the mechanism discussed in Sect. 29.2 without further assumptions.

29.4.1 The Prandtl–Tomlinson Model at Finite Temperature

Let us focus again on the energy profile discussed in Sect. 29.2.1. For the sake of simplicity, we will suppose that $\eta \gg 1$. At a given time $t < t^*$, the tip jump is prevented by the energy barrier $\Delta E = U(x_{\max}, t) - U(x_{\min}, t)$, where $x_{\max}$ corresponds to the first maximum observed in the energy profile and $x_{\min}$ is the actual position of the tip (Fig. 29.17a). The quantity ΔE decreases with time or, equivalently, with the frictional force F_{L} until it vanishes when $F_{\mathrm{L}} \approx F^*$ (Fig. 29.17b).

At finite temperature T, the lateral force required to induce a jump is lower than F^*. To estimate the most probable value of F_{L} at this point, we first consider the probability p that the tip does *not* jump. The probability p changes with time t according to the master equation

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = -f_0 \exp\left(-\frac{\Delta E(t)}{k_{\mathrm{B}}T}\right) p(t)\,, \qquad (29.15)$$

where f_0 is a characteristic attempt frequency in the order of fewtens of kHz (for typical FFM measurements on the atomic scale). Note that the probability of a reverse slip is neglected, since in this case the energy barrier that must be overcome is much higher than ΔE. Assuming a linear dependence of ΔE on the lateral force, a logarithmic dependence of friction on the scan velocity is obtained

$$F_{\mathrm{L}}(v) = F^* - \frac{k_{\mathrm{B}}T}{\lambda} \ln \frac{v_{\mathrm{c}}}{v}\,, \qquad (29.16)$$

where λ is in the order of the lattice constant.

However, the linear approximation for ΔE does not hold when the slip occurs very close to the critical point $x = x^*$, which is the case at high velocities. In this case thermal activation does not play any role and friction (due to stick–slip motion) becomes constant, in agreement with the classical Coulomb's law of friction. We

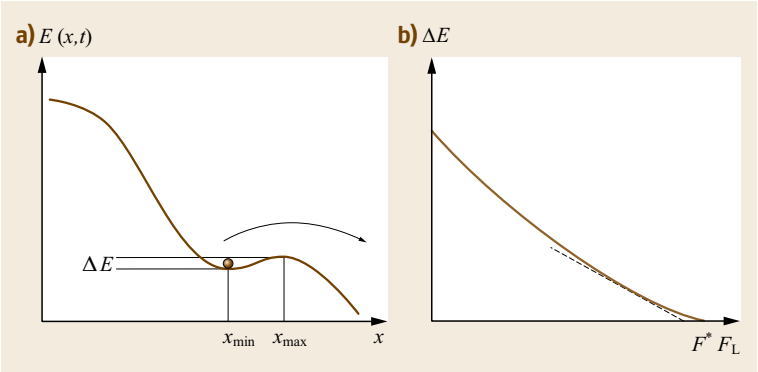


Fig. 29.17 **(a)** Energy barrier that hinders the tip jump in the PT model. **(b)** Energy barrier ΔE as a function of the lateral force F_L. The *dashed line* close to the critical value corresponds to the linear approximation (29.16)

should also mention that a better approximation for the energy barrier is $\Delta E \propto (F^* - F_L)^{3/2}$ [29.48]. The corresponding relations for the velocity dependence of friction are more complex and difficult to distinguish from the experimental results. Furthermore, the logarithmic law (29.15) becomes invalid at very low values of v. In this case reverse slip cannot be excluded, and, as a result, static friction is found to be proportional to v [29.49]. At room temperature such a *thermolubric* regime would be observed for unrealistic velocity values well below 1 pm/s (but the critical value increases rapidly with temperature). The thermally activated PT model has also been discussed in 2-D [29.50].

29.4.2 Velocity and Temperature Dependence of Atomic-Scale Friction

The first velocity-dependent atomic-scale FFM measurements in UHV were performed by *Bennewitz* et al. [29.51] and *Gnecco* et al. [29.37] with surfaces on copper and sodium chloride; in both cases a logarithmic increase of friction with the sliding velocity was observed up to $v < 1\,\mu$m/s in perfect agreement with (29.16). It is worth noting that a similar logarithmic dependence was also reported by *Bouhacina* et al. in polymer layers grafted on silicon oxide and interpreted as the result of stress-aided thermal activation of viscous flow [29.52]. We should also emphasize that the so-called *velocity strengthening* of friction can be masked by *velocity weakening* in FFM measurements performed in a humid environment. Here the nucleation of capillary bridges can be assisted by thermal activation and lead to friction forces logarithmically increasing with the sliding velocity [29.53].

Thus far we have used thermal activation to explain the velocity dependence of friction. The same mechanism also predicts that friction should change with temperature. The master equation (29.15) shows indeed that the probability of a tip slip is reduced at low temperatures T until it vanishes when $T = 0$. In this case thermal activation is excluded, and the lateral force F_L is equal to F^*, independent of the scanning velocity v.

Zhao et al. [29.54] observed that friction on graphite decreases as $1/T$ over a wide temperature range (140–750 K), supporting the hypothesis of thermal activation of the stick–slip process. However, it was only recently that the group of *Schirmeisen* reported atomic-scale FFM measurements in UHV at different temperatures [29.55]. When silicon, SiC, ionic crystals, and graphite surfaces were cooled from room temperature to cryogenic conditions, a good agreement with the thermally activated PT model was found down to a peak or a plateau, appearing between 50 and 200 K. Below these values, the friction was found to decrease with temperature, which the authors attributed to the competition between thermally activated rupture and formation of chemical bonds [29.56].

29.5 Friction on the Nanometer Scale

Friction is ultimately related to the real shape of the contact between the sliding surfaces. On the macroscopic scale, the contact between two bodies is studied within the context of continuum mechanics, which is based on the elasticity theory developed by Hertz in the nineteenth century. Various FFM experiments have shown that continuum mechanics is still valid down to contact areas just a few nanometers in size. Only when contact is reduced to a few atoms does the continuum frame become unsuitable, and other approaches like molecular dynamics become necessary.

29.5.1 Continuum Mechanics of Single Asperities

The lateral force F_L between two surfaces in reciprocal motion depends on the size of the real area of contact A, which can be a few orders of magnitude smaller than the apparent area of contact. The simplest assumption is that friction is proportional to A; the proportionality factor is called the *shear strength* τ [29.57]

$$F_L = \tau A \,. \tag{29.17}$$

For plastic deformation, the asperities are compressed until the pressure p equals a certain yield value p^*. The resulting contact area is thus $A = F_N/p^*$, and the well-known Amonton's law is obtained: $F_L = \mu F_N$, where $\mu = \tau/p^*$ is the *coefficient of friction*. The same idea can be extended to contacts formed by many asperities, and it leads again to Amonton's law. The simplicity of this analysis explains why most friction processes were related to plastic deformation for a long time. Such a mechanism, however, should provoke quick disruption of surfaces, which is not observed in practice.

Elastic deformation can be easily studied in the case of a sphere of radius R pressed against a flat surface. In this case, the contact area is

$$A(F_N) = \pi \left(\frac{R}{K}\right)^{2/3} F_N^{2/3} \,, \tag{29.18}$$

where $K = 3E^*/4$ and E^* is an effective Young's modulus, related to the Young's moduli (E_1 and E_2) and the Poisson numbers (ν_1 and ν_2) of the sphere and plane, by the following relation [29.58]

$$\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \,. \tag{29.19}$$

The result $A \propto F_N^{2/3}$ contrasts with Amonton's law. However, a linear relation between F_L and F_N can be obtained for contacts formed from several asperities in particular cases. For example, the area of contact between a flat surface and a set of asperities with an exponential height distribution and the same radius of curvature R depends linearly on the normal force F_N [29.59]. The same conclusion holds approximately even for a Gaussian height distribution. However, the hypothesis that the radii of curvature are the same for all asperities is not realistic. A general model was recently proposed by *Persson*, who analytically derived the proportionality between contact area and load for a large variety of elasto-plastic contacts formed by surfaces with arbitrary roughness [29.60]. However, the discussion of this theory is not straightforward and goes beyond the purpose of this section.

Further effects are observed if adhesive forces between the asperities are taken into account. If the range of action of these forces is smaller than the elastic deformation, (29.18) is extended to the Johnson–Kendall–Roberts (JKR) relation

$$A(F_N) = \pi \left(\frac{R}{K}\right)^{2/3} \left(F_N + 3\pi\gamma R + \sqrt{6\pi R F_N + (3\pi\gamma R)^2}\right)^{2/3} , \tag{29.20}$$

where γ is the surface tension of the sphere and plane [29.61]. The real contact area at zero load is finite and the sphere can be detached only by pulling it away with a certain force. This is also true in the opposite case, in which the range of action of adhesive forces is larger than the elastic deformation. In this case, the relation between contact area and load takes the simple form

$$A(F_N) = \pi \left(\frac{R}{K}\right)^{2/3} (F_N - F_{off})^{2/3} \,, \tag{29.21}$$

where F_{off} is the negative load required to break the contact. The *Hertz-plus-offset* relation (29.21) can be derived from the Derjaguin–Muller–Toporov (DMT) model [29.62]. To discriminate between the JKR or DMT models, *Tabor* introduced a nondimensional parameter

$$\Phi = \left(\frac{9R\gamma^2}{4K^2 z_0^3}\right)^{1/3} , \tag{29.22}$$

where z_0 is the equilibrium distance during contact. The JKR model can be applied if $\Phi > 5$; the DMT model holds when $\Phi < 0.1$ [29.63]. For intermediate values of Φ, the Maugis–Dugdale model [29.64] can reasonably explain experimental results.

29.5.2 Dependence of Friction on Load

The FFM tip represents a single asperity sliding on a surface. The previous discussion suggests a nonlinear dependence of friction on the applied load, provided that continuum mechanics is applicable. *Schwarz* et al. [29.65] observed the Hertz-plus-offset relation (29.21) on graphite, diamond, amorphous carbon, and C_{60} in an argon atmosphere (Fig. 29.18). In their measurements, they used well-defined spherical tips with radii of curvature of tens of nanometers, obtained by contaminating silicon tips with amorphous carbon in a transmission electron microscope.

A good agreement with JKR theory has been reported in earlier UHV-FFM measurements on ionic crystals, mica, and metals [29.66–68]. In order to correlate the lateral and normal forces with improved statistics, *Meyer* et al. [29.66] applied an original 2-D histogram technique (Fig. 29.19). *Carpick* et al. extended the JKR relation (29.19) to include nonspherical tips [29.67]. In the case of an axisymmetric tip profile $z \propto r^{2n}$ $(n > 1)$, it can be proven analytically that the increase in friction becomes less pronounced with increasing n.

29.5.3 Estimation of the Contact Area

In contrast to other tribological instruments, such as the surface force apparatus [29.69], the contact area cannot be measured directly by FFM. Indirect methods are provided by contact stiffness measurements. The contact between the FFM tip and the sample can be modeled by a series of two springs (Fig. 29.20). The effective constant k^z_{eff} of the series is given by

$$\frac{1}{k^z_{\text{eff}}} = \frac{1}{k^z_{\text{contact}}} + \frac{1}{c_{\text{N}}} , \qquad (29.23)$$

where c_{N} is the normal spring constant of the cantilever and k^z_{contact} is the normal stiffness of the contact. This quantity is related to the radius of the contact area (a) by the simple relation [29.70]

$$k^z_{\text{contact}} = 2aE^* , \qquad (29.24)$$

where E^* is the effective Young's modulus introduced previously. However, typical values of k^z_{contact} are 1–2 orders of magnitude larger than c_{N} and a direct application of (29.23) is difficult. Alternatively, one can measure the normal resonance frequency f_{N} of the cantilever contacting the surface while scanning. *Steiner* et al. reported atomically resolved measurements of f_{N} on a sodium chloride surface [29.71]. Corresponding variations of the lateral force were simultaneously acquired. The normal stiffness k_{N} could be estimated from continuum mechanics relations, which were modified to account for the inclination of the cantilever with respect to the sample surface.

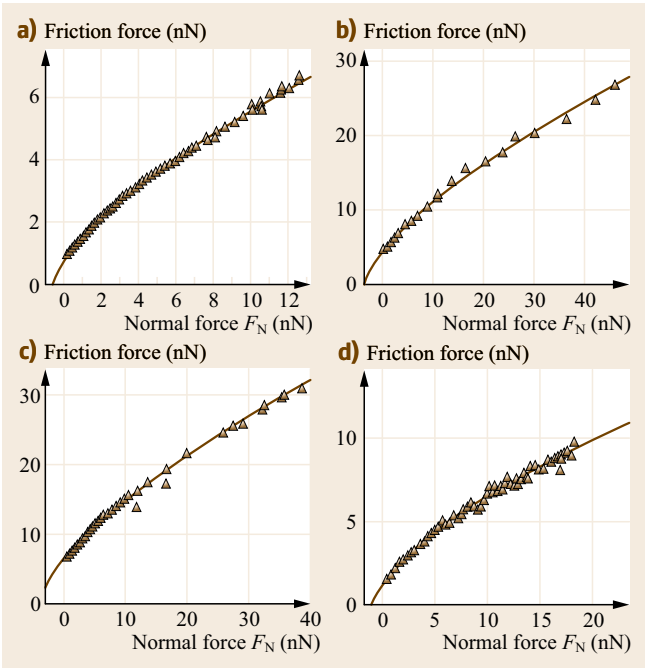


Fig. 29.18a–d Friction versus load curve on amorphous carbon in argon atmosphere. Curves **(a,b)** refer to tips with different radii of curvature (after [29.65])

The *lateral* contact stiffness of the contact between a sphere and a flat surface is [29.70]

$$k^x_{\text{contact}} = 8aG^* , \qquad (29.25)$$

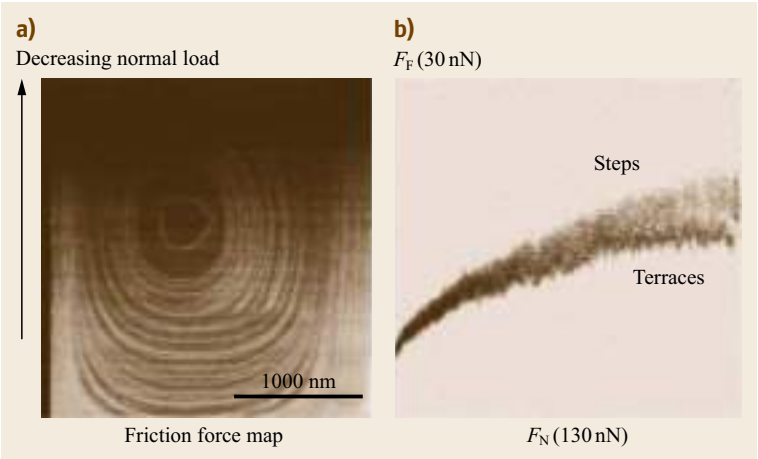


Fig. 29.19 **(a)** Friction force map on NaCl(100). The load is decreased from 140 to 0 nN (jump-off point) during imaging. **(b)** 2-D histogram of **(a)** (after [29.66])

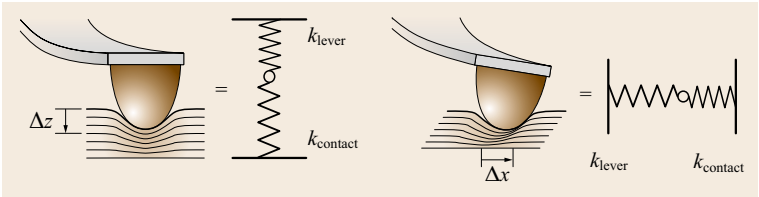


Fig. 29.20 Sketch of normal and lateral stiffness of the contact between tip and surface (after [29.72])

where the effective shear stress G^* is defined by

$$\frac{1}{G^*} = \frac{2-\nu_1^2}{G_1} + \frac{2-\nu_2^2}{G_2} \,. \tag{29.26}$$

G_1, G_2 are the shear moduli of the sphere and the plane, respectively. The contact between the FFM tip and the sample can again be modeled by a series of springs (Fig. 29.20). The effective constant k^x_{eff} of the series is given by

$$\frac{1}{k^x_{\text{eff}}} = \frac{1}{k^x_{\text{contact}}} + \frac{1}{k^x_{\text{tip}}} + \frac{1}{c_{\text{L}}} \,, \tag{29.27}$$

where c_{L} is the lateral spring constant of the cantilever and k^x_{contact} is the lateral stiffness of the contact. As suggested by *Lantz* et al., (29.27) also includes the lateral stiffness of the tip k^x_{tip} which can be comparable to the lateral spring constant [29.73]. The effective spring constant k^x_{eff} is simply given by the slope of the friction loop (Sect. 29.2.1). Once k^x_{contact} is determined, the contact radius a is easily estimated by (29.25).

The lateral stiffness method was applied to contacts between silicon nitride and muscovite mica in air and between $NbSe_2$ and graphite in UHV. The dependences of both the spring constant k^x_{eff} and the lateral force F_{L} on the load F_{N} were explained within the same models (JKR and Maugis–Dugdale, respectively), which confirms that friction is proportional to the contact area for the range of loads applied (up to $F_{\text{N}} = 40\,\text{nN}$ in both experiments).

Enachescu et al. estimated the contact area by measuring the contact conductance on diamond as a function of the applied load [29.74, 75]. Their experimental data were fitted with the DMT model, which was also used to explain the dependence of friction on load. Since the contact conductance is proportional to the contact area, the validity of the hypothesis (29.17) was confirmed again.

29.6 Wear on the Atomic Scale

If the normal force F_{N} applied by the FFM exceeds a critical value, which depends on the tip shape and on the material under investigation, the surface topography is permanently modified by abrasive wear and related processes. On the nanoscale wear may result in patterns with well-defined shapes. Here we will focus on the few experiments which demonstrated the unique ability of the FFM to both scratch and image surfaces down to the atomic scale.

In Fig. 29.21a, a small mound piled up at the end of a groove on KBr(100) is shown at different magnifications. The groove (800 nm long) was produced by repeated scratching with a normal force F_{N} of about 20 nN. When the groove was formed (in about 20 min) a rather continuous increase of friction with time was observed, which is well fitted by the exponential law

$$F_{\text{L}} = F_0 + F_1(1 - \text{e}^{-t/\tau}). \tag{29.28}$$

If $t \gg \tau$ a saturation of the contact area (proportional to friction) is expected, corresponding to the limit value at which the applied load can be perfectly balanced by the tip. In spite of the apparent continuity of the process on long time scales, the friction loops acquired while scraping (not shown) suggest that the atoms (or ion pairs) are removed one by one. Note that this kind of *atom-by-atom attrition* process has been observed recently in TEM images of nanotips getting worn when

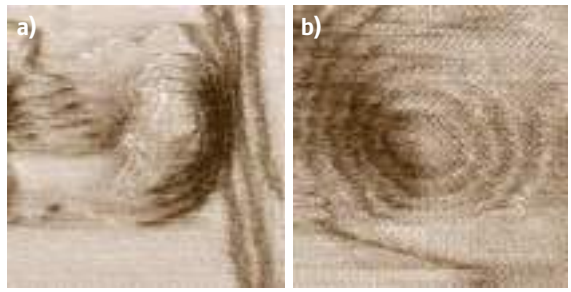


Fig. 29.21 **(a)** Lateral force image acquired at the end of a groove scratched 256 times with a normal force $F_{\text{N}} = 21\,\text{nN}$. Frame sizes: 39 nm. **(b)** Lateral force image of a pit produced with $F_{\text{N}} = 22.8\,\text{nN}$. Frame size: 20 nm (after [29.76])

scraped on harder surfaces [29.77]. No fracture or plastic deformation was found when Si volumes up to few tens of nm^3 were progressively removed from the tip.

Another example is given in Fig. 29.21b, where a pit produced by repeated low-frequency tip oscillations on a 5 nm long line is shown. From a comparison of the number of missing atoms (extracted from the figure) and the energy dissipated during scanning (extracted from the friction loops) it could be estimated that only 30% of the energy dissipation went into rupture of ionic bonds. Furthermore, the damage was observed to increase with the applied load, but did not vary significantly with the scan velocity.

In another study on KBr films on Cu(100) *Filleter* et al. [29.78] reported significant wear at intrinsic step edges of the films, where atomic coordination is lower. In contrast, low friction and no wear were observed across metal steps covered by KBr, which indicates a stabilizing effect of the alkali halide coating on the metal surface.

A different kind of wear is observed on layered materials. *Kopta* et al. [29.79] removed layers from a mica surface by scraping it with large normal forces of few hundreds of nN. Fourier-filtered images acquired on very small areas revealed the different periodicities of the underlying layers, which reflect the complex structure of the substrate. To interpret their results, *Kopta* et al. assumed that wear is initiated by atomic defects. When the defects accumulate beyond a critical concentration, they grow to form the scars observed by AFM. This process can be once again thermally activated. The number of defects created in the contact area $A(F_N)$ is

$$N_{def}(F_N) = t_{res} n_0 A(F_N) f_0 \exp\left(-\frac{\Delta E}{k_B T}\right), \qquad (29.29)$$

where t_{res} is the residence time of the tip, n_0 is the surface density of atoms, and f_0 is the frequency of attempts to overcome the energy barrier ΔE to break a Si−O bond, which depends on the applied load. When the defect density reaches a critical value, a hole is nucleated.

29.7 Noncontact Friction

As discussed in Sect. 29.1.3 forces and dissipation can be measured by recording Δf and A_{exc} simultaneously while scanning a surface by NC-AFM. Earlier experiments showed true atomic contrast in topography in the dissipation signal A_{exc} [29.80–83]. Figure 29.22a shows the border of an ultrathin NaCl island grown on Cu(111).

The strongly enhanced contrast of the step edges and kink sites could be attributed to a slower decay of the electric field and to easier relaxation of the positions of the ions at these locations. The dissipation image shown in Fig. 29.22b was recorded at the same time. Depending on the structure of the tip apex, the topography and damping signal were found to be in phase, partially shifted, or even in antiphase. This is impressively confirmed by the tip change observed in the experimental images in Fig. 29.22b. The dissipation contrast is seriously enhanced, while the topography contrast remains almost unchanged.

More directly related to friction measurements, where the tip is sliding in contact with the sample, are NC-AFM experiments in the pendulum geometry (Sect. 29.1.3). In the first measurements of this kind *Stowe* et al. [29.84] could image the dopant density for n- and p-type silicon samples with 150 nm spatial resolution. The damping turned out to scale as U_{ts}^2 with the tip–sample voltage as expected for electric field Joule dissipation [29.85]. *Stipe* et al. [29.86] measured the noncontact friction between a Au(111) surface and a gold-coated tip using the same setup. They observed the same U_{ts}^2 dependence of the bias voltage and a distance dependence that follows the power law $1/d^n$, where n is between 1.3 and 3. Note that a considerable electric field was present even with zero external bias voltage, which is frequently the case in this kind of measurement.

In order to distinguish phononic and electronic contribution to friction *Kisiel* et al. [29.87] oscillated

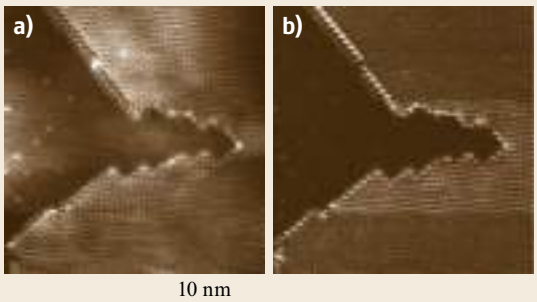


Fig. 29.22 (a) Topography and (b) damping images of an NaCl island on Cu(111). The tip changed after 1/4 of the scan, thereby enhancing the contrast in A_{exc}. After 2/3 of the scan, the contrast in the lower part of the image was restored, indicating a reversible tip change (after [29.82])

a pendulum cantilever over a 150 nm thick niobium film. The measurement was performed at two different temperatures, thus allowing them to investigate both the superconducting state of Nb, where the phononic channel of energy dissipation prevails, as well as the normal metal state where an electronic contribution to friction is superimposed. The dependence of the damping on the temperature is shown in Fig. 29.23a. During the measurement the tip–sample distance was kept constant. The noncontact friction coefficient rose by a factor of three when the critical temperature $T_c = 9.2$ K was approached from below and leveled off in the normal state. The variation of the friction coefficient in the vicinity of T_c is smooth as predicted by the Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity [29.89]. In the superconducting phase, the electrons lose their identity and form Cooper pairs. Therefore, the energy is lost only in the emission of phonons. Close to the transition temperature the gradually growing normal electron population causes acoustic wave attenuation rises due to electron-phonon scattering, which ultimately results in dissipation. Further measurements of voltage V and distance d dependent friction support the hypothesis of separation of phononic and electronic channels.

The idea of replacing a passive substrate with an *active* material undergoing a phase transition for the purpose of friction study was further accomplished on an $SrTiO_3$ sample. According to *Benassi* et al. [29.90] the friction coefficient is peaking when a sliding object undergoes a phase transition. In this context $SrTiO_3$ is a perfect candidate, because of the presence of a structural second-order phase transition at $T = 110$ K. A recent pendulum AFM experiment [29.88] showed that the critical fluctuations in a bulk crystal may affect the dissipation of mechanical probes even if completely external to the crystal surface. NC-AFM dissipation bears clear evidence of a most classic bulk transition, that to the antiferrodistortive state of $SrTiO_3$, as is shown in Fig. 29.23b. It also showed that the friction coefficient peaks at temperatures of 110–130 K. In this range the system becomes very *soft* owing to proliferation of domain walls, defects, and disorder in the distortive part of the crystal structure. Moreover, the frictional response close to the transition was governed by critical exponents characteristic for the three-dimensional (3-D) Ising model.

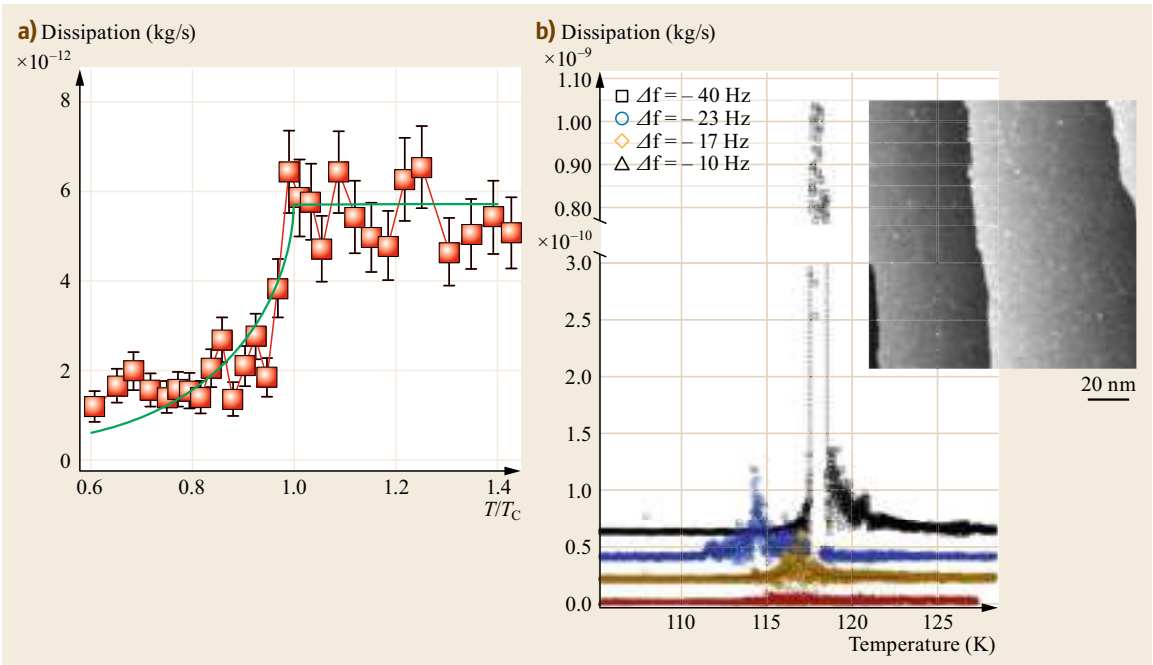


Fig. 29.23 **(a)** Temperature dependence of noncontact friction across the critical temperature of an Nb sample. The *green line* is a BCS theory fit to the measured data (*red squares*) (after [29.87]). **(b)** Experimental AFM dissipation as a function of temperature and tip–$SrTiO_3$ sample distance. The *black color* corresponds to the closest approach and the sharp peak is located at the critical temperature of $SrTiO_3$ in the bulk region under the tip. *Inset* shows a low-temperature ($T = 5$ K) STM image of the $SrTiO_3$ surface. (Reprinted with permission from [29.88]. Copyright 2015 by the American Physical Society)

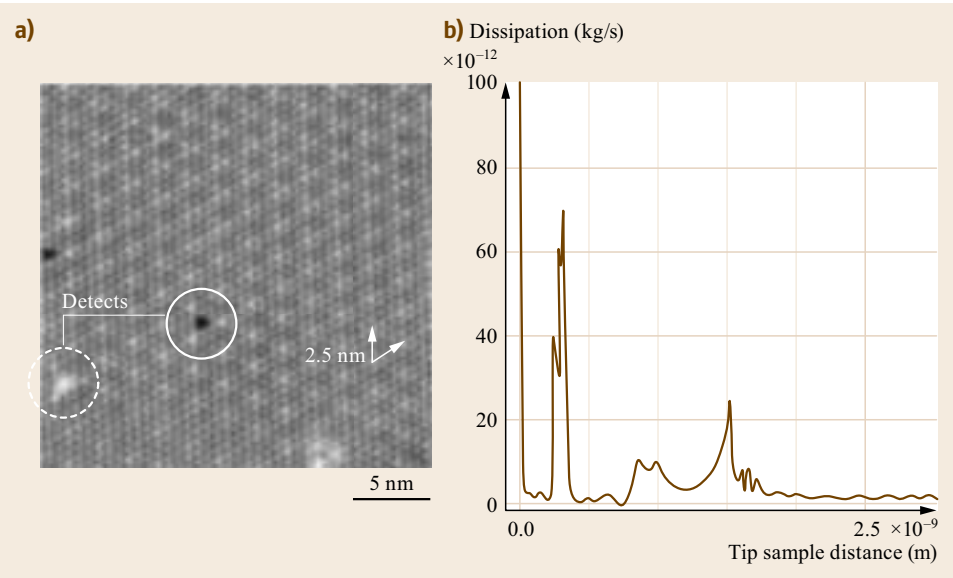


Fig. 29.24 **(a)** STM image of an $NbSe_2$ surface at $T = 6\,K$ and **(b)** Energy dissipation as a function of tip–sample distance. Three giant dissipation peaks are visible (after [29.91])

Pendulum AFM experiments have also addressed the coupling of a mechanical oscillator to *charge-density-waves* (CDWs) on an $NbSe_2$ sample, an elusively subtle long-range property of matter [29.91]. Figure 29.21a shows an atomically resolved surface topography with the additional CDW-induced moiré pattern, which is incommensurate with the underlying lattice. The distance dependence of the friction coefficient $\Gamma(d)$ shows striking multiple dissipation peaks at a few nm distance from the surface (Fig. 29.24b). The temperature evolution of the peaks suggested a strong connection with CDWs. In order to understand the origin of the dissipation spikes a theoretical model was proposed, where a CDW is considered to be an elastic medium, whose phase is perturbed locally by the attractive potential of the cantilever tip [29.92]. This is shown in Fig. 29.24c. For a given perturbation the charge peak below the tip displaces (red solid line) and the resulting CDW phase is getting distorted (red dashed line). As the force reaches a critical value the central peak disappears and an extra 2π phase is pumped locally below the cantilever tip (green line). The energies for the two discussed phase deformations ($N = 0$ and $N = 1$) are shown by the red and green solid lines and a crossover energy is marked by a circle. Oscillations close to a crossover point will cause pumping in and out of an extra 2π phase and thus be accompanied by a hysteresis cycle in the tip mechanics. The hysteresis explains the giant size of dissipation peaks and approaching the tip further down, might produce another phase slip explaining multiple dissipation peaks.

Lateral oscillations can also be excited in normal AFM geometry when exciting the torsional oscillation of commercial, rectangular AFM cantilevers. Also in this case the tip is oscillating approximately parallel to the surface. This mode, although much less force sensitive than the pendulum AFM, allowed one to measure lateral forces acting on the tip at step edges and near impurities quantitatively [29.93]. Enhanced energy dissipation was also observed at the impurities. When the tip was moved further toward the sample, the contact formation transformed the nearly free torsional oscillation of the cantilever into a different mode, with the tip–sample contact acting as a hinge. This caused a rapid increase in the power required to maintain a constant tip oscillation amplitude and a positive frequency shift.

29.8 Single Molecule Friction

The experiments described so far focused on the friction forces between a sharp tip and a crystal surface either in contact or in close proximity to each other. In this section we will address the problem of the frictional response of a nano-object as small as a single molecule sliding on a surface. Such experiments require us to intentionally pick up the molecule previously deposited on a solid surface with the apex of an AFM tip and to manipulate it laterally or vertically. This can be also done using quartz tuning forks as a force

sensor (QPlus) operated in NC mode. The high stiffness ($\approx 2\,\mathrm{kN/m}$) of the sensor allows one to achieve subangstrom oscillation amplitudes. Furthermore, low-temperature conditions are mandatory to ensure the stability of the molecule–surface junction while performing the friction experiments.

Ternes et al. first succeeded in measuring *the force to move an adatom on a surface* by means of force spectroscopy [29.94, 95]. First attempts to measure the force accompanying the lifting of a PTCDA (perylenetetracarboxylic dianhydride) molecule on Au(111) were conducted by the group of *Temirov* [29.96, 97]. Systematic vertical manipulations demonstrated reproducible features of the force curves which were related to the detachment of the molecule, i. e., adhesion, as well as to the mechanics of the molecular junction.

To investigate the mechanical degrees of freedom of a molecular structure during vertical displacements, *Kawai* et al. performed pulling manipulation experiments of long polymeric chains [29.98]. The chains were synthesized on a Au(111) surface by Ullmann-coupling of several dibromoterfluorene (DBTF) molecules previously adsorbed on the surface. A single chain, having a length of a few nm and lying flat on the surface, is bound to the tip of an AFM as shown in Fig. 29.25a. Tip retraction results in vertical manipulation of the molecular chain. During such lifting, force–distance $\Delta f(z)$ curves (Sect. 29.1.3) were acquired (Fig. 29.25b). The effective stiffness, calculated as the derivative of the curve, is indicated on the right scale. A periodic variation is observed, with each period terminated by an abrupt jump. The number of periods (24) exactly corresponds to the number of the fluorene units forming the chain, as previously estimated by STM, and the periodicity of about 0.9 nm corresponds to the length of a single unit. The chain thus detached step-wise from the surface during pulling. The experimental results were interpreted within an *edge-driven Frenkel–Kontorova* (FK) model. The polymer chain was schematized as a series of particles connected by springs of stiffness k and equilibrium length b. Each particle interacts with the substrate via a periodic potential reproducing the symmetry of the Au(111) surface and expressed by a formula introduced by Steele to describe gas-atoms solid interaction [29.99]. As shown in Fig. 29.25c, the main features of the experimental $\Delta f(z)$ curves were satisfactorily reproduced in such a way. Simulations also pointed out the fact that the effective stiffness of the pulled segments must be large compared to the stiffness of the chemical bond formed between tip and chain. Remarkably, they also allowed one to estimate the normal force required to detach one unit, which cannot be directly measured in the experiment. This force (in the order of 0.25 nN) is nearly independent of the length of the chain as well as of the unit position in the chain.

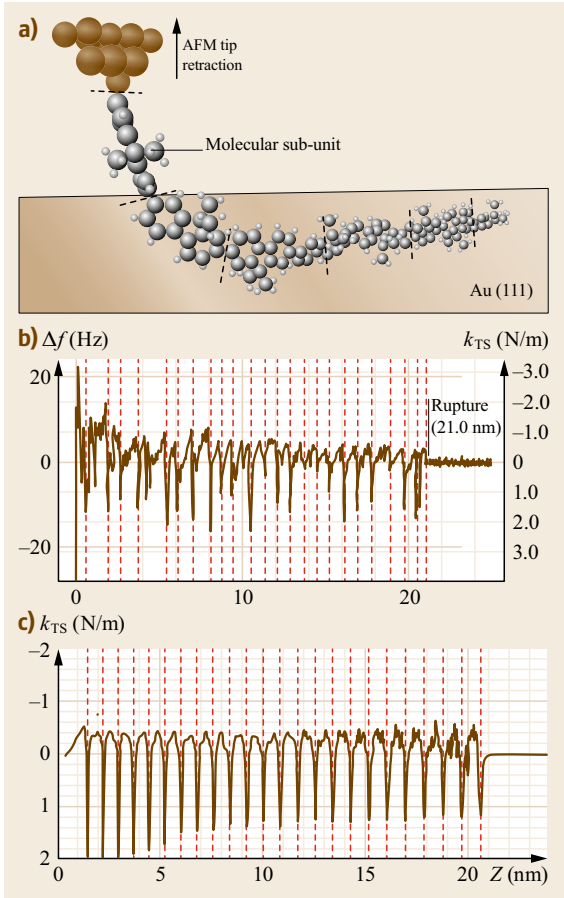


Fig. 29.25 **(a)** On-surface Ullmann coupling of DBTF molecules leads to long polymeric chains on Au(111). The end of the chain can be attached to the AFM tip and then retracted from the surface while acquiring interaction forces. **(b)** Experimental $\Delta f(z)$ spectra obtained during the lifting of the macromolecule (f_0 is the free resonance frequency of the tuning fork and k, its spring constant). The detachment is marked by the arrow. **(c)** Calculated dependence of the normal force gradient k_{TS} versus tip–surface distance z according the extended Frenkel–Kontorova model (Reprinted from [29.98]. Copyright 2014 National Academy of Sciences)

In another investigation *Kawai* et al. were able to measure the friction force acting on graphene nanoribbons (GNRs) sliding on a Au(111) surface using a similar approach [29.100]. The GNRs were again produced by Ullmann-coupling, and first manipulated without attaching them to the tip. The GNRs started to move in correspondence to exceptionally low friction force values of a few 100 pN. The static friction

per unit length was found to decrease with increasing GNR length, consistently with a strong superlubric behavior of graphene when forming an incommensurate contact. As a next step, the ends of selected GNRs were anchored to the tip and dragged back and forth in a controlled way (Fig. 29.26a). The corresponding variations of $\Delta f(x)$ are shown in Fig. 29.26c,d for increasing values of the tip–sample separation z. The frequency shift oscillates with a periodicity of 0.28 nm, except when the GNR was driven backward with $z = 5$ nm. The amplitude of the Δf oscillations is not constant along x but is modulated on distances of a few nm. Molecular dynamics (MD) simulations were also employed to reproduce the manipulation process. In this way the herringbone reconstruction of Au(111) was found to be responsible not only for the modulation of $\Delta f(x)$, but also for a significant reduction of commensurability between GNR and substrate. As a consequence, the GNR-Au(111) contact is almost superlubric, with both static and kinetic friction force values in the range of 100 pN.

The frictional response of a single organic molecule sliding on a metal surface was recently investigated by *Pawlak* et al. [29.101]. In this experiment, a porphyrin molecule was attached to the AFM tip via its core as shown in Fig 29.27a. The surface–molecule junction was formed by a single carbonitrile end-group of the porphyrin structure resulting in a single-point contact with a Cu(111) surface. While dragging the molecule, the junction was robust enough to maintain the contact although mechanical deformation of the molecular spring was expected. A friction $\Delta f(x, y)$ map of the Cu(111) surface obtained in such conditions is shown in Fig. 29.27b. The atomic-scale contrast resembles the frictional patterns obtained by conventional FFM on Cu(111) [29.51]. A dark depression corresponding to a Cu vacancy in the bottom right part of the image was also observed. In the line profiles of the tip–sample stiffness $k_{TS}(x)$ extracted from the Δf data (Fig. 29.21c) a clear saw-tooth modulation is detected resembling the standard atomic-scale friction pattern in FFM. Here, the hysteresis of the friction loop between forward and backward traces is small suggesting superlubric sliding. To verify this assumption, a generalized PT model including the Steele potential was used to simulate the friction process [29.99]. The intrinsic parameters characterizing the relevant stretching and bending rigidities of the molecule structure were extracted from first-principles calculations and used to parameterize the simulations. The calculations demonstrated that the variation of the normal force gradient during sliding is coupled to the friction force acting on the cyano end-group via the molecular entity thus supporting the experimental hypothesis.

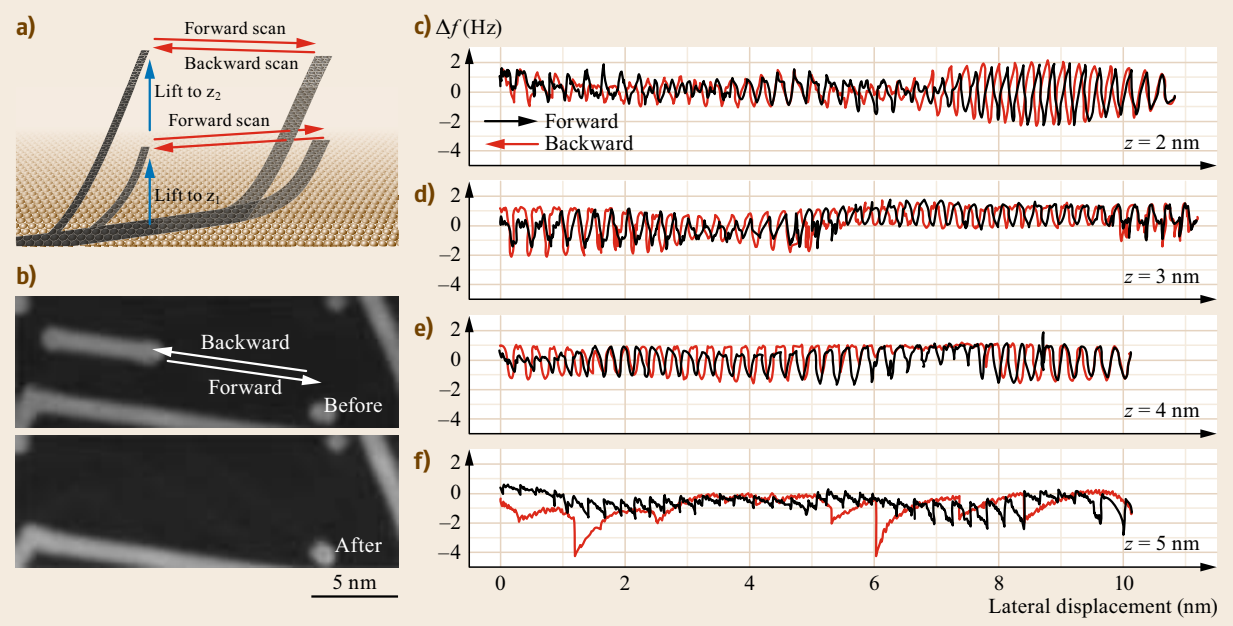


Fig. 29.26a–f Lateral manipulation of a GNR on a Au(111) surface. **(a)** Schematic drawing of the manipulation protocol at different tip–sample separations. **(b)** STM topographies before and after removal of a GNR moved on the Au(111) surface in the direction of the yellow arrows. **(c–f)** Frequency shifts accompanying the lateral motion of another GNR at different distances from the surface ($z = 2, 3, 4$ and 5 nm). (After [29.100], reprinted with permission from AAAS)

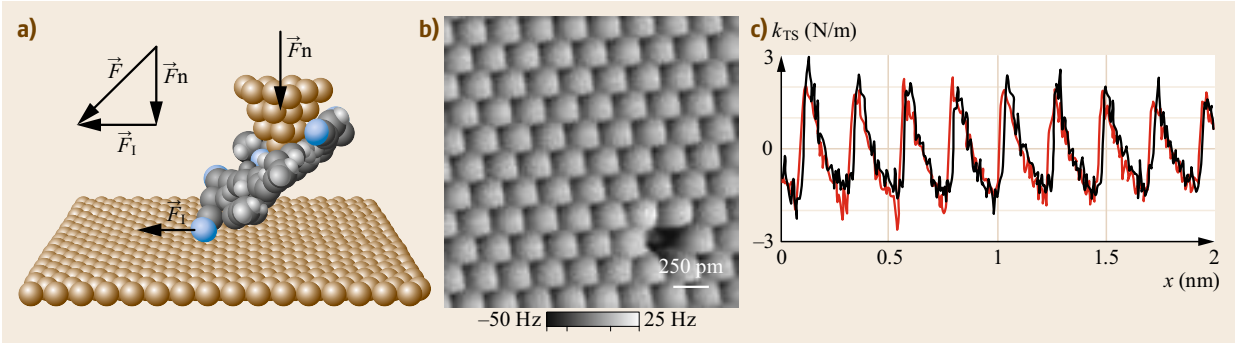


Fig. 29.27a–c Single molecule sliding on a metal substrate: **(a)** a porphyrin is attached to an AFM tip by vertical manipulation and dragged over a Cu(111) surface. **(b)** Constant-height $\Delta f(x, y)$ map of the Cu(111) surface revealing a typical stick–slip pattern. **(c)** Cross section showing the saw tooth variation. (Reprinted from [29.101]. Copyright 2016 American Chemical Society)

29.9 Conclusion

Over the last 25 years, the establishment of friction force microscopy has led to exceptional development in the field of nanotribology. The sliding motion of single asperities with only a few atoms forming the contact can be precisely investigated, and the friction forces accompanying the motion can be studied as a function of time, but also of load, velocity, and temperature. Stick–slip motion on the atomic scale is properly described by the Prandtl–Tomlinson model. Extensions of this model predict superlubric transitions and thermally activated slip, both of which have been experimentally verified. Friction force microscopy can also be applied to investigate initial stages of abrasive wear. Conversely, noncontact friction has recently become an active topic of research thanks to instrumental refinements such as pendulum AFM and tuning fork sensors. Distinguishing phononic and electronic contributions to noncontact friction is nowadays possible, as well as estimating the friction forces acting on polymer chains, graphene ribbons, and even single molecules dragged or lifted off a solid surface.

References

29.1 C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surfce, Phys. Rev. Lett. **59**, 1942–1945 (1987)

29.2 O. Marti, J. Colchero, J. Mlynek: Combined scanning force and friction microscopy of mica, Nanotechnology **1**, 141–144 (1990)

29.3 G. Meyer, N. Amer: Simultaneous measurement of lateral and normal forces with an optical-beam-deflection atomic force microscope, Appl. Phys. Lett. **57**, 2089–2091 (1990)

29.4 G. Neubauer, S.R. Cohen, G.M. McClelland, D.E. Horn, C.M. Mate: Force microscopy with a bidirectional capacitance sensor, Rev. Sci. Instrum. **61**, 2296–2308 (1990)

29.5 G.M. McClelland, J.N. Glosli: Friction at the atomic scale. In: *NATO ASI Series E*, Vol. 220, ed. by L. Singer, H.M. Pollock (Kluwer, Dordrecht 1992) pp. 405–425

29.6 R. Linnemann, T. Gotszalk, I.W. Rangelow, P. Oesterschuze: Atomic force microscopy and lateral force microscopy using piezoresistive cantilevers, J. Vac. Sci. Technol. B **14**, 856–860 (1996)

29.7 M. Nonnenmacher, J. Greschner, O. Wolter, R. Kassing: Scanning force microscopy with micromachined silicon sensors, J. Vac. Sci. Technol. B **9**, 1358–1362 (1991)

29.8 J. Cleveland, S. Manne, D. Bocek, P.K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64**, 403–405 (1993)

29.9 J.L. Hutter, J. Bechhoefer: Caliberation of atomic-force microscope tips, Rev. Sci. Instrum. **64**, 1868–1873 (1993)

29.10 H.J. Butt, M. Jaschke: Calculation of thermal noise in atomic-force microscopy, Nanotechnology **6**, 1–7 (1995)

29.11 J.M. Neumeister, W.A. Ducker: Lateral, normal and longitudinal spring constants at atomic-force microscopy cantilevers, Rev. Sci. Instrum. **65**, 2527–2531 (1994)

29.12 D.F. Ogletree, R.W. Carpick, M. Salmeron: Calibration of frictional forces in atomic force microscopy, Rev. Sci. Instrum. **67**, 3298–3306 (1996)

29.13 U.D. Schwarz, P. Köster, R. Wiesendanger: Quantitative analysis of lateral force microscopy experiments, Rev. Sci. Instrum. **67**, 2560–2567 (1996)

29.14 J.E. Sader, C.P. Green: In-plane deformation of cantilever plates with applications to lateral force microscopy, Rev. Sci. Instrum. **75**, 878–883 (2004)

29.15 S.A. Edwards, W.A. Ducker, J.E. Sader: Influence of atomic force microscope cantilever tilt and induced torque on force measurements, J. Appl. Phys. **103**, 064513 (2008)

29.16 J.S. Villarrubia: Algorithms for scanned probe microscope image simulation, surface reconstruction and tip estimation, J. Res. Natl. Inst. Stand. Technol. **102**, 425–454 (1997)

29.17 L. Howald, E. Meyer, R. Lüthi, H. Haefke, R. Overney, H. Rudin, H.-J. Güntherodt: Multifunctional probe microscope for facile operation in ultrahigh vacuum, Appl. Phys. Lett. **63**, 117–119 (1993)

29.18 B. Gotsmann, C. Seidel, B. Anczykowski, H. Fuchs: Conservative and dissipative tip-sample interaction forces probed with dynamic AFM, Phys. Rev. B **60**, 11051–11061 (1999)

29.19 B. Gotsmann: Tribology: Sliding on vacuum, Nat. Mat. **10**, 87–88 (2011)

29.20 C. Loppacher, R. Bennewitz, O. Pfeiffer, M. Guggisberg, R. Schär, M. Bammerlin, V. Barwich, A. Baratoff, E. Meyer: Experimental aspects of dissipation force microscopy, Phys. Rev. B **62**, 13674–13679 (2000)

29.21 T.R. Albrecht, P. Grütter, D. Horne, D. Rugar: Frequency modulation detection using high-Q cantilevers for enhanced force microscope senitivity, J. Appl. Phys. **69**, 668 (1991)

29.22 G. Vilhena, C. Pimentel, P. Pedraz, F. Luo, P.A. Serena, C.M. Pina, E. Gnecco, R. Perez: Atomic-scale sliding friction on grapheme in water, ACS Nano **10**, 4288–4293 (2016)

29.23 L. Prandtl: A conceptual model for the kinetic theory of solids, Z. Angew. Math. Mech. **8**, 85 (1928), in German

29.24 A. Socoliuc, R. Bennewitz, E. Gnecco, E. Meyer: Transition from stick–slip to continuous sliding in atomic friction: Entering a new regime of ultralow friction, Phys. Rev. Lett. **92**, 134301 (2004)

29.25 E. Gnecco, R. Roth, A. Baratoff: Analytical expressions for the kinetic friction in the Prandtl–Tomlinson model, Phys. Rev. B **86**, 035443 (2012)

29.26 T. Gyalog, M. Bammerlin, R. Lüthi, E. Meyer, H. Thomas: Mechanism of atomic friction, Europhys. Lett. **31**, 269–274 (1995)

29.27 P. Steiner, R. Roth, E. Gnecco, A. Baratoff, E. Meyer: Angular dependence of static and kinetic friction on alkali halide surfaces, Phys. Rev. B **82**, 205417 (2010)

29.28 E. Gnecco: Quasi-isotropy of static friction on hexagonal surface lattices, Europhys. Lett. **91**, 66008 (2010)

29.29 T. Gyalog, H. Thomas: Friction between atomically flat surfaces, Europhys. Lett. **37**, 195–200 (1997)

29.30 M. Weiss, F.J. Elmer: Dry friction in the Frenkel–Kontorova–Tomlinson model: Static properties, Phys. Rev. B **53**, 7539–7549 (1996)

29.31 M.H. Müser, M.O. Robbins: Conditions for static friction between flat crystalline surfaces, Phys. Rev. B **61**, 2335 (2000)

29.32 M.H. Müser: Structural lubricity: Role of dimension and symmetry, Europhys. Lett. **66**, 97 (2004)

29.33 S. Maier, Y. Sang, T. Filleter, M. Grant, R. Bennewitz, E. Gnecco, E. Meyer: Fluctuations and jump dynamics in atomic friction experiments, Phys. Rev. B **72**, 245418 (2005)

29.34 R. Roth, T. Glatzel, P. Steiner, E. Gnecco, A. Baratoff, E. Meyer: Multiple slips in atomic-scale friction: An indicator for the lateral contact damping, Tribol. Lett. **39**, 63 (2010)

29.35 A. Socoliuc, E. Gnecco, S. Maier, O. Pfeiffer, A. Baratoff, R. Bennewitz, E. Meyer: Atomic-scale control of friction by actuation of nanometer-sized contacts, Science **313**, 207–210 (2006)

29.36 R. Roth, O.Y. Fajardo, J.J. Mazo, E. Meyer, E. Gnecco: Lateral vibration effects in atomic-scale friction, Appl. Phys. Lett. **104**, 083103 (2014)

29.37 E. Gnecco, R. Bennewitz, T. Gyalog, C. Lopacher, M. Bammerlin, E. Meyer, H.-J. Güntherodt: Velocity dependence of atomic friction, Phys. Rev. Lett. **84**, 1172–1175 (2000)

29.38 T. Filleter, R. Bennewitz: Structural and frictional properties of graphene on SiC(0001) studied by atomic force microscopy, Phys. Rev. B **81**, 155412 (2010)

29.39 M. Goryl, J. Budzioch, F. Krok, M. Wojtaszek, M. Kolmer, L. Walczak, J. Konior, E. Gnecco, M. Szymonski: Probing atomic-scale friction on reconstructed surfaces of single-crystal semiconductors, Phys. Rev. B **85**, 085308 (2012)

29.40 S. Maier, E. Gnecco, A. Baratoff, E. Meyer: Atomic-scale friction modulated by a buried interface: Combined atomic and friction force microscopy experiments, Phys. Rev. B **78**, 045432 (2008)

29.41 P. Steiner, E. Gnecco, F. Krok, J. Budzioch, L. Walczak, J. Konior, M. Szymonski, E. Meyer: Atomic-scale friction on stepped surfaces of ionic crystals, Phys. Rev. Lett. **106**, 186104 (2011)

29.42 S. Fujisawa, E. Kishi, Y. Sugawara, S. Morita: Atomic-scale friction observed with a two-dimensional frictional-force microscope, Phys. Rev. B **51**, 7849–7857 (1995)

29.43 H. Kawakatsu, T. Saito: Scanning force microscopy with two optical levers for detection of deformations of the cantilevers, J. Vac. Sci. Technol. B **14**, 872–876 (1996)

29.44 M. Dienwiebel, G. Verhoeven, N. Pradeep, J. Frenken, J. Heimberg, H. Zandbergen: Superlubricity of graphite, Phys. Rev. Lett. **92**, 126101 (2004)

29.45 G. Fessler, I. Zimmermann, T. Glatzel, E. Gnecco, P. Steiner, R. Roth, T.D. Keene, S.X. Liu, S. Decurtins, E. Meyer: Orientation dependent molecular friction on organic layer compound crystals,

Appl. Phys. Lett. **98**, 083119 (2011)
29.46 C.M. Pina, R. Miranda, E. Gnecco: Anisotropic surface coupling while sliding on dolomite and calcite crystals, Phys. Rev. B **85**, 073402 (2012)
29.47 L.P. Nita, C. Pimentel, F. Luo, B. Milian-Medina, J. Gierschner, C.M. Pina, E. Gnecco: Molecular resolution friction microscopy of Cu phtalocyanine films on dolomite(104) in water, Nanoscale **6**, 8344 (2014)
29.48 Y. Sang, M. Dubé, M. Grant: Thermal effects on atomic friction, Phys. Rev. Lett. **87**, 174301 (2001)
29.49 S.Y. Krylov, K.B. Jinesh, H. Valk, M. Dienwiebel, J.W.M. Frenken: Thermally induced suppression of friction at the atomic scale, Phys. Rev. E **71**, 65101 (2005)
29.50 C. Fusco, A. Fasoline: Velocity dependence of atomic-scale friction: A comparative study of the one- and two-dimensional Tomlinson model, Phys. Rev. B **71**, 45413 (2005)
29.51 R. Bennewitz, T. Gyalog, M. Guggisberg, M. Bammerlin, E. Meyer, H.-J. Güntherodt: Atomic-scale stick–slip processes on Cu(111), Phys. Rev. B **60**, R11301–R11304 (1999)
29.52 T. Bouhacina, J.P. Aimé, S. Gauthier, D. Michel, V. Heroguez: Tribological behavior of a polymer grafted on silanized silica probed with a nanotip, Phys. Rev. B **56**, 7694–7703 (1997)
29.53 E. Riedo, F. Lévy, H. Brune: Kinetics of capillary condensation in nanoscopic sliding friction, Phys. Rev. Lett. **88**, 185505 (2002)
29.54 X. Zhao, M. Hamilton, W.G. Sawyer, S.S. Perry: Thermally activated friction, Tribol. Lett. **27**, 113–117 (2007)
29.55 L. Jansen, H. Hölscher, H. Fuchs, A. Schirmeisen: Temperature dependence of atomic-scale stick–slip friction, Phys. Rev. Lett. **104**, 256101 (2010)
29.56 I. Barel, M. Urbakh, L. Jansen, A. Schirmeisen: Multibond dynamics of nanoscale friction: The role of temperature, Phys. Rev. Lett. **104**, 066104 (2010)
29.57 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids* (Oxford Univ. Press, Oxford 1950)
29.58 L.D. Landau, E.M. Lifshitz: *Introduction to Theoretical Physics*, Vol. 7 (Nauka, Moscow 1998)
29.59 J.A. Greenwood, J.B.P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. Lond. A **295**, 300 (1966)
29.60 B.N.J. Persson: Elastoplastic contact between randomly rough surfaces, Phys. Rev. Lett. **87**, 116101 (2001)
29.61 K.L. Johnson, K. Kendall, A.D. Roberts: Surface energy and contact of elastic solids, Proc. R. Soc. Lond. A **324**, 301 (1971)
29.62 B.V. Derjaguin, V.M. Muller, Y.P. Toporov: Effect of contact deformations on adhesion of particles, J. Colloid Interface Sci. **53**, 314–326 (1975)
29.63 D. Tabor: Surface forces and surface interactions, J. Colloid Interface Sci. **58**, 2–13 (1977)
29.64 D. Maugis: Adhesion of spheres: The JKR-DMT transition using a Dugdale model, J. Colloid Interface Sci. **150**, 243–269 (1992)
29.65 U.D. Schwarz, O. Zwörner, P. Köster, R. Wiesendanger: Quantitative analysis of the frictional properties of solid materials at low loads, Phys. Rev. B **56**, 6987–6996 (1997)
29.66 E. Meyer, R. Lüthi, L. Howald, M. Bammerlin, M. Guggisberg, H.-J. Güntherodt: Site-specific friction force spectroscopy, J. Vac. Sci. Technol. B **14**, 1285–1288 (1996)
29.67 R.W. Carpick, N. Agrät, D.F. Ogletree, M. Salmeron: Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscopic, J. Vac. Sci. Technol. B **14**, 1289–1295 (1996)
29.68 C. Polaczyk, T. Schneider, J. Schöfer, E. Santer: Microtribological behavior of Au(001) studied by AFM–FFM, Surf. Sci. **402**, 454–458 (1998)
29.69 J.N. Israelachvili, D. Tabor: Measurement of van der Waals dispersion forces in range 1.5 to 130 nm, Proc. R. Soc. Lond. A **331**, 19 (1972)
29.70 K.L. Johnson: *Contact Mechanics* (Cambridge Univ. Press, Cambridge 1985)
29.71 P. Steiner, R. Roth, E. Gnecco, T. Glatzel, A. Baratoff, E. Meyer: Modulation of contact resonance frequency accompanying atomic-scale stickslip in friction force microscopy, Nanotechnology **20**, 495701 (2009)
29.72 R.W. Carpick, D.F. Ogletree, M. Salmeron: Lateral stiffness: A new nanomechanical measurement for the determination of shear strengths with friction force microscopy, Appl. Phys. Lett. **70**, 1548–1550 (1997)
29.73 M.A. Lantz, S.J. O'Shea, M.E. Welland, K.L. Johnson: Atomic-force-microscopic study of contact area and friction on $NbSe_2$, Phys. Rev. B. **55**, 10776–10785 (1997)
29.74 M. Enachescu, R.J.A. van den Oetelaar, R.W. Carpick, D.F. Ogletree, C.F.J. Flipse, M. Salmeron: Atomic force microscopy study of an ideally hard contact: The diamond(111)–tungsten carbide interface, Phys. Rev. Lett. **81**, 1877–1880 (1998)
29.75 M. Enachescu, R.J.A. van den Oetelaar, R.W. Carpick, D.F. Ogletree, C.F.J. Flipse, M. Salmeron: Observation of proportionality between friction and contact area at the nonometer scale, Tribol. Lett. **7**, 73–78 (1999)
29.76 E. Gnecco, R. Bennewitz, E. Meyer: Abrasive wear on the atomic scale, Phys. Rev. Lett. **88**, 215501 (2002)
29.77 T.D.B. Jacobs, R. Carpick: Nanoscale wear as a stress-assisted chemical reaction, Nature Mat **8**, 108 (2013)
29.78 T. Filleter, W. Paul, R. Bennewitz: Atomic structure and friction of ultrathin films of KBr on Cu(100), Phys. Rev. B **77**, 035430 (2000)
29.79 S. Kopta, M. Salmeron: The atomic scale origin of wear on mica and its contribution to friction, J. Chem. Phys. **113**, 8249–8252 (2000)
29.80 M. Guggisberg, M. Bammerlin, A. Baratoff, R. Lüthi, C. Loppacher, F.M. Battiston, J. Lü, R. Bennewitz, E. Meyer, H.J. Güntherodt: Dynamic force microscopy across steps on the Si(111)-(7 × 7) surface, Surf. Sci. **461**, 255–265 (2000)

29.81 C. Loppacher, R. Bennewitz, O. Pfeiffer, M. Guggisberg, M. Bammerlin, S. Schär, V. Barwich, A. Baratoff, E. Meyer: Experimental aspects of dissipation force microscopy, Phys. Rev. B **62**, 13674–13679 (2000)

29.82 R. Bennewitz, A.S. Foster, L.N. Kantorovich, M. Bammerlin, C. Loppacher, S. Schär, M. Guggisberg, E. Meyer, A.L. Shluger: Atomically resolved edges and kinks of NaCl islands on Cu(111): Experiment and theory, Phys. Rev. B **62**, 2074–2084 (2000)

29.83 R. Bennewitz, S. Schär, V. Barwich, O. Pfeiffer, E. Meyer, F. Krok, B. Such, J. Kolodzej, M. Szymonski: Atomic-resolution images of radiation damage in KBr, Surf. Sci. **474**, 197–202 (2001)

29.84 T.D. Stowe, T.W. Kenny, J. Thomson, D. Rugar: Silicon dopant imaging by dissipation force microscopy, Appl. Phys. Lett. **75**, 2785–2787 (1999)

29.85 W. Denk, D.W. Pohl: Local electrical dissipation imaged by scanning force microscopy, Appl. Phys. Lett. **59**, 2171–2173 (1991)

29.86 B.C. Stipe, H.J. Mamin, T.D. Stowe, T.W. Kenny, D. Rugar: Noncontact friction and force fluctuations between closely spaced bodies, Phys. Rev. Lett. **87**, 96801 (2011)

29.87 M. Kisiel, E. Gnecco, U. Gysin, L. Marot, S. Rast, E. Meyer: Suppression of electronic friction on Nb films in the superconducting state, Nat. Mat. **10**, 119 (2011)

29.88 M. Kisiel, F. Pellegrini, G.E. Santoro, M. Samadashvili, R. Pawlak, A. Benassi, U. Gysin, R. Buzio, A. Gerbi, E. Meyer, E. Tosatti: Noncontact atomic force microscope dissipation reveals a central peak of $SrTiO_3$ structural phase transition, Phys. Rev. Lett. **115**, 046101 (2015)

29.89 J. Bardeen, L.N. Cooper, J.R. Schrieffer: Theory of superconductivity, Phys. Rev. **108**, 1175 (1957)

29.90 A. Benassi, A. Vanossi, G.E. Santoro, E. Tosatti: Sliding over a phase transition, Phys. Rev. Lett. **106**, 256102 (2011)

29.91 M. Langer, M. Kisiel, R. Pawlak, F. Pellegrini, G.E. Santoro, R. Buzio, A. Gerbi, G. Balakrishnan, A. Baratoff, E. Tosatti, E. Meyer: Giant frictional dissipation peaks and charge-density-wave slips at the $NbSe_2$ surface, Nat. Mat. **13**, 173 (2013)

29.92 F. Pellegrini, G.E. Santoro, E. Tosatti: Charge-density-wave surface phase slips and noncontact nanofriction, Phys. Rev. B **89**, 245416 (2014)

29.93 O. Pfeiffer, R. Bennewitz, A. Baratoff, E. Meyer, P. Grütter: Lateral-force measurements in dynamic force microscopy, Phys. Rev. B **65**, 161403 (2002)

29.94 M. Ternes, C.P. Lutz, C.F. Hirjibehedin, F.J. Giessibl, A.J. Heinrich: The force needed to move an atom, Science **319**, 1066–1069 (2008)

29.95 M. Emmrich, F. Schneiderbauer, F. Huber, A.J. Weymouth, N. Okabayashi, F.J. Giessibl: Force field analysis suggests a lowering of diffusion barriers in atomic manipulation due to presence of STM tip, Phys. Rev. Lett. **114**, 146101 (2015)

29.96 N. Fournier, C. Wagner, C. Weiss, R. Temirov, F.S. Tautz: Force-controlled lifting of molecular wires, Phys. Rev. B **84**, 035435 (2011)

29.97 C. Wagner, N. Fournier, F.S. Tautz, R. Temirov: The role of surface corrugation and tip oscillation in single-molecule manipulation with non-contact atomic force microscope, Beilstein J. Nanotechnol. **5**, 202–209 (2014)

29.98 S. Kawai, M. Koch, E. Gnecco, A. Sadeghi, R. Pawlak, T. Glatzel, J. Schwarz, S. Goedecker, S. Hecht, A. Baratoff, E. Meyer: Quantifying the atomic-level mechanics of single long physisorbed molecular chains, Proc. Natl. Acad. Sci. USA **111**, 3968–3972 (2014)

29.99 W.A. Steele: The physical interaction of gases with crystalline solids: I. Gas–solid energies and properties of isolated adsorbed atoms, Surf. Sci. **36**, 317 (1973)

29.100 S. Kawai, A. Benassi, E. Gnecco, H. Söde, R. Pawlak, X. Feng, K. Müllen, D. Passerone, C.A. Pignedoli, P. Ruffieux, R. Fasel, E. Meyer: Superlubricity of graphene nanoribbons on gold surfaces, Science **351**, 957 (2016)

29.101 R. Pawlak, W. Ouyang, A.E. Filippov, L. Kalikhman-Razvozov, S. Kawai, T. Glatzel, E. Gnecco, A. Baratoff, Q.-S. Zheng, O. Hod, M. Urbakh, E. Meyer: Single molecule tribology: Force microscopy manipulation of a porphyrin derivative on a copper surface, ACS Nano **10**, 713–722 (2016)

30. Computer Simulations of Nanometer-Scale Indentation and Friction

Susan B. Sinnott, Seong-Jun Heo, Donald W. Brenner, Judith A. Harrison, Douglas L. Irving

Engines and other machines with moving parts are often limited in their design and operational lifetime by friction and wear. This limitation has motivated the study of fundamental tribological processes with the ultimate aim of controlling and minimizing their impact. The recent development of miniature apparatus, such as microelectromechanical systems (MEMSs) and nanometer-scale devices, has increased interest in atomic-scale friction, which has been found to, in some cases, be due to mechanisms that are distinct from the mechanisms that dominate in macroscale friction.

Presented in this chapter is a review of computational studies of tribological processes at the atomic and nanometer scale. In particular, a review of the findings of computational studies of nanometer-scale indentation, friction and lubrication is presented, along with a review of the salient computational methods that are used in these studies, and the conditions under which they are best applied.

30.1 **Computational Details** 1014
30.1.1 Energies and Forces 1014
30.1.2 Important Approximations 1016
30.2 **Indentation** 1019
30.2.1 Surfaces 1019
30.2.2 Thin Films 1028
30.3 **Friction and Lubrication** 1033
30.3.1 Bare Surfaces 1033
30.3.2 Decorated Surfaces 1039
30.3.3 Thin Films 1041
30.4 **Conclusions** 1058
References 1058

Engines and other machines with moving parts are often limited in their design and operational lifetime by friction and wear. This limitation has motivated the study of tribological processes with the aim of controlling and minimizing the impact of these processes. There are numerous historical examples that illustrate the importance of friction to the development of civilizations, including the ancient Egyptians who invented technologies to move the stones used to build the pyramids [30.1]; Coulomb, who was motivated to study friction by the need to move ships easily and without wear from land to the water [30.1]; and *Johnson* et al. [30.2], who developed an improved understanding of contact mechanics and surface energies through the study of automobile windshield wipers. At present, substantial research and development is aimed at microscale and nanoscale machines with moving parts that at times challenge our fundamental understanding of friction and wear. This has motivated the study of atomic-scale friction and has, consequently, led to new discoveries such as self-lubricating surfaces and wear-resistant materials. While there are similarities between friction at the macroscale and the atomic scale, in many instances the mechanisms that lead to friction at these two scales are quite different. Thus, as devices such as magnetic storage disks and microelectromechanical systems (MEMS) [30.3] continue to shrink in size, it is expected that new phenomena associated with atomic-scale friction, adhesion and wear will dominate the functioning of these devices.

The last two decades have seen considerable scientific effort expended on the study of atomic-scale friction [30.4–17]. This effort has been facilitated by the development of new advanced experimental tools to measure friction over nanometer-scale distances at low loads, rapid improvements in computer power, and the maturation of computational methodologies for the modeling of materials at the atomic scale. For example, friction-force and atomic-force microscopes (FFMs and AFMs) allow the frictional properties of solids to be characterized with atomic-scale resolution under single-asperity indentation and sliding condi-

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_30

tions [30.18–21]. In addition, the surface force apparatus (SFA) provides data about the tribological and lubrication responses of many liquid and solid systems with atomic resolution [30.22], and the quartz crystal microbalance (QCM) provides information about the atomic-scale origins of friction [30.4, 23]. These and related experimental methods allow researchers to study sliding surfaces at the atomic scale and relate the observed phenomena to macroscopically observed friction, lubrication and wear.

Analytic models and computational simulations have played an important role in characterizing and understanding friction. They can, for example, assist in the interpretation of experimental data or provide predictions that subsequent experiments can confirm or refute. Analytic models have long been used to study friction, including early studies by *Tomlinson* [30.24] and *Frenkel* and *Kontorova* [30.25] and more recent studies by *McClelland* and *Glosli* [30.26], *Sokoloff* [30.13, 27–33], *Persson* et al. [30.34–37] and others [30.38–44]. Most of these idealized models divide the complex motions that create friction into more fundamental components defined by quantities such as spring constants, the curvature and magnitude of potential wells, and bulk phonon frequencies. While these simplifications provide these approaches with some predictive capabilities, many assumptions must be made in order to be able to apply these models to study friction, which may lead to incorrect or incomplete results.

In atomic-scale molecular dynamics (MD) simulations, atom trajectories are calculated by numerically integrating coupled classical equations of motion. Interatomic forces that enter these equations are typically calculated either from total energy methods that include electronic degrees of freedom, or from simplified mathematical expressions that give the potential energy as a function of interatomic displacements. MD simulations can be considered numerical experiments that provide a link between analytic models and experiments. The main strength of MD simulations is that they can reveal unanticipated phenomena or unexpected mechanisms for well-known observations. Weaknesses include a lack of quantum effects in classical atomistic dynamics, and perhaps more importantly, the fact that meaningless results can be obtained if the simulation conditions are chosen incorrectly. The next section contains a review of MD simulations, including the approximations that are inherent in their application to the study of friction, and the conditions under which they should and should not be applied.

30.1 Computational Details

Molecular dynamics simulations are straightforward to describe: given a set of initial conditions and a way of mathematically modeling interatomic forces, Newton's (or equivalent) classical equation of motion is numerically integrated [30.45]

$$\boldsymbol{F} = m\boldsymbol{a}\,, \tag{30.1a}$$

$$-\nabla E = m\left(\frac{\partial^2 \boldsymbol{r}}{\partial t^2}\right)\,, \tag{30.1b}$$

where F is the force on each atom, m is the atomic mass, a is the atomic acceleration, E is the potential energy felt by each atom, r is the atomic position, and t is time. The forces acting on any given atom are calculated, and then the atoms move a short increment ∂t (called a time step) forward in time in response to these applied forces. This is accompanied by a change in atomic positions, velocities and accelerations. The process is then repeated for some specified number of time steps.

The output of these simulations includes new atomic positions, velocities, and forces that allow additional quantities such as temperature and pressure to be determined. As the size of the system increases, it is useful to render the atomic positions in animated movies that reveal the responses of the system in a qualitative manner. Quantitative data can be obtained by analyzing the numerical output directly.

The following sections review the way in which energies and forces are calculated in MD simulations and the important approximations that are used to realistically model the friction that occurs in experiments with smaller systems of only a few tens of thousands of atoms in simulations. The reader is referred to additional sources [30.46–52] for a more comprehensive overview of MD simulations (including computer algorithms), analysis methods, and the potentials that are used to calculate energies and forces in MD simulations.

30.1.1 Energies and Forces

There are several different approaches by which interatomic energies and forces are determined in MD simulations. The most theoretically rigorous methods are those that are classified as ab initio or first principles. These techniques, which include density functional theory [30.53, 54] and quantum chemical ab initio [30.55]

methods, are derived from quantum mechanical principles and are generally both the most accurate and the most computationally intensive. They are therefore limited to a small number of atoms (< 500), which has limited their use in the study of friction. Alternatively, empirical methods are functions containing parameters that are determined by fitting to experimental data or the results of ab initio calculations [30.50]. These techniques can usually be relied on to correctly describe qualitative trends and are often the only choice available for modeling systems containing tens of thousands, millions, or billions of atoms. Empirical methods have therefore been widely used in studies of friction. Semiempirical methods, including tight-binding methods, include some elements of both empirical methods and ab initio methods. For instance, they require quantum mechanical information in the form of, for example, on-site and hopping matrix elements, and include fits to experimental data [30.56].

Empirical methods simplify the modeling of materials by treating the atoms as spheres that interact with each other via repulsive and attractive terms that can be either pairwise additive or many-body in nature. In this approach, electrons are not treated explicitly, although it is understood that the interatomic interactions are ultimately dependent on them. As discussed in this section, some empirical methods explicitly include charge through classical electrostatic interactions, although most methods assume charge-neutral systems. The repulsive and attractive functional forms generally depend on interatomic distances and/or angles and contain adjustable parameters that are fit to ab initio results and/or experimental data.

The main strength of empirical potentials is their computational speed. Recent simulations with these approaches have modeled billions of atoms [30.57], something that is not possible with ab initio or semiempirical approaches at this time. The main weakness of empirical potentials is their lack of quantitative accuracy, especially if they are poorly formulated or applied to systems that are too far removed from the fitting database used in their construction. Furthermore, because of the differences in the nature of chemical bonding in various materials, such as covalent bonding in carbon versus metallic bonding in gold, empirical methods have been historically derived for particular classes of materials. They are therefore generally nontransferable, although some methods have been shown to be theoretically equivalent [30.51, 58], and in recent years there has been progress towards the development of empirical methods that can model heterogeneous material systems [30.59–64].

Several of the most important and common general classes of empirical methods used for calculating interatomic energies and forces in materials, the so-called potentials, are reviewed here. The first to be considered are the potentials that are used to model covalently bound materials, including the bond-order potential and the Stillinger–Weber potential.

The bond-order potential was first formulated by *Abell* [30.65] and subsequently developed and parameterized by *Tersoff* for silicon and germanium [30.66, 67], *Brenner* and coworkers for hydrocarbons [30.52, 68, 69], *Dyson* and *Smith* for carbon-silicon-hydrogen systems [30.70], *Sinnott* and coworkers for carbon-oxygen-hydrogen systems [30.71], and *Graves* and coworkers [30.72] and *Sinnott* and coworkers [30.73] for fluorocarbons, *Schall* and coworkers for pure Si [30.74], and *Hu* and coworkers on C-N-O-H [30.75].

The bond-order potential has the general functional form

$$E = \sum_i \sum_{j(>i)} [V_R(r_{ij}) - b_{ij} V_A(r_{ij})] \qquad (30.2)$$

where $V_R(r)$ and $V_A(r)$ are pair-additive interactions that model the interatomic repulsion and electron–nuclear attraction, respectively. The quantity r_{ij} is the distance between pairs of nearest-neighbor atoms i and j, and b_{ij} is a bond-order term that takes into account the many-body interactions between atoms i and j, including those due to nearest neighbors and angle effects. The potential is short-ranged and only considers nearest neighbor bonds. To model long-range nonbonded interactions, the bond-order potential is combined with pair-wise potentials either directly through splines [30.76] or indirectly with more sophisticated functions [30.77].

The *Stillinger–Weber* potential [30.78] potential was formulated to model silicon, with a particular emphasis on the liquid phases of silicon. It includes many-body interactions in the form of a sum of two- and three-body interactions

$$E = \sum_{ij} V_{ij}^2(r_{ij}) + \sum_{jik} V_{jik}^3(r_{ij}, r_{ik}) \,, \qquad (30.3)$$

where V^2 is a pair-additive interaction and V^3 is a three-body term. The three-body term includes an angular interaction that minimizes the potential energy for tetrahedral angles. This term favors the formation of open structures, such as the diamond cubic crystal structure.

The second potential is the embedded atom method (EAM) approach [30.79, 80] and related methods [30.81], which were initially developed for model-

ing metals and alloys. The functional form in the EAM is

$$E = \sum_i F(\rho_i) + \sum_{i>j} \Phi(r_{ij}) \,, \tag{30.4}$$

where F is called the embedding energy. This term models the energy due to embedding an atom into a uniform electron gas with a uniform compensating positive background (jellium) of density ρ_i that is equal to the actual electron density of the system. The term $\Phi(r_{ij})$ is a pairwise functional form that corrects for the jellium approximation. Several parameterizations of the EAM exist (see, for example, [30.79, 80, 82–84]) and it has recently been extended to model nonmetallic systems. For example, the modified EAM (MEAM) approach [30.64, 85, 86] was developed so that EAM could be applied to metal oxides [30.60] and covalently bound materials [30.86].

The third method is the general class of Coulomb or multipole interaction potentials used to model charged ionic materials or molecules [30.87]. In this formalism, an energy term is given as

$$E = \sum_i \sum_{j(>i)} \left(\frac{q(r_i)q(r_j)}{r_{ij}} \right) , \tag{30.5}$$

where $q(r_i)$ is the charge on atom i and r_{ij} is the distance between atoms i and j. More complex formalisms that take into account, say, the Madelung constant in the case of ionic crystals, are used in practice. In general, the charges are held fixed, but methods that allow charge to vary in a realistic manner have been developed [30.61, 88].

Lastly, long-range van der Waals or related forces are typically modeled with pairwise additive potentials. A widely used approximation is the Lennard-Jones (LJ) potential [30.89], which has the following functional form

$$E = 4\varepsilon \sum_i \sum_{j(>i)} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] . \tag{30.6}$$

In this approach ε and σ are parameters and r_{ij} is the distance between atoms i and j.

All of these potentials are widely used in MD simulations of materials, including studies of friction, lubrication, and wear.

30.1.2 Important Approximations

Several approximations are typically used in MD simulations of friction. The first is the use of periodic boundary conditions (PBCs) and the minimum image convention for interatomic interactions [30.48]. In both cases the simulation supercell is surrounded by replicas of itself so that atoms (or phonons, etc.) that exit one side of the supercell remerge into the simulation through the opposite side of the supercell. In the minimum image convention an atom interacts either with another atom in the supercell or its equivalent atom in a surrounding cell depending on which distance to the atom is shortest. This process is illustrated in Fig. 30.1. In this convention supercells must be large enough that atoms do not interact with themselves over the periodic boundaries. In computational studies of friction and wear, PBCs are usually applied in the two dimensions within the plane(s) of the sliding surface(s). The strength of this approach is that it allows a finite number of atoms to model an infinite system. However, the influence of boundaries on system dynamics is not completely eliminated; for example, phonon scattering due to the periodic boundaries can influence heat transport and therefore frictional properties of sliding interfaces.

Another important tool that is often used in MD simulations of friction is thermostats to regulate system temperature. In macroscopic systems, heat that is generated from friction is dissipated rapidly from the surface to the bulk phonon modes. Because atomistic computer simulations are limited systems that are many orders of magnitude smaller than systems that are generally studied experimentally, thermostats are needed to prevent the system temperature from rising

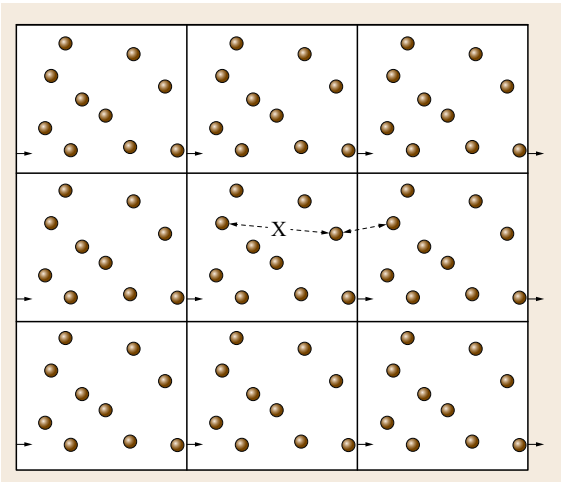


Fig. 30.1 Illustration of periodic boundary conditions consisting of a central simulation cell surrounded by replica systems. The *solid arrows* indicate an atom leaving the central box and reentering on the opposite side. The *dotted arrows* illustrate the minimum image convention

in a nonphysical manner. Typically in simulations of indentation or friction, the thermostat is applied to a region of the simulation cell that is well removed from the interface where friction and indentation is taking place. In this way, local heating of the interface that occurs as work is done on the system, but excess heat is efficiently dissipated from the system as a whole. In this manner the adjustment of atomic temperatures occurs away from the processes of interest, and simplified approximations for the friction term can be used without unduly influencing the dynamics produced by the interatomic forces.

There are several different formalisms for atomistic thermostats. The simplest of these controls the temperature by intermittently rescaling the atomic velocities to values corresponding to the desired temperature [30.90] such that

$$\left(\frac{\boldsymbol{v}_{\text{new}}}{\boldsymbol{v}_{\text{old}}}\right)^2 = \frac{T}{T_{\text{ins}}} , \tag{30.7}$$

where $\boldsymbol{v}_{\text{new}}$ is the rescaled velocity, and $\boldsymbol{v}_{\text{old}}$ is the velocity before the rescaling. This approach, which is called the velocity rescaling method, is both simple to implement and effective at maintaining a given temperature over the course of an MD simulation. It was consequently widely used in early MD simulations. The velocity rescaling approach does have some significant disadvantages, however. First, there is little theoretical basis for the adjustment of atomic velocities, and the system dynamics are not time-reversible, which is inconsistent with classical mechanics. Second, the rate and mode of heat dissipation are disconnected from system properties, which may affect system dynamics. Lastly, for typical MD simulation system sizes, the averaged quantities that are obtained, such as pressure for instance, do not correspond to values in any thermodynamic ensemble.

For these reasons, more sophisticated methods for maintaining system temperatures in MD simulations have been developed. The Langevin dynamics approach [30.48], which was originally developed from the theory of Brownian motion, falls into this category. In this approach, terms are added to the interatomic forces that correspond to a random force and a frictional term [30.46, 91, 92]. Therefore, Newton's equation of motion for atoms subjected to Langevin thermostats is given by the following equation rather than (30.1a) and (30.1b)

$$m\boldsymbol{a} = \boldsymbol{F} - m\xi\boldsymbol{v} + R(t) , \tag{30.8}$$

where $\boldsymbol{F}$ are the forces due to the interatomic potential, ξ is a friction coefficient, m and $\boldsymbol{v}$ are the particle's mass and velocity, respectively, and $R(t)$ is a random force that acts as *white noise*. The friction term can be formulated in terms of a memory kernal, typically for harmonic solids [30.93–95], or a friction coefficient can be approximated using the Debye frequency. The random force can be given by a Gaussian distribution where the width, which is chosen to satisfy the fluctuation-dissipation theorem, is determined from the equation

$$\langle R(0)R(t)\rangle = 2mk_{\text{B}}T\xi\delta(t) . \tag{30.9}$$

Here, the function R is the random force in (30.8), m is the particle mass, T is the desired temperature, k_{B} is Boltzmann's constant, t is time, and ξ is the friction coefficient. It should be noted that the random forces are uncoupled from those at previous steps, which is denoted by the delta function. Additionally, the width of the Gaussian distribution from which the random force is obtained varies with temperature. Thus, the Langevin approach does not require any feedback from the current temperature of the system as the random forces are determined solely from (30.9).

In the early 1980s, *Nosé* developed a new thermostat that corresponds directly to a canonical ensemble (system with constant temperature, volume and number of atoms) [30.96, 97], which is a significant advance from the methods described so far. In this approach, Nosé introduces a degree of freedom s that corresponds to the heat bath and acts as a time scaling factor, and adds a parameter Q that may be regarded as the heat bath *mass*. A simplified form of Nosé's method was subsequently implemented by *Hoover* [30.46] that eliminated the time scaling factor whilst introducing a thermodynamic friction coefficient ζ. Hoover's formulation of Nosé's method is therefore easy to use and is commonly referred to as the Nosé–Hoover thermostat.

When this thermostat is applied to a system containing N atoms, the equations of motion are written as (dots denote time derivatives)

$$\begin{aligned}
\dot{\boldsymbol{r}}_i &= \frac{\boldsymbol{p}_i}{m_i} , \\
\dot{\boldsymbol{p}}_i &= \boldsymbol{F}_i - \zeta\boldsymbol{p}_i , \\
\dot{\zeta} &= \frac{1}{Q}\left(\sum_{i=1}^{N}\frac{p_i^2}{m_i} - N_{\text{f}}k_{\text{B}}T\right) ,
\end{aligned} \tag{30.10}$$

where $\boldsymbol{r}_i$ is the position of atom i, $\boldsymbol{p}_i$ is the momentum and $\boldsymbol{F}_i$ is the force applied to each atom. The last equation in (30.10) contains the temperature control mechanism in the Nosé–Hoover thermostat. In particular, the term between the parentheses on the right-hand

side of this equation is the difference between the system's instantaneous kinetic energy and the kinetic energy at the desired temperature. If the instantaneous value is higher than the desired one, the friction force will increase to lower it and vice versa.

It should be pointed out that the choice of the heat bath *mass* Q is arbitrary but crucial to the successful performance of the thermostat. For example, a small value of Q leads to rapid temperature fluctuation while large Q values result in inefficient sampling of phase space. Nosé recommended that Q should be proportional to $N_f k_B T$ and should allow the added degree of freedom s to oscillate around its averaged value at a frequency of the same order as the characteristic frequency of the physical system [30.96, 97]. If ergodic dynamic behavior is assumed, the Nosé–Hoover thermostat will maintain a well-defined canonical distribution in both momentum and coordinate space. However, for small systems where the dynamic is not ergodic, the Nosé–Hoover thermostat fails to generate a canonical distribution. Therefore, more sophisticated algorithms based on the Nosé–Hoover thermostat have been proposed to fix its ergodicity problem; for example, the *Nosé–Hoover chain* method of *Martyna* et al. [30.98]. However, these complex thermostats are not as easy to apply as the Nosé–Hoover thermostat due to the difficult evaluation of the coupling parameters for each different case and the significant computational cost [30.99]. From a practical point of view, if the molecular system is large enough that the movements of the atoms are sufficiently chaotic, ergodicity is guaranteed and the performance of the Nosé–Hoover thermostat is satisfactory [30.25].

In an alternative approach, *Schall* et al. recently introduced a hybrid continuum-atomistic thermostat [30.100]. In this method, an MD system is divided into grid regions, and the average kinetic energy in the atomistic simulation is used to define a temperature for each region. A continuum heat transfer equation is then solved stepwise on the grid using a finite difference approximation, and the velocities of the atoms in each grid region are scaled to match the solution of the continuum equation. To help account for a time lag in the transfer of kinetic to potential energy, Hoover constraining forces are added to those from the interatomic potential. This process is continued, leading to an ad hoc feedback between the continuum and atomistic simulations. The main advantage of this approach is that the experimental thermal diffusivity can be used in the continuum expression, leading to heat transfer behavior that matches experimental data. For example, in metals the majority of the thermal properties at room temperature arise from electronic degrees of freedom that are neglected with strictly classical potentials. This thermostat is relatively straightforward to implement, and requires only the interatomic potential and the bulk thermal diffusivity as input. It is also appropriate for nonequilibrium heat transfer, such as occurs as heat is dissipated from sliding surfaces moving at high relative velocities. Other localized phenomena, such as Joule heating and melting in current carrying applications, can also be simulated by using a recent extension to the hybrid thermostat [30.101]. This modification allows the ability to model degradation of interfaces under high electrical load at the atomic level. Relevant examples are hot switched radio frequency MEMS and metal/metal contacts in electromagnetic launchers.

Cushman et al. [30.102, 103] developed a unique alternative to the grand canonical ensemble by performing a series of grand canonical Monte Carlo simulations [30.48, 104] at various points along a hypothetical sliding trajectory. The results from these simulations are then used to calculate the correct particle numbers at a fixed chemical potential, which are then used as inputs to nonsliding, constant-*NVE* MD simulations at each of the chosen trajectory points. The sliding speed can be assumed to be infinitely slow because the system is fully equilibrated at each step along the sliding trajectory. This approach offers a useful alternative to continuous MD simulations that are restricted to sliding speeds that are orders of magnitude larger than most experimental studies (about 1 m/s or greater).

To summarize, this section provides a brief review and description of components that are used in atomistic, molecular dynamics simulation of many of the processes related to friction, such as indentation, sliding, and wear. The components discussed here include the potential energy expression used to calculate energies and forces in the simulations, periodic boundary conditions and thermostats. Each of these components has their own strengths and weaknesses that should be well-understood both prior to their use and in the interpretation of results. For example, general principles related to liquid lubrication in confined areas may be most easily understood and generalized from simulations that use pair potentials and may not require a thermostat. On the other hand, if one wants to study the wear or indentation of a surface of a particular metal, then EAM or other semiempirical potentials, together with a thermostat, would be expected to yield more reliable results. If one requires information on electronic effects, ab initio or semi-empirical approaches that include the evaluation of electronic degrees of freedom must be used. Thus, the best combination of components for a particular study depends on the chemical nature of the system of interest, the processes being simulated, the type of information desired, and the available computational resources.

30.2 Indentation

It is critical to understand the nanometer-scale properties of materials that are being considered for use as new coatings with specific friction and wear behavior. Experimental determination of these properties is most frequently done with the AFM, which provides a variety of data related to the interaction of the microscope tips with the sample surface [30.105–107]. In AFM experiments, the tip has a radius of about 1−100 nm and is pressed against the surface under ambient conditions (in air), ultrahigh vacuum (UHV) conditions, or in a liquid. The microscope tip can either move in the direction normal to the surface, which is the case in nanoindentation studies, or raster across the surface, which is the case in surface imaging or friction studies. Sliding rates of 1 nm/s–1 μm/s are typically used, which are many orders of magnitude slower than the rates used in MD simulations of sliding or indentation of around 1−100 m/s. As discussed in the previous section, the higher rates used in computational simulations are a consequence of modeling full atomic motion, which occurs on a femto- to picosecond timescale, and the stepwise solution of the classical equations of motion, which makes the large number of simulation steps needed to reach experimental timescales computationally impossible with current processor speeds.

As the tip moves either normal to or across the surface, the forces acting upon it as a result of its interactions with the surface are measured. When the tip is moved in the surface normal direction, it can penetrate the surface on the nanometer scale and provide information on the nanometer-scale mechanical properties of the surface [30.108, 109]. The indentation process also causes the force on the tip to increase, and the rate of increase is related to both the depth of indentation and the properties of the surface. The region of the force curve that reflects this high force is known as the repulsive wall region [30.105], or, when considered without any lateral motion of the tip, an indentation curve. When the tip is retracted after indentation, enhanced adhesion between the tip and surface relative to the initial contact can result. This phenomenon is indicated by hysteresis in the force curve.

Tip–surface adhesion can result from the formation of chemical bonds between the tip and the sample, or from the formation of liquid capillaries between the microscope tip and the surface caused by the interaction of the tip with a layer of liquid contamination on the surface. The latter case is especially prevalent in AFM studies conducted in ambient environments. In the case of clean metallic systems, the sample can wet the tip or the tip can wet the sample in the form of a connective *neck* of metal atoms between the surface and the tip that can lead to adhesion. In the case of polymeric or molecular systems, entanglement of molecules that are anchored on the tip with molecules anchored on the sample can be responsible for force curve hysteresis.

In the case of horizontal rastering of AFM tips across surfaces, the force curve data provide a map of the surface that is indicative of the surface topography [30.110]. If the deflection of the tip in the lateral direction is recorded while the tip is being rastered, a friction map of the surface [30.20] is produced.

The rest of this section discusses some of the important insights and findings that have been obtained from MD simulations of nanoindentation. These studies have not only provided insight into the physical phenomena responsible for the qualitative shapes of AFM force curves, they have also revealed a wealth of atomic-scale phenomena that occur during nanoindentation that was not previously known.

30.2.1 Surfaces

The nature of adhesive interactions between clean, deformable metal tips indenting metal surfaces have been identified and clarified over the course of the last decade through the use of MD simulations [30.107, 111–115]. In particular, the high surface energies associated with clean metal surfaces can lead to strongly attractive interactions between surfaces in contact. The strength of this attraction can be so large that when the tip gets close enough to the surface to interact with it, surface atoms *jump* upwards to wet the tip in a phenomena known as jump-to-contact (JC). This wetting mechanism was first discovered in MD simulations [30.114] and has been confirmed experimentally [30.108, 116–118] using the AFM, as shown in Fig. 30.2.

The MD simulations of *Landman* et al. [30.114, 119–121] using EAM potentials revealed that the JC phenomenon in metallic systems is driven by the need of the atoms at the tip–surface interface to optimize their interaction energies while maintaining their individual material cohesive binding. When the tip advances past the JC point it indents the surface, which causes the force to increase. This behavior is indicated in Fig. 30.3, points D to M. This region of the computer-generated force curve has a maximum not present in the force curve generated from experimental data (Fig. 30.3, point L). This is due to tip-induced flow of the metal atoms in the surface that causes *pile-up* of the surface atoms around the edges of the indenter. Hysteresis on the withdrawal of the tip, shown in Fig. 30.3, points M to X, is present due to adhesion between the tip and the substrate. In particular, as the tip re-

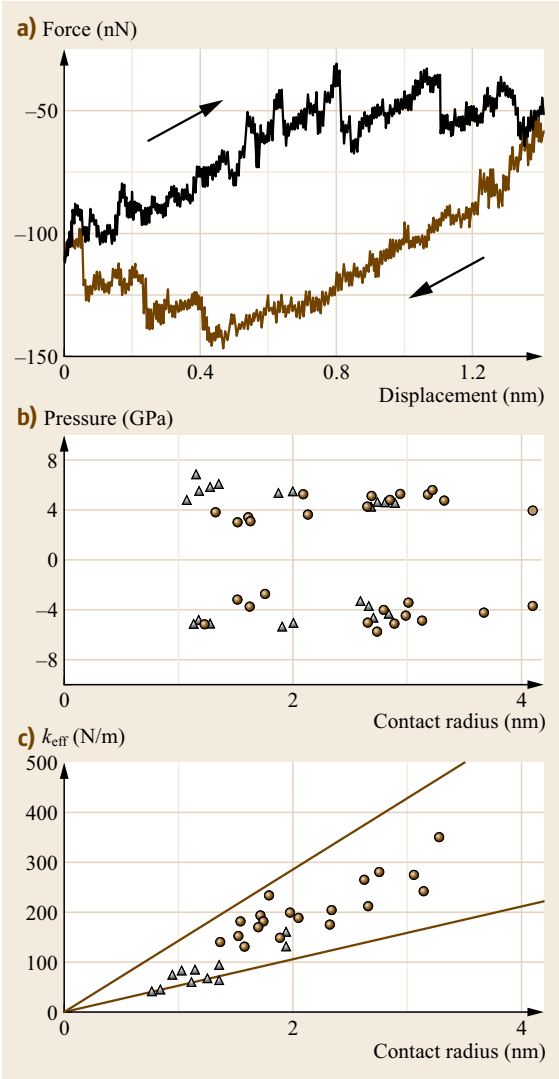


Fig. 30.2 **(a)** The experimental values for the force between a tip and a surface that have a connective neck between them. **(b)** The neck contracts and extends without breaking on the scales shown. **(c)** The effective spring constant k_{eff} determined experimentally for the connective necks and corresponding maximum pressures, versus contact radius of the tip. The *triangles* indicate measurements taken at room temperature; the *circles* are the measurements taken at liquid He temperatures (after [30.118], © ACS 1996)

tracts from the sample, a connective *neck* or nanowire of atoms forms between the tip and the substrate that is primarily composed of metal atoms from the surface with some atoms from the metal indenter that have diffused into the structure. A snapshot from the MD simulations that illustrates this behavior is shown in Fig. 30.4.

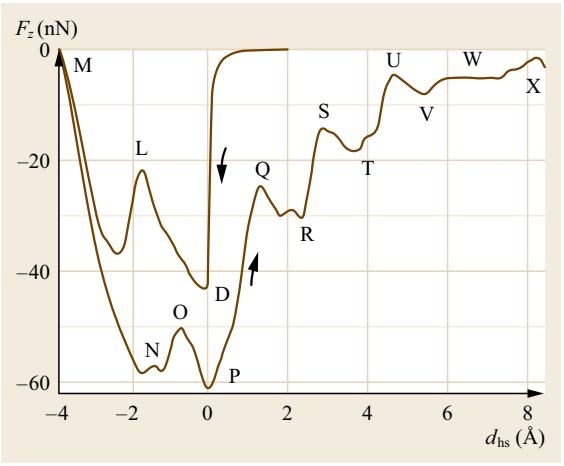


Fig. 30.3 Computationally derived force F_z versus tip-to-sample distance d_{hs} curves for approach, contact, indentation, then separation using the same tip–sample system shown in Fig. 30.4. These data were calculated from an MD simulation (after [30.114], © AAAS 1990)

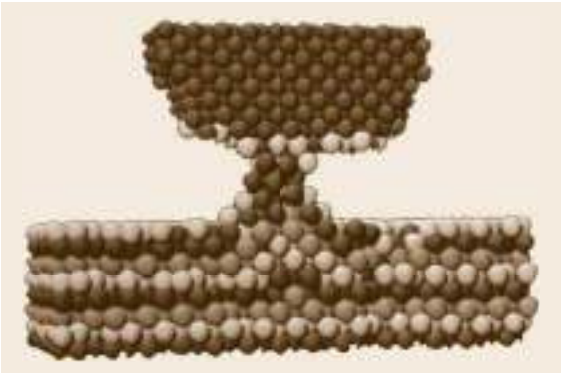


Fig. 30.4 Illustration of atoms in the MD simulation of a Ni tip being pulled back from an Au substrate. This causes the formation of a connective neck of atoms between the tip and the surface (after [30.114], © AAAS 1990)

As the tip is withdrawn farther, the magnitude of the force increases (becomes more negative) until, at a critical force, the atoms in adjacent layers of the connective nanowire rearrange so that an additional row of atoms is created. This process causes elongation of the connective nanowire and is responsible for the fine structure (apparent as a series of maxima) present in the retraction portion of the force curve. These elongation and rearrangement steps are repeated until the connection between the tip and the surface is broken. Similar elongation events have been observed experimentally. For example, scanning tunneling microscopy (STM) experiments demonstrate that the metal nanowires between

metal tips and surfaces can elongate ≈ 2500 Å without breaking [30.122].

The JC process has been shown to affect the temperature at the tip–surface interface. For instance, the constant-energy MD simulations of *Tomagnini* et al. [30.123] predicted that the energy released due to the wetting of the tip by surface atoms increases the temperature of the tip by about 15 K at room temperature and is accompanied by significant structural rearrangement. At temperatures high enough to cause the first few metal surface layers to be liquid, the distance at which the JC occurs increases, as does the contact area between the tip and the surface and the amount of nanowire elongation prior to breakage.

Simulations by *Komvopoulos* and *Yan* [30.124] using LJ potentials showed how metallic surfaces respond to single and repeated indentation by metallic, or covalently bound, rigid tips. The simulations predicted that a single indentation event produces hysteresis in the force curve as a result of surface plastic deformation and heating. The repulsive force decreases abruptly during surface penetration by the tip and surface plastic deformation. Repeated indentation results in the continuous decrease of the elastic stiffness, surface heating, and mean contact pressures at maximum penetration depths to produce behaviors that are similar to cyclic work hardening and softening by annealing observed in metals at the macroscale.

When the tip is much stiffer than the surface, pile-up of surface atoms around the tip occurs to relieve the stresses induced by nanoindentation. In contrast, when the surface is much stiffer than the tip, the tip can be damaged or destroyed. Simulations by *Belak* and *Stowers* [30.126] using perfectly rigid tips showed the mechanism by which the surface yields plastically after its elastic threshold is exceeded. The simulations showed how nanoindentation causes surface atoms to move on to the surface but under the tip and thus cause atomic pile-up. In this study, variations in the indentation rate reveal that point defects created as a result of nanoindentation relax by moving through the surface if the rate of indentation is slow enough. If the indentation rate is too high, there is no time for the point defects to relax and move away from the indentation area and so strain builds up more rapidly. The rigid indenters considered in these MD simulations are analogous to experiments that use surface passivation to prevent JC between the tip and the surface [30.125, 127], the results of which agree with the predicted results of pile-up and crater formation, as shown in Fig. 30.5 [30.125].

In short, MD simulations are able to explain the atomic-scale mechanisms behind measured experimental force curves produced when metal tips indent homogeneous metal surfaces to nanometer-scale depths. This preliminary work has spawned much of the current interest in using the JC to produce metal nanowires [30.128–130].

MD simulations have also been used to examine the relationship between nanoindentation and surface structure. This is most apparent in a series of computational studies that consider the indentation of a surface with a *virtual* hard-sphere indenter in a manner that is independent of the rate of indentation, as shown in Fig. 30.6. The virtual indenter is modeled through the application of a repulsive force to the surface rather than through the presence of an actual atomic tip. *Kelchner* et al. [30.131], rather than use MD, pushed the indenter against the surface a short distance and then allowed the system to relax using standard energy minimization methods in combination with EAM potentials. The

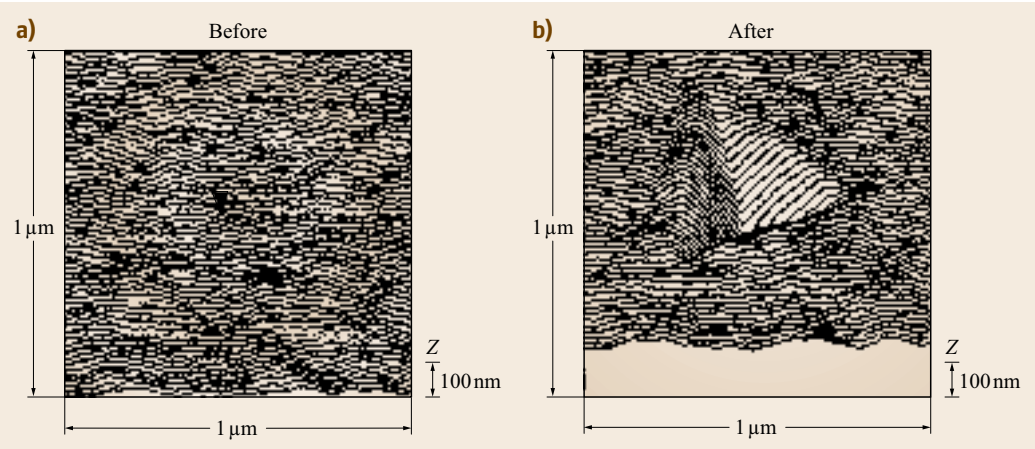


Fig. 30.5a,b Images of a gold surface *before* (**a**) and *after* (**b**) being indented with a pyramidal shaped diamond tip in air. The indentation created a surface crater. Note the pile-up around the crater edges (after [30.125], © Elsevier 1993)

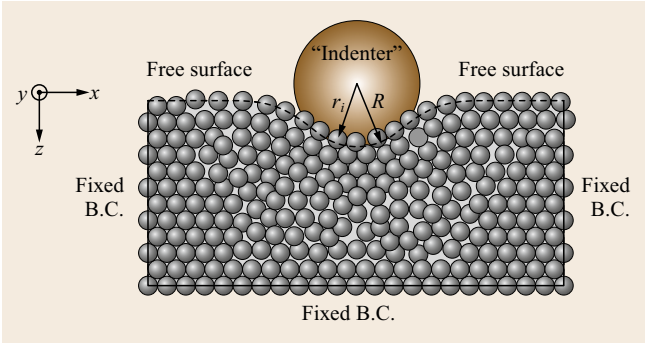


Fig. 30.6 A schematic of a spherical, virtual tip indenting a metal surface (after [30.132], © Elsevier 1993) (B.C. – boundary conditions)

system is fully relaxed when the energy of the surface system is minimized. After relaxation, the tip is pushed further into the surface and the process is repeated. As the tip generates more stresses in the surface, dislocations are generated and plastic deformation occurs. If the tip is pulled back after indenting less than a specific critical value, the atoms that were plastically deformed are healed during the retraction and the surface recovers its original structure. In contrast, if the tip is indented past the critical depth, additional dislocations are created that interfere with the surface healing process on tip withdrawal. In this case, a surface crater is left on the surface following nanoindentation.

A similar study by *Lilleodden* et al. [30.132] considered the generation of dislocations in perfect crystals and near grain boundaries in gold. Analysis of the relationship between the load and the tip displacement in the perfect crystal shows discrete load drops followed by elastic behavior. These load drops are shown to correspond to the homogeneous nucleation of dislocations, as illustrated in Fig. 30.7, which is a snapshot taken just after the load drop. When nanoindentation occurs close to a grain boundary, similar relationships between the load and tip displacement are predicted to occur as were seen for the perfect crystal. However, the dislocations responsible for the load drop are preferentially emitted from the grain boundaries, as illustrated in Fig. 30.8.

Simulations can also show how atomic structure and stresses are affected by nanoindentation. For instance, MD simulations with a virtual indenter by *Hasnaoui* et al. [30.133] using semi-empirical tight-binding methods showed the interaction between the grain boundaries under the indenter and the dislocations generated by the indentation, as illustrated in Fig. 30.9. This study shows that if the size of the indenter is smaller than the grain size, the grain boundaries can emit, absorb, and reflect the dislocations in a manner that depends on atomic structure and the distribution of stresses.



Fig. 30.7 Snapshot of two partial dislocations separated by a stacking fault. The *dark spheres in the center* of the structure indicate atoms in perfect crystal positions after both partial dislocations have passed (after [30.132], © Elsevier 1993)



Fig. 30.8 Snapshot of the high-energy atoms only after a load drop caused by dislocation generation during the nanoindentation of gold near a grain boundary (after [30.132], © Elsevier 1993)

Zimmerman et al. considered the indentation of a single-crystal gold substrate both near and far from a surface step [30.134]. The results of these simulations, which used EAM potentials, showed that the onset of plastic deformation depends to a significant degree on the distance of indentation from the step, and whether the indentation is on the plane above or below the step. In a related set of simulations, *Shenderova* et al. [30.135] examined whether ultrashallow elastic nanoindentation can nondestructively probe surface stress distributions associated with surface structures

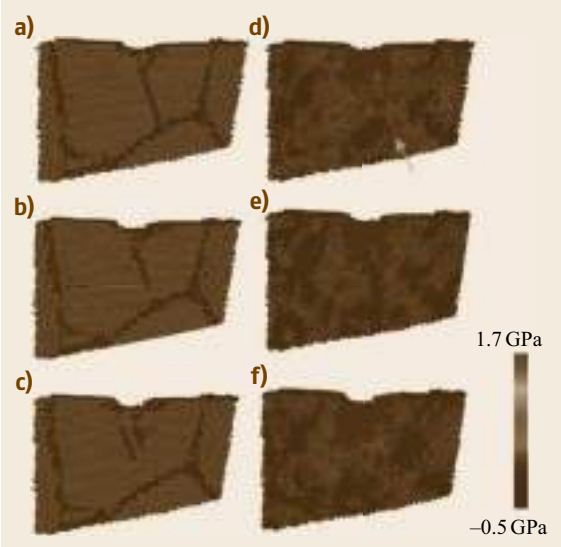


Fig. 30.9a–d Snapshots showing the atomic stress distribution and atomic structures in a gold surface. Figures **(a–c)** show the atomic structure at indentation depths of 7.9, 8.6, and 9.6 Å, respectively, with a virtual spherical indenter. A dislocation is represented by the two parallel {111} planes (*two dark lines*) that show the stacking fault left behind after the leading partial dislocation has passed. Figures **(d–f)** show the atomic stress distribution of the same system at the same indentation depths. Here the *dark color* indicates compressive hydrostatic pressures of 1.7 GPa and higher while the *gray* color indicates tensile pressures of -0.5 GPa and lower. The *arrow* in **(d)** shows the region of the system where a dislocation interacts with a grain boundary (after [30.133], © Elsevier 2004)

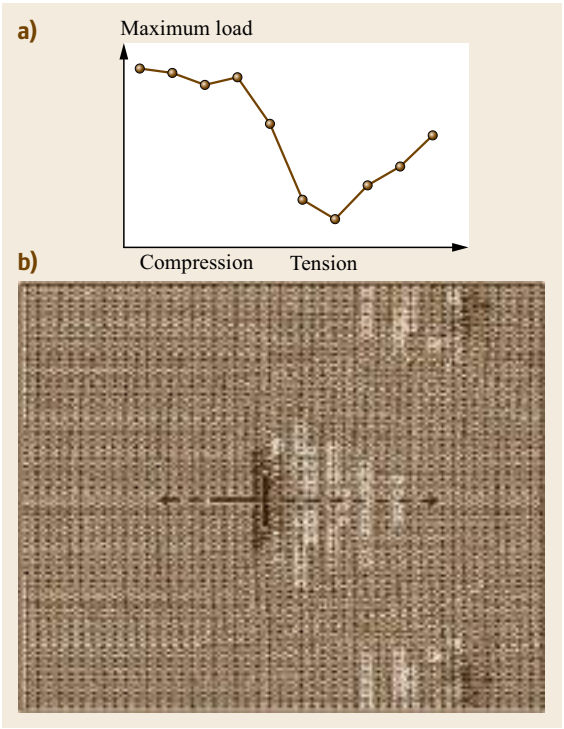


Fig. 30.10a,b Data and system illustration from a simulation of a gold surface containing a dislocation. **(a)** Maximum load for simulated shallow indentation at several points along the *dotted line* in the bottom illustration. **(b)** Top view of the simulated surface. The dislocation is denoted by the *solid black lines*

such as a trench and a dislocation intersecting a surface. The simulations carried out the nanoindentation to a constant depth. They predicted maximum loads that reflect the in-plane stresses at the point of contact between the indenter and the substrate, as illustrated in Fig. 30.10.

Since the 1930s, studies have been performed using hardness measurement techniques [30.136–139] and indentation methods [30.140] that suggest that the hardness of a material depends on applied in-plane uni- and bi-axial strain. In general, tensile strain appeared to decrease hardness while increases in hardness under compressive in-plane strain were reported. This behavior had traditionally been attributed to the contribution of stresses from the local strain from the indentation to the resolved shear stresses and the in-plane strain [30.137, 139]. However, in 1996, *Pharr* and coworkers determined that changes in elastic modulus determined from unloading curves of strained substrates using contact areas estimated via an elastic model are too large to have physical significance, a result that brought into question the interpretation of prior hardness data [30.140, 141]. They hypothesized that the apparent change in modulus (and hardness) with in-plane strain is mainly due to changes in contact area that are not typically taken into consideration in elastic half-space models. This hypothesis was based on experimental nanoindentation studies of a strained polycrystalline aluminum alloy and finite element calculations on an isotropic solid [30.140, 141]. They further suggested that in-plane compression increases pile-up around the indenter that, when not taken into account in the analysis of unloading curves, implies a nonphysical increase in modulus. Likewise, they suggested that in-plane tensile strain reduces the amount of material that is piled up around an indenter, which leads to a corresponding reduced (nonphysical) modulus when interpreting unloading curves using elastic models.

To explore in more detail the issue of pile-up and its influence on the interpretation of loading curves, *Schall* and *Brenner* used MD simulations and EAM potentials to model the plastic nanoindentation of a single-crystal

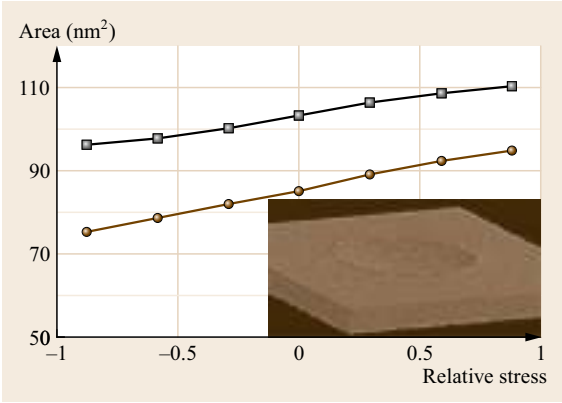
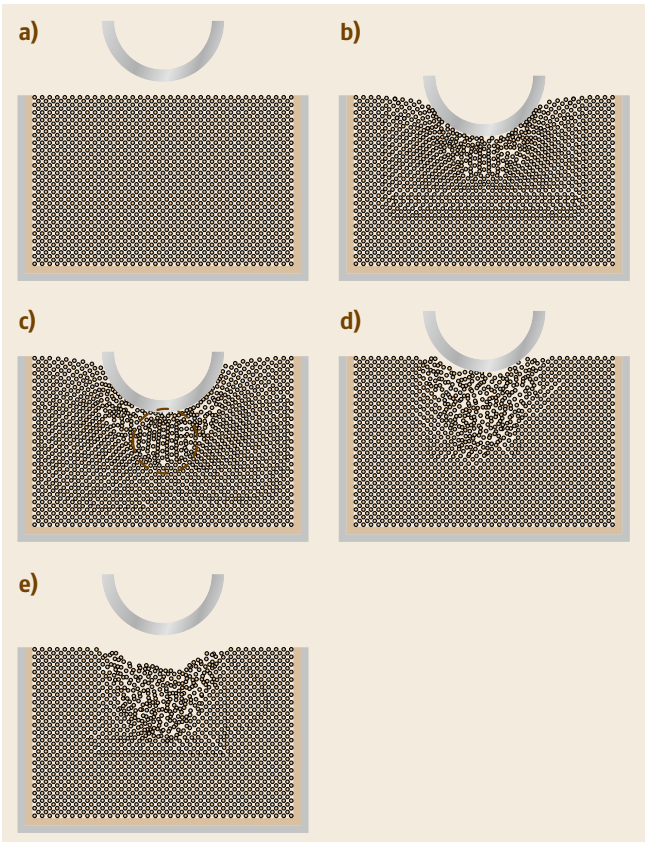


Fig. 30.11 Contact area projected in the plane at a maximum load for simulated indention of a gold surface as a function of in-plane biaxial stress. The stress is normalized to the theoretical yield stress. The *top curve* is from an atomistic simulation; the *bottom curve* is from an elastic model. *Inset*: Illustration of the region near the indention from the simulation. The tip is not shown for clarity. Initial formation of pile-up around the edge of the indentation is apparent



gold surface under an applied in-plane strain [30.142]. These simulations predicted that the mean pressure, calculated from true contact areas that take into account plastic pile-up around the indenter, varies only slightly with applied pre-stress. They also predicted that the higher values occur in compression rather than in tension, and that the modulus calculated from the true contact area is essentially independent of the pre-stress level in the substrate. In contrast, if the contact area is estimated from approximate elastic formulae, the contact area is underestimated, which leads to a strong, incorrect dependence of apparent modulus on the pre-stress level. The simulations also showed larger pile-up in compression than in tension, in agreement with the Pharr model, and both regimes produced contact areas larger than those typically assumed in elastic analyses. These findings are illustrated in Fig. 30.11.

Nanometer-scale indentation of ceramic systems has also been investigated with MD simulations. Ceramics are stiffer and more brittle than metals at the macroscale and examining the nanoindentation of ceramic surfaces provides information about the nanometer-scale properties. They also reveal the manner by which defects form in covalent and ionic materials. For example, *Landman* et al. [30.113, 143] considered the interaction of a CaF_2 tip with a CaF_2 substrate in MD simulations using empirical potentials. As the tip approaches the surface, the attractive force between them steadily increases. This attractive force increases dramatically at the critical distance of 2.3 Å as the interlayer spacing of the tip increases (the tip is elongated) in a process that is similar to the JC phenomenon observed in metals. An important difference, however, is the amount of elongation, which is 0.35 Å in the case of the ionic ceramics and several angstroms in the case of metals. As the distance between the tip and the surface decreases further, the attractive nature of their interaction increases until a maximum value is reached. Indentation beyond this point results in a repulsive tip–substrate interaction, compression of the tip, and ionic bonding between the tip and substrate. These bonds are responsible for the hysteresis predicted to occur in

Fig. 30.12a–e Snapshots of a silicon sample during indentation. The *smaller dots* are diamond atoms. **(a)** Crystalline silicon prior to indentation. **(b)** Atoms beneath the indenter are displaced as a result of indentation. **(c)** The system at maximum indentation. Some of the atoms are in a crystalline arrangement (*circled region*) that is different from the diamond structure. **(d)** The surface structure is largely amorphous as the tip is withdrawn. **(e)** The surface after indentation. Note the amorphous region at the site of the indentation process (after [30.133], © IOP 2000) ◄

the force curve on retraction, which ultimately leads to plastic deformation of the tip followed by fracture.

The responses of covalently bound ceramics such as diamond and silicon to nanoindentation have been heavily studied with MD simulations. One of the first of these computational studies was carried out by *Kallman* et al. who used the Stillinger–Weber potential to examine the indentation of amorphous and crystalline silicon [30.144]. The motivation for this study came from experimental data that indicated a large change in electrical resistivity during indentation of silicon, which led to the suggestion of a load-induced phase transition below the indenter. *Clarke* et al., for example, reported forming an Ohmic contact under load, and using transmission electron microscopy they observed an amorphous phase at the point of contact after indentation [30.147]. Using micro-Raman microscopy, *Kailer* et al. identified a metallic β-Sn phase in silicon near the interface of a diamond indenter during hardness loading [30.148]. Furthermore, upon rapid unloading they detected amorphous silicon as in the *Clarke* et al. [30.147] experiments, while slow unloading resulted in a mixture of high-pressure polymorphs near the indent point. At the highest indentation rate and the lowest temperature, the simulations by *Kallman* et al. [30.144] showed that amorphous and crystalline silicon have similar yield strengths of 138 and 179 kbar, respectively. In contrast, at temperatures near the melting temperature and at the slowest indentation rate, both amorphous and crystalline silicon are predicted to have lower yield strengths of 30 kbar. The simulations thus show how the predicted yield strength of silicon at the

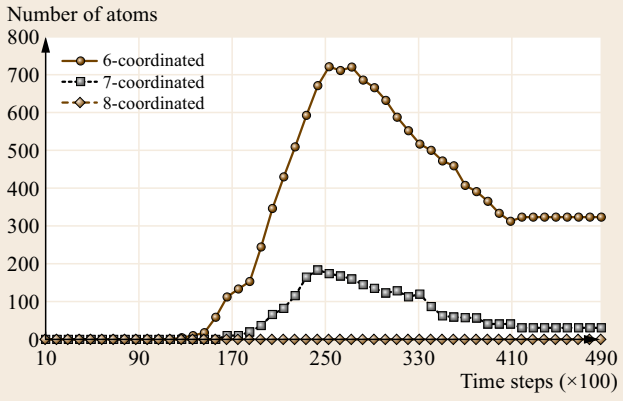


Fig. 30.13 The coordination of the silicon atoms shown in Fig. 30.12 as a function of time during nanoindentation (after [30.145], © IOP 2000)

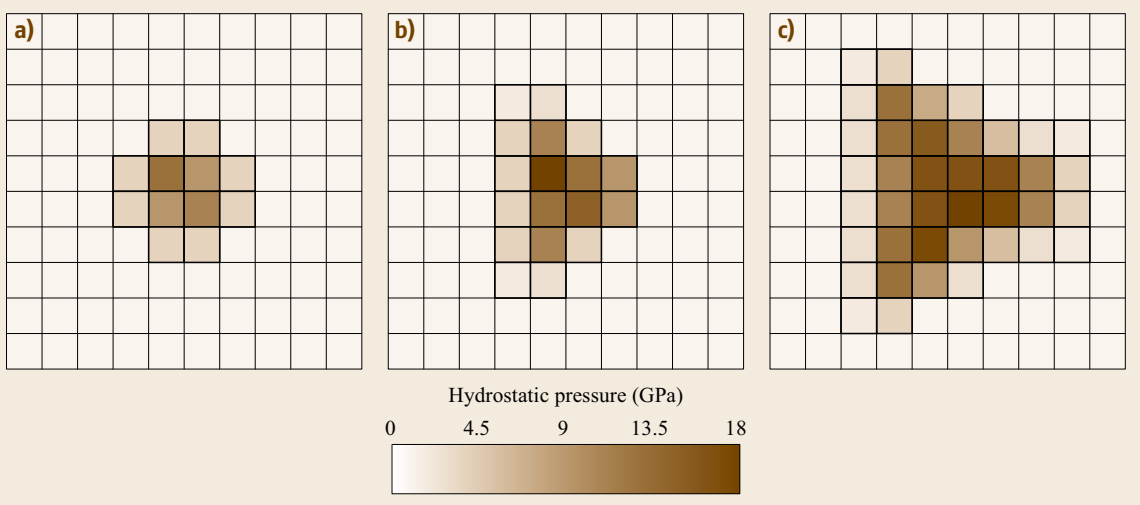


Fig. 30.14a–c Calculated hydrostatic pressure of surface cells at indentation depths of **(a)** 8.9 Å, **(b)** 15.7 Å, and **(c)** 25.3 Å (after [30.146], © IOP 2000)

nanometer scale depends on structure, rate of deformation, and surface temperature.

Interestingly, *Kallman* et al. [30.144] found that amorphous silicon does not crystallize upon indentation, but indentation of crystalline silicon at temperatures near the melting point transforms the surface structure near the indenter to the amorphous phase. The simulations do not predict transformation to the β-Sn structure under any of the conditions considered. These results agree with the outcomes of scratching experiments [30.151] that showed that amorphous silicon emerges from room-temperature scratching of crystalline silicon.

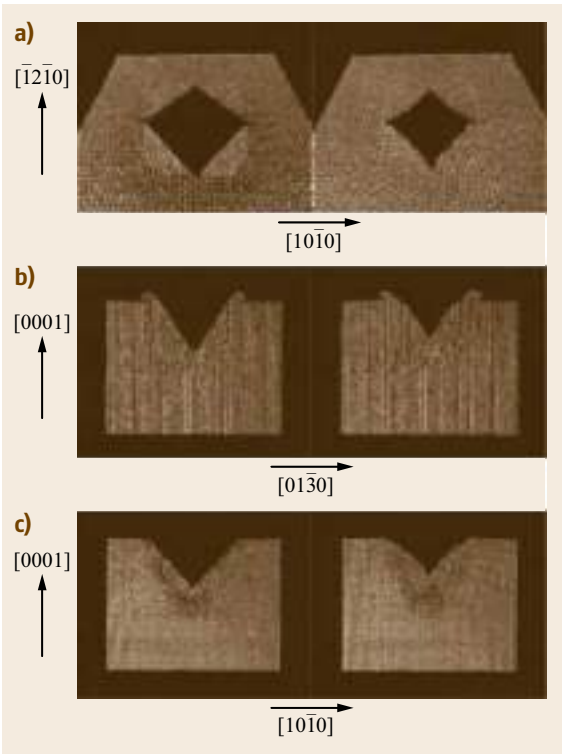


Fig. 30.15a–c Snapshots of the silicon nitride **(a)** surface, **(b)** slide parallel to the edges of the indenter, and **(c)** slide across the indenter diagonal. The *left-hand side* shows the surface when it is fully loaded, while the *right-hand side* shows the surface after the tip has been withdrawn (after [30.149], © AIOP 2003)

Kaxiras and coworkers revisited the silicon nanoindentation issue using a quasi-continuum model that couples interatomic forces from the Stillinger–Weber potential to a finite element grid [30.152]. They report good agreement between simulated loading curves and experiment provided that the curves are scaled by the indenter size. Rather than the β-Sn structure, however, atomic displacements suggest formation of a metallic structure with fivefold coordination below the indenter upon loading, and a residual simple cubic phase near the indentation site after the load is released rather than the mix of high-pressure phases characterized experimentally. Smith et al. attribute this discrepancy to shortcomings of the Stillinger–Weber potential inadequately describing the high-pressure phases of silicon. They also used a simple model for changes in electrical resistivity with loading involving contributions from both a Schottky barrier and spreading resistance. Simulated resistance-versus-loading curves agree well with experiment despite possible discrepancies between the high-pressure phases under the indenter, suggesting that the salient features of the experiment are not dependent on the details of the high-pressure phases produced.

Additional MD simulations of the indentation of silicon were carried out by *Cheong* and *Zhang* [30.145]. Their simulations provide further details about the phase transformations that occur in silicon as a result of nanoindentation. In particular, they find that the diamond cubic silicon is transformed into a body-centered tetragonal structure (β-Si) upon loading of the indenter, as illustrated in Fig. 30.12. Figure 30.13 shows that the coordination numbers of silicon atoms also coincide with that of the theoretical β-Si structure. The body-centered tetragonal structure is transformed into

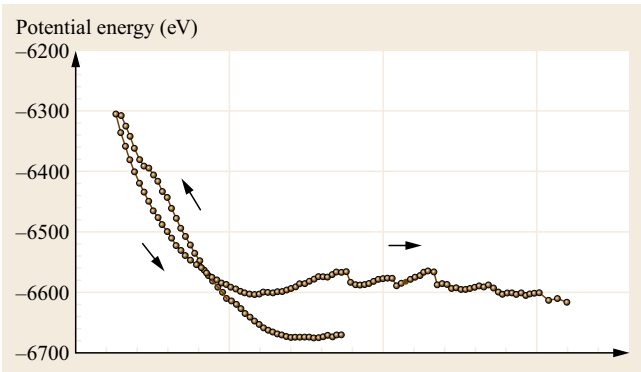


Fig. 30.16 Potential energy as a function of rigid-layer separation generated from an MD simulation of an elastic (nonadhesive) indentation of a hydrogen-terminated diamond (111) surface using a hydrogen-terminated, sp^3-hybridized tip (after [30.150], © Elsevier 1992)

amorphous silicon during the unloading stage. A second indentation simulation again predicted that that this is a reversible process. Atomistic simulations by *Sanz-Navarro* et al. [30.146] shows the relation between the indentation of silicon and the hydrostatic pressure on surface cells due to the nanoindentation, as illustrated in Fig. 30.14. These simulations further predict that the transformation of diamond silicon into the β-Si structure can occur if the hydrostatic pressure is somewhat over 12 GPa.

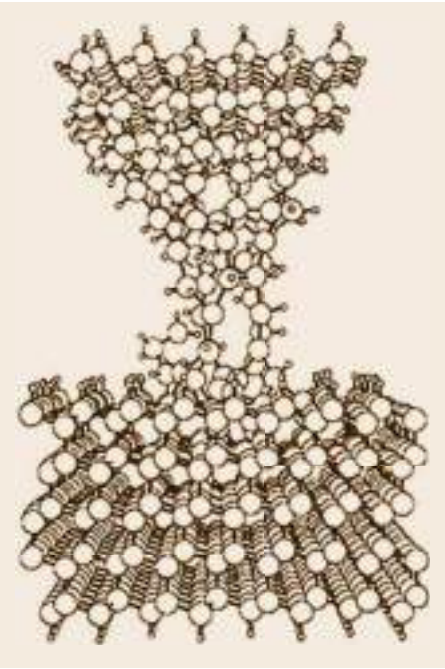


Fig. 30.17 Illustration of atoms in the MD simulation of the indentation of a hydrogen-terminated diamond (111) substrate with a hydrogen-terminated, sp^3-hybridized tip at selected time intervals. The figure illustrates the tip–substrate system as the tip was being withdrawn from the sample. *Large* and *small spheres* represent carbon and hydrogen atoms, respectively (after [30.150], © Elsevier 1992)

Multimillion atom simulations of the indentation of silicon nitride were recently carried out by *Walsh* et al. [30.149]. The elastic modulus and hardness of the surface was calculated using load–displacement relationships. Snapshots from the simulations, illustrated in Fig. 30.15, show that pile-up occurs on the surface along the edges of the tip. Plastic deformation of the surface is predicted to extend a significant distance beyond the actual contact area of the indenter, as illustrated in Fig. 30.15.

The indentation of bare and hydrogen-terminated diamond (111) surfaces beyond the elastic limit was investigated by *Harrison* et al. [30.150] using a hydrogen-terminated sp^3-bonded tip in MD simulations that utilized bond-order potentials. The simulations identified the depth and applied force at which the diamond (111) substrate incurred plastic deformation due to indentation. At low indentation forces, the tip–surface interaction is purely elastic, as illustrated in Fig. 30.16. This finding agrees with the findings of *Cho* and *Joannopoulos* [30.153], who examined the atomic-scale mechanical hysteresis experienced by an AFM tip indenting Si(100) with density functional theory. The calculations predicted that at low rates it is possible to cycle repeatedly between two buckled configurations of the surface without adhesion.

When the nanoindentation process of diamond (111) is plastic, connective strings of atoms are formed between the tip and the surface, as illustrated in Fig. 30.17. These strings break as the distance between the tip and crystal increases and each break is accompanied by a sudden drop in the potential energy at large positive values of tip–substrate separation. The simulations further predict that the tip end twists to minimize interatomic repulsive interactions between the hydrogen atoms on the surface and the tip. This behavior is

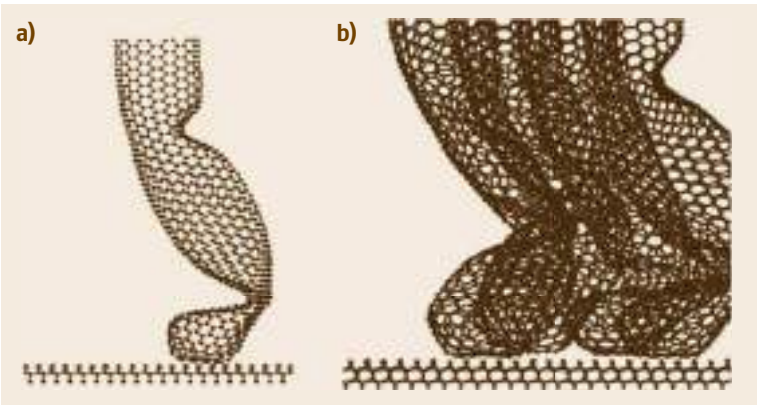


Fig. 30.18a,b Snapshots of the indentation of a single-wall nanotube **(a)** and a bundle of nanotubes **(b)** on hydrogen-terminated diamond (111)

predicted to lead to new covalent bond formation between the tip and the carbon atoms below the first layer of the surface and connective strings of atoms between the tip and the surface when the tip is retracted. Not surprisingly, when the surface is bare and not terminated with hydrogen atoms, the repulsive interactions between the tip and the surface are minimized and the tip indents the substrate without twisting [30.150]. Because carbon–carbon bonds are formed between the tip and the first layer of the substrate, the indentation is ordered (the surface is not disrupted as much by interacting with the tip) and the eventual fracture of the tip during retraction results in minimal damage to the substrate. The concerted fracture of all bonds in the tip gives rise to a single maximum in the potential versus distance curve at large distances.

Harrison et al. [30.154, 155] and *Garg* et al. [30.156, 157] considered the indentation of hydrogen-terminated diamond and graphene surfaces with AFM tips of carbon nanotubes and nanotube bundles using MD simulations and bond-order potentials. Tips consisting of both single-wall nanotubes and multiwall nanotubes were considered. The simulations predicted that nanotubes do not plastically deform during tip crashes on these surfaces. Rather, they elastically deform, buckle, and slip as shown in Fig. 30.18. However, as is the case for diamond tips indenting reactive diamond surfaces discussed above, strong adhesion can occur between the nanotube and the surface that destroys the nanotube in the case of highly reactive surfaces, as illustrated in Fig. 30.19.

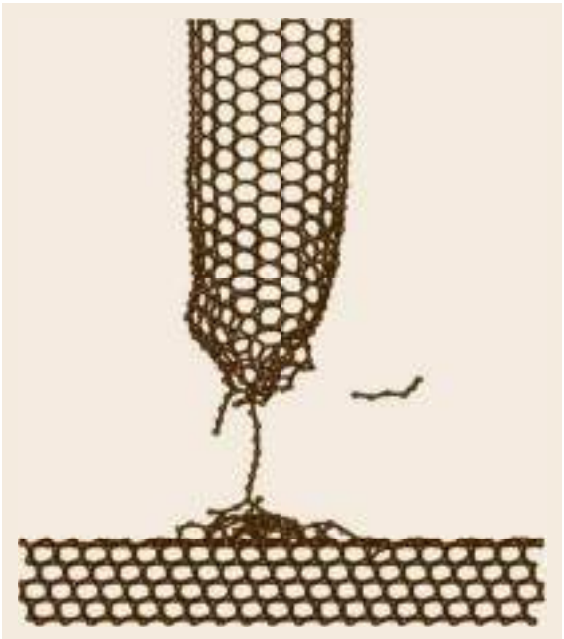


Fig. 30.19 Snapshot of a single-wall carbon nanotube as it is withdrawn following indentation on a bare diamond (111) surface (after [30.156], © APS 1999)

To summarize, MD simulations reveal the properties of ceramic tips and surfaces with covalent or ceramic bonding that are most important for nanometer-scale indentation. They predict that brittle fracture of the tip can occur that is sometimes accompanied by strong adhesion with the surface. They also reveal the conditions under which neither the tip nor the surface is affected by the nanoindentation process. The insight gained from these simulations helps in the interpretation of experimental data, and it also reveals the nanometer-scale mechanisms by which, for example, tip buckling and permanent modification of the surface occur.

30.2.2 Thin Films

In many instances, surfaces are covered with thin films that can range in thickness from a few atomic layers to several μm. These films are more likely to have properties that differ from the properties of bulk materials of similar composition, and the likelihood of this increases as the film thickness decreases. Nanoinden-

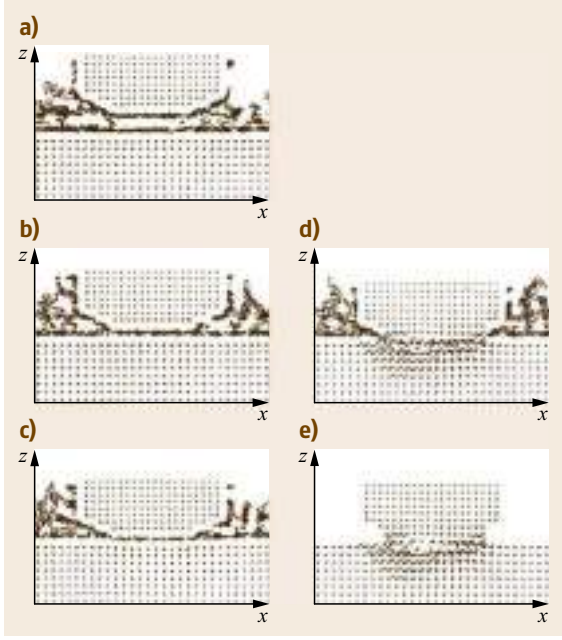


Fig. 30.20a–e Cutaways of the side view from molecular dynamics simulations of a Ni tip indenting a Au(001) surface covered with a hexadecane film. In (e) only the metal atoms are shown. Note how the hexadecane is forced out from between the metal surfaces (after [30.113], © Elsevier 1995)

tation is one of the best approaches to determining the properties of these films. Consequently, numerous computational simulations of this process have been carried out.

For example, MD simulations have been used to study the indentation of metal surfaces covered with liquid *n*-hexadecane films, as illustrated in Fig. 30.20. As the metal tip touches the film, some of the molecules from the surface transfer to the tip and this causes the film to *swell*. As the tip continues to push against the surface, the hydrocarbon film wets the side of the tip. The simulations show how the hydrocarbon film passivates the surface and prevents the strong attractive interactions discussed above for clean metal surfaces and tips from occurring.

In a series of MD simulations, *Tupper* and *Brenner* modeled the compression of a thiol self-assembled monolayer (SAM) on a rigid gold surface using both a smooth compressing surface [30.158] and a compressing surface with an asperity [30.159]. These simulations showed that compression with the smooth surface produced a compression-induced structural change that led to a change in slope of the simulated force versus compression curve. This transition is reversible and involves a change in the ordered arrangement of the sulfur head groups on the gold surface. A similar change in slope seemed to be present in the experimental indentation curves of *Houston* and coworkers [30.160], but was not discussed by the authors. The simulations with the asperity showed that the asperity is able to penetrate the tail groups of the SAM, as illustrated in Fig. 30.21, before an appreciable load is apparent on the compressing surface. This result indicates that it is possible to image the head groups of a thiol self-assembled monolayer that are adsorbed onto the surface of a gold substrate using STM, and consequently ordered images of these systems may not be indicative of the arrangement of the tail groups.

Zhang et al. [30.161] used a hybrid MD simulation approach, where a dynamic element model for the AFM cantilever was merged with a MD relaxation approach for the rest of the system, to study the frictional properties of alkanethiol SAMs on gold. They investigated the effect of several variables like chain length, terminal group, scan direction, and scan velocity. Their results show that friction forces decrease as the chain length of the SAMs increase. In the case of shorter chains such as C_7CH_3, the SAMs near the tip can be deformed by indentation, as illustrated in Fig. 30.22. This behavior is predicted to be the cause of higher friction that occurs for the short-length chains.

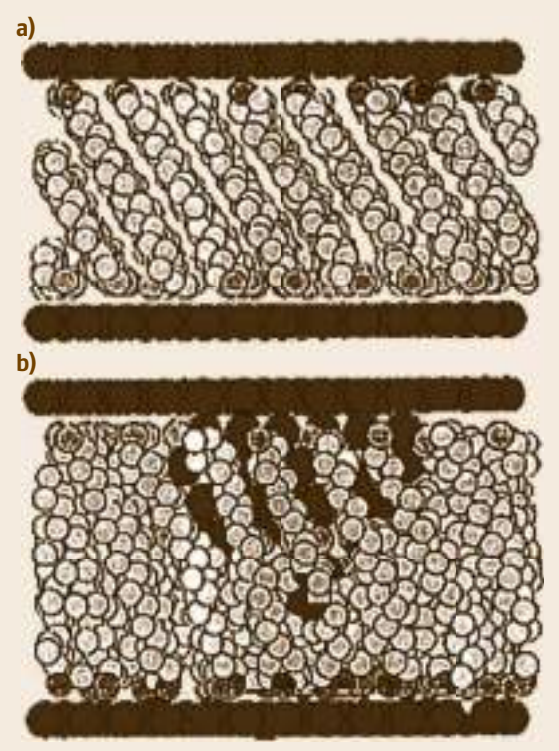


Fig. 30.21a,b Snapshots illustrating the compression of a self-assembled thiol film on gold for a smooth surface **(a)** and a surface containing an asperity **(b)**. The asperity can penetrate and disorder the film tail groups before appreciable load occurs

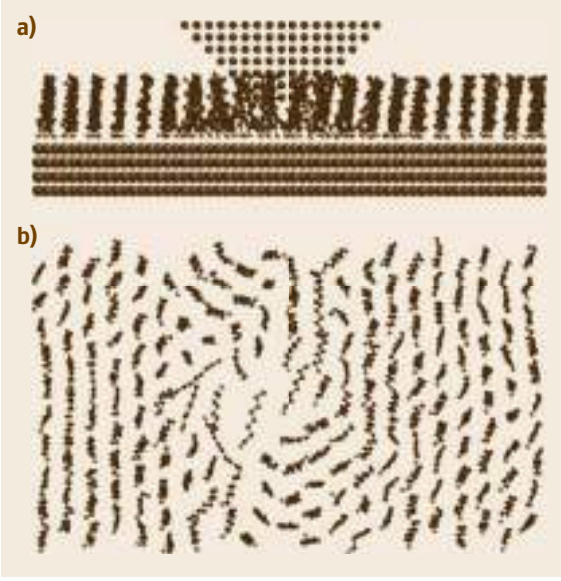


Fig. 30.22 **(a)** Side and **(b)** top views of the final configuration of a C_7CH_3 self-assembled monolayer on Au(111) under a high normal load of 1.2 nN at 300 K. The tip is not shown in **(b)** for clarity (after [30.161], © ACS 2003)

Harrison and coworkers have used classical MD simulations [30.155, 162] to examine the indentation of monolayers composed of linear hydrocarbon chains that are chemically bound (or anchored) to a diamond substrate. Both flexible and rigid single-wall, capped nanotubes were used as tips. The simulations showed that indentation causes the ordering of the monolayer to be disrupted regardless of the type of tip used. Indentation results in the formation of gauche defects within the monolayer and, for deep indents, results in the pinning of selected hydrocarbon chains beneath the tube. Flexible nanotubes tilt slightly as they begin to indent the softer monolayers. This small distortion is due to the fact that nanotubes are stiff along their axial direction and more flexible in the transverse direction. In contrast, when the nanotubes encounter the hard diamond substrate, after *pushing* through the monolayer, they buckle. This process is illustrated in Fig. 30.23 and the force curves are shown in Fig. 30.24. The buckling of the nanotube was previously observed when single-wall, capped nanotubes were brought into contact with hydrogen-terminated diamond (111) surfaces [30.154, 156]. In the absence of the monolayer, the nanotube tips encounter the hard substrate in an almost vertical position. This interaction with the diamond substrate causes the cap of the nanotubes to be *pushed* inside the nanotube (they invert). Increasing the load on the nanotubes causes the walls of the tube to buckle. Both the cap inversion and the buckling are reversible processes. That is, when the load on the tube is removed, it recovers its original shape.

Deep indents of the hydrocarbon monolayers using rigid nanotubes result in rupture of chemical bonds. The

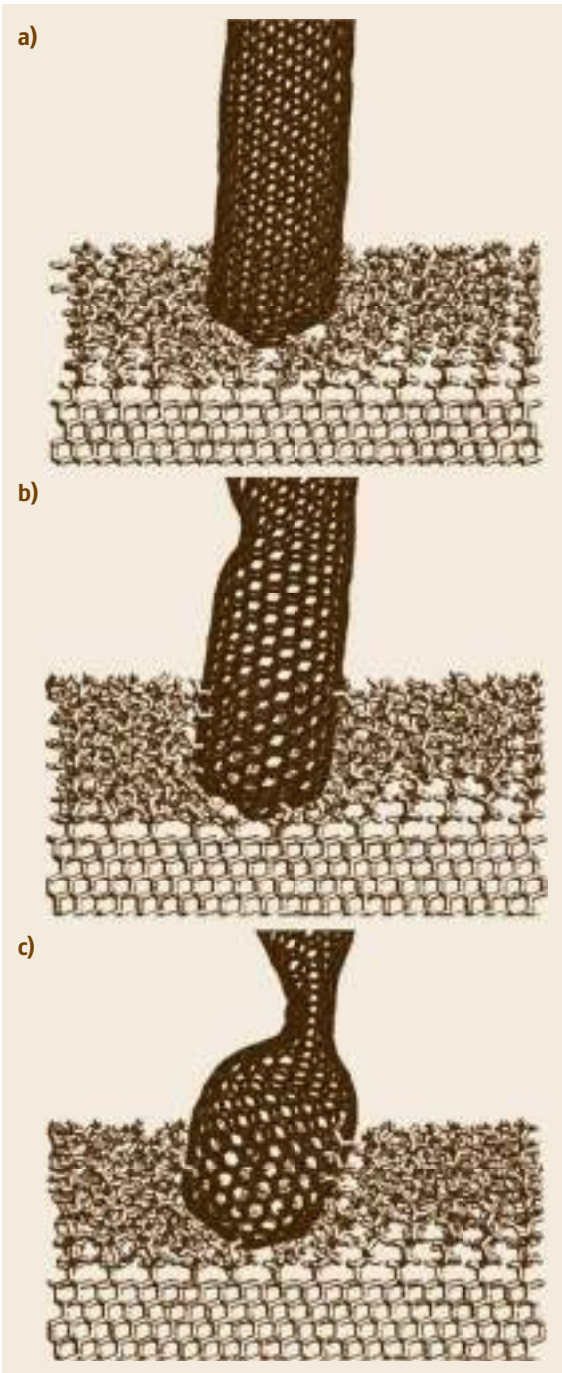


Fig. 30.23a–c Snapshots from the simulation of the interaction of a flexible single-walled carbon nanotube with a monolayer of C_{13} chains on diamond. The loads are **(a)** 19.8 nN, **(b)** 41.2 nN, and **(c)** 36.0 nN (after [30.162], © ACS 2003) ◄

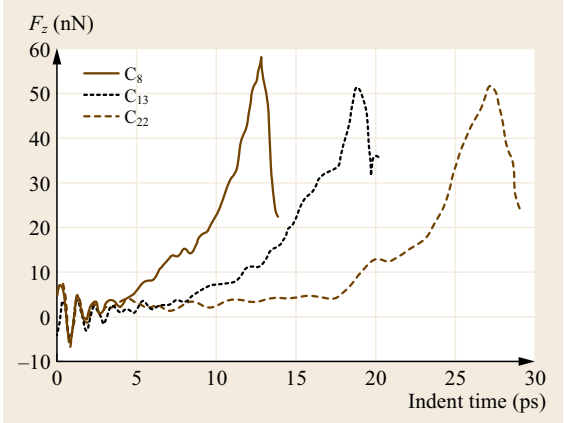


Fig. 30.24 The load on the upper two layers of the flexible carbon nanotube indenter shown in Fig. 30.23 as a function of indentation time for the nanoindentation of the indicated hydrocarbon monolayes on diamond (after [30.162], © ACS 1999)

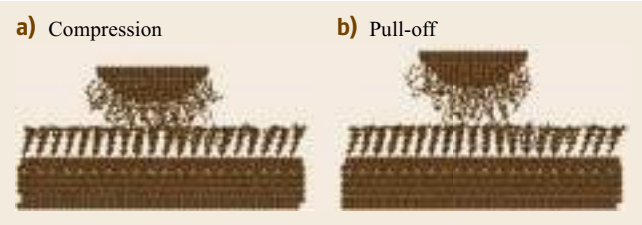


Fig. 30.25 Snapshots from the OH/OH pair interaction during **(a)** compression and **(b)** pull-off (after [30.163], © ACS 2002)

simulations also show that the number of gauche defects generated by the indentation is a linear function of penetration depth and equal for C_{13} and C_{22} monolayers. Thus, it is the tip that governs the number of gauche defects generated.

Leng and *Jiang* [30.163] investigated the effect of using tips coated with SAMs containing hydrophobic methyl (CH_3) or hydrophilic hydroxyl (OH) terminal groups to nanoindent gold surfaces that also are covered with SAMs with the identical terminal groups as the tip. Figure 30.25 contains snapshots for the indentation process predicted to occur for terminal OH/OH interactions during compression and the pull-off. The adhesion force of OH/OH pairs is calculated to be about four times larger than that of CH_3/CH_3 pairs, as shown in Fig. 30.26. This is due to the formation of hydrogen bonding between OH/OH pairs. This interaction is also expected to increase the frictional force between monolayers with OH terminations.

Related MD simulations by *Mate* predict that the end groups on polymer lubricants have a significant influence on the lubrication properties of polymers [30.164]. For instance, fluorinated end groups are predicted to be less reactive than regular alcohol end groups. When fluorinated films are indented, the normal force becomes more attractive as the distance between the tip and film decreases until the hard wall limit is reached and the interactions become repulsive. In contrast, when AFM tips indent hydrogenated films, the forces become increasingly repulsive as the distance between them decreases, as shown in Figs. 30.27 and 30.28. This predicted behavior is due to the compression of the end group beneath the tip. For the lubricant molecules to be squeezed out from between the tip and the surface, the hydrogen bonding between the two must first be broken and this increases the force needed to indent the system. As a result, a major effect of the presence of alcohol end groups is to dramatically increase the load that a liquid lubricant can support before failure (solid–solid contact) occurs.

When atomically sharp tips are used to indent solid-state thin films where there is a large mismatch in the mechanical properties of the film and the substrate, it is difficult to determine the true contact area between the tip and the surface during nanoindentation. In the case of soft films on hard substrates, pile-up can occur around the tip that effectively increases the contact area. In contrast, with hard films on soft substrates, *sink-in* is

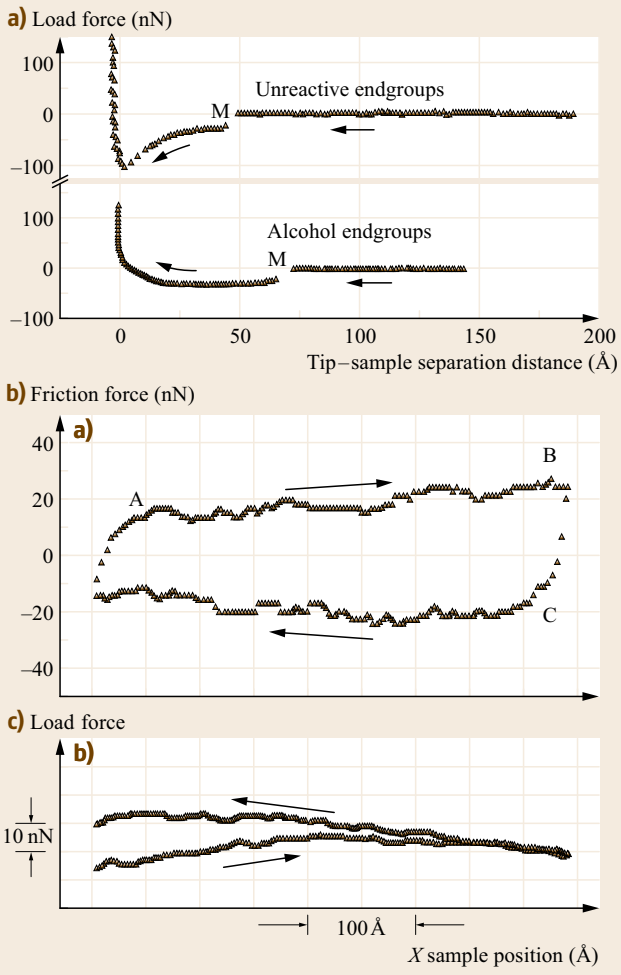
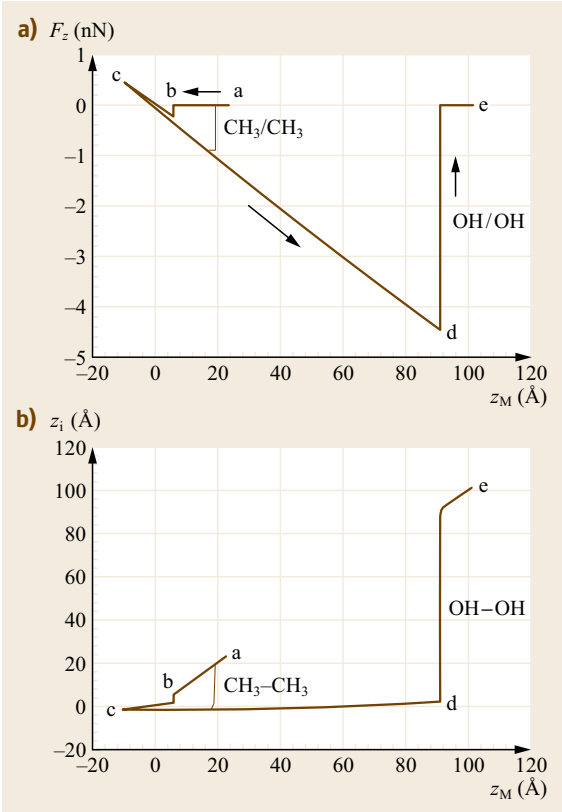


Fig. 30.26 **(a)** The force versus distance curve (indentation part only) for unbonded perfluoropolyether on Si(100). The unreactive end groups were from a 10 Å-thick film; the reactive alcohol end groups were from a 30 Å-thick film. The negative forces represent attractive interactions between the tip and the surface. Measured plots of friction **(b)** and load forces **(c)** of the tip as it slides over the sample with the alcohol end groups (after [30.164], © APS 1992)



experienced around the tip that decreases the true contact area.

A class of coatings that has received much attention is diamondlike amorphous carbon (DLC) coatings. DLC coatings are almost as hard as crystalline diamond and may have very low friction coefficients (< 0.01) depending upon the growth conditions [30.165–168]. They have therefore generated much interest in the tribological community and there have been several MD simulation studies to determine the mechanical and atomic-scale frictional properties of DLC coatings. MD simulations with bond-order potentials by *Sinnott* et al. [30.169] examined the differences in indentation behavior of a hydrogen-terminated diamond tip on hydrogen-terminated single-crystal diamond surfaces and diamond surfaces covered with DLC. In the former case, the tip goes through shear and twist deformations at low loads that change to plastic deformation and adhesion with the surface at high loads. When the surface is covered with the DLC film, the tip easily penetrates the film, as illustrated in Fig. 30.29, which *heals* easily when the tip is retracted so that no crater or other evidence of the indentation is left behind.

Fig. 30.27 **(a)** Force–distance curves and **(b)** tip position (z_i) versus support position (z_M) for the OH/OH contact pair and the CH_3/CH_3 contact pair (after [30.163], © ACS 2002) ◀

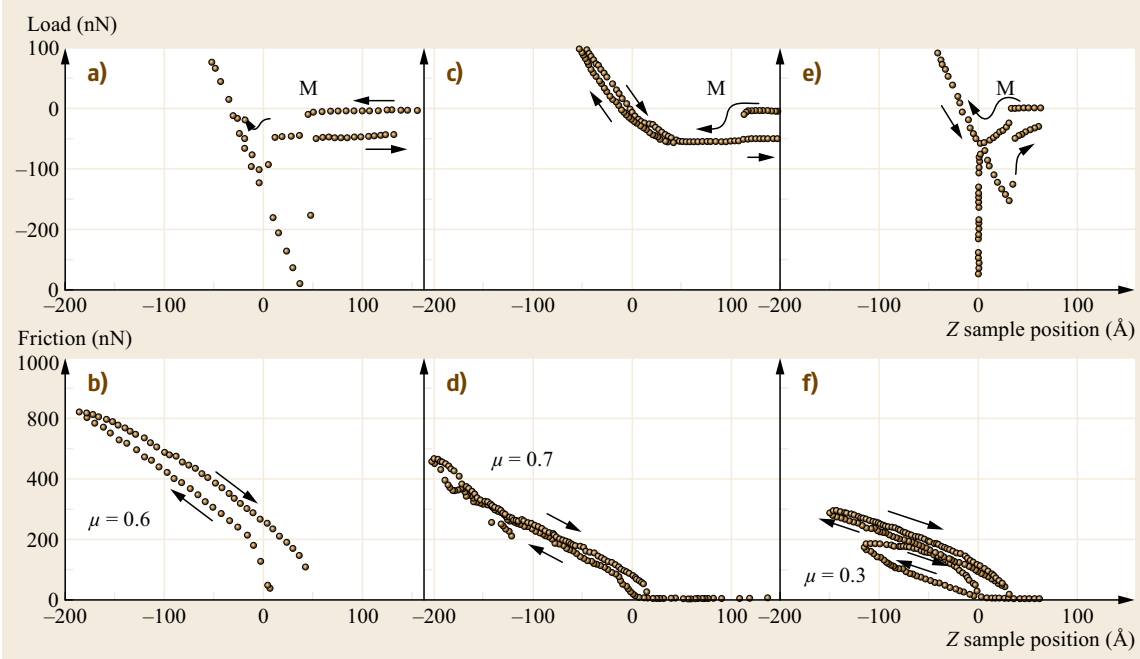


Fig. 30.28a–f Measured values for friction and load as an atomic-force microscope tip is scanned across a 30 Å-thick sample of perfluoropolyether on Si(100). **(a,b)** The unbonded polymer with unreactive end groups. **(c,d)** The unbonded polymer with alcohol end groups. **(e,f)** A bonded polymer (after [30.164], © APS 1992)

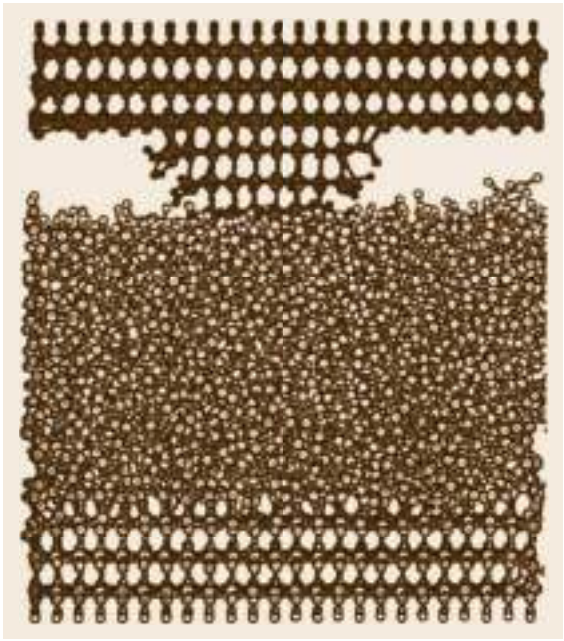


Fig. 30.29 Snapshot from a molecular dynamics simulation where a pyramidal diamond tip indented an amorphous carbon thin film that is 20 layers thick. The simulation took place at room temperature and the carbon atoms in the film were 21% sp^3-hybridized and 58% sp^2-hybridized (the remaining atoms were on the surface and were not counted) (after [30.169], © AIP 1997) ◄

MD simulations by *Glosli* et al. [30.170] of the indentation of DLC films that are about 20 nm-thick give similar results. In this case a larger, rigid diamond tip was used in the indentations and was also slid across the surface. During sliding, the tip plows the surface, which causes some changes to the film not seen during indentation. However, because the tip is perfectly rigid, adhesion between the film and surface is not allowed which influences the results.

This section shows that repulsive interactions between surfaces covered with molecular films and proximal probe tips are minimized relative to interactions between bare surfaces and indentation tips. The lubrication properties of polymers and SAMs can vary with chain length, the rigidity of the tip, and the chemical properties of the end groups. In some cases, indentations can disrupt the initial ordering of polymers and SAMs, which affects their responses to nanoindentation and friction.

30.3 Friction and Lubrication

Work is required to slide two surfaces against one another. When the work of sliding is converted to a less ordered form, as required by the first law of thermodynamics, friction will occur. For instance, if the two surfaces are strongly adhering to one another, the work of sliding can be converted to damage that extends beyond the surfaces and into the bulk. If the adhesive force between the two surfaces is weaker, the conversion of work results in damage that is limited to the area at or near the surface and produces transfer films or wear debris [30.171, 172]. While the thermodynamic principles of the conversion of work to heat are well known, the mechanisms by which this takes place at sliding surfaces are much less well established despite their obvious importance for a wide variety of technological applications.

Atomic-scale simulations of friction are therefore important tools for achieving this understanding. They have consequently been applied to numerous materials in a wide variety of structures and configurations, including atomically flat and atomically rough diamond surfaces [30.173–175], rigid substrates covered with monolayers of alkane chains [30.176], perfluorocarboxylic acid and hydrocarboxylic Langmuir–Blodgett (L–B) monolayers [30.177], between contacting copper surfaces [30.178, 179], between a silicon tip and a silicon substrate [30.119, 143], and between contacting diamond surfaces that have organic molecules absorbed on them [30.180]. These and several other studies are discussed below.

30.3.1 Bare Surfaces

Sliding friction that takes place between two surfaces in the absence of lubricant is termed *dry* friction even if the process occurs in an ambient environment. Simple models have been developed to model dry sliding friction that, for example, consider the motion of a single atom over a monoatomic chain [30.181]. Results from these models reveal how elastic deformation of the substrate from the sliding atom affects energy dissipation and how the average frictional force varies with changes in the force constant of the substrate in the direction normal to the scan direction. Much of the correct behavior involved in dry sliding friction is captured by these types of simple models. However, more detailed models and simulations, such as MD simulations, are required to provide information about more complex phenomena.

MD simulations have been used to study the sliding of metal tips across clean metal surfaces by numerous groups [30.179, 182–186]. An illustrative case is shown in Fig. 30.30 for a copper tip sliding across a copper surface [30.179]. Adhesion and wear occur when the attractive force between the atoms on the tip and the atoms at the surface becomes greater than the attractive forces within the tip itself. Atomic-scale stick and slip can occur through nucleation and subsequent motion of dislocations, and wear can occur if part of the tip gets left behind on the surface (Fig. 30.30). The simulations can further provide data on how the characteristic *stick–slip* friction motion can depend on the area of contact, the rate of sliding, and the sliding direction (Fig. 30.31).

An additional study of stick–slip in the sliding of much larger, square-shaped metal tips across metal surfaces was carried out by *Li* et al. [30.187] using EAM potentials. The initial structure of a NiAl tip and surface system is shown in Fig. 30.32. This study predicted that collective elastic deformation of the surface layers in response to sliding is the main cause of the stick–slip behavior shown in Fig. 30.33. The simulations also predicted that stick–slip produces phonons that propagate through the surface slab.

Large-scale simulations using pairwise Morse potentials that are similar in form to (30.6) were used to study the wear of metal surfaces caused by metal tips that plow the surface, as illustrated in Fig. 30.34. They provide insight into the wear track dependence of the sliding rate [30.188] and how variations in the scratching force, friction coefficient, and other quantities depend on the scratch depth [30.189], as illustrated in Fig. 30.35.

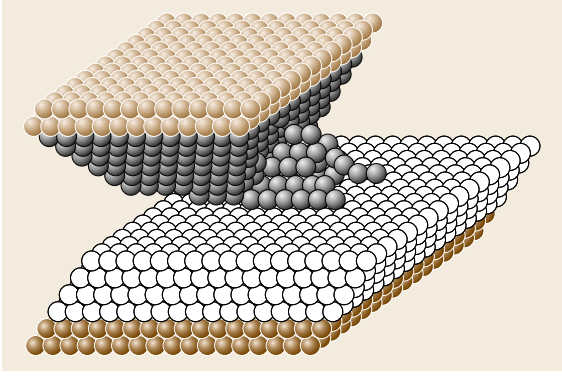


Fig. 30.30 Snapshot from a molecular dynamics simulation of a copper tip sliding across a Cu(100) surface. A connective neck between the two is sheared during the sliding, leading to wear of the tip. The simulation was performed at a temperature of 0 K (after [30.179], © APS 1996)

On the whole, the results of experimental studies show good agreement with the results of the computational studies described above. This is true despite the fact that all of these MD simulations use empirical potentials that do not include electronic effects and thus effectively assume that the electronic contributions to friction on metal surfaces are negligible. However, experiments have measured a nonnegligible contribution of conduction electrons to friction [30.190]. Thus, future simulations of metal-tip–metal-substrate interactions using more sophisticated tight-binding or first principles methods that include electronic effects are encouraged.

Layered ceramics, such as mica, graphite and MoS_2, that have structures that include strongly bound layers that interact with one another through weak van der Waals bonds, have long been known to have good lubricating properties because of the ease with which the layers slide over one another. They have, therefore, been the focus of some of the earliest experimental studies of nanometer-scale friction [30.19, 191]. The results of these early studies lead researchers to hypothesize that at high loads measured friction forces were related to *incipient sliding* [30.192, 193] caused by a small flake from the surface becoming attached to the end of the

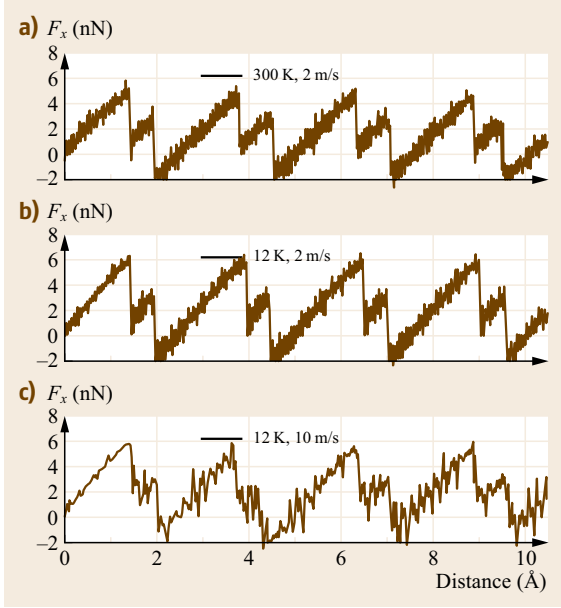


Fig. 30.31a–c Plots of the lateral force versus distance from a simulation similar to that shown in Fig. 30.30. The plots illustrate the dependence of the force on temperature and sliding velocity. **(a)** Temperature of 300 K and a sliding velocity of 2 m/s, **(b)** temperature of 12 K and a sliding velocity of 2 m/s, and **(c)** temperature of 12 K and a sliding velocity of 10 m/s (after [30.179], © APS 1996)

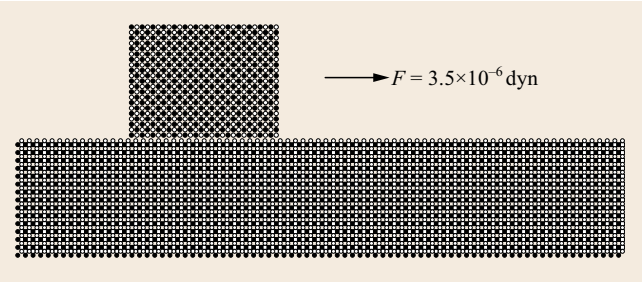


Fig. 30.32 Starting configuration of sliding NiAl tip on a NiAl surface (after [30.187], © AIP 2001)

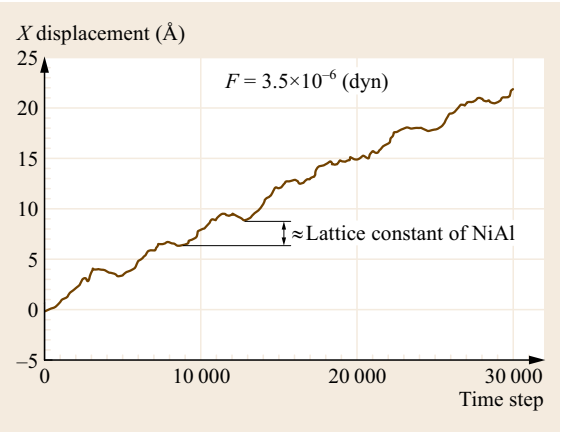


Fig. 30.33 A structured curve of frictional dynamics of an atom in the *upper right corner* that is indicative of stick–slip (after [30.187], © AIP 2001)

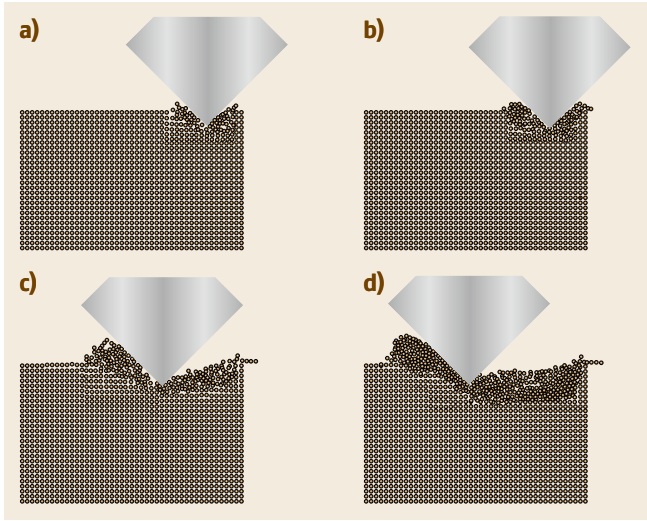


Fig. 30.34a–d Snapshots of the scratching of an aluminum surface with a rigid tip at a depth of 0.8 nm (after [30.189], © APS 2000)

tip. If true, this would mean that all measured interactions were between the surface and the flake, which has a larger contact area than the clean tip. However, subsequent simulations of constant force AFM images of graphite by *Tang* et al. [30.194] showed that there is no need for the assumption of a graphite flake under the tip to reproduce the experimental images of a graphite surface.

Surprisingly strong localized fluctuations in atomic-scale friction are displayed by layered ceramics [30.195–198]. For instance, square-well signals with sub-angstrom lateral width are obtained in FFM scans on $MoS_2(001)$ in the direction across the scan direction, while sawtooth signals are detected along the scan direction, as shown in Fig. 30.36. This finding can be explained by a stick–slip model by *Mate* et al. [30.19] and *Erlandsson* et al. [30.191] that assumes that the tip does a zigzag walk along the scan. Measured variations in the frictional force with the periodicity of cleavage planes [30.191] are consistent with the results of this simple model. However, additional experiments indicate a more complex tip–surface interaction, such as changes in the intrinsic lateral force between the substrate and the AFM tip [30.199] or sliding-induced chemistry between the tip and the surface [30.200, 201].

MoS_2 has proven to be a very successful solid lubricant for applications that operate in vacuum but its performance quickly deteriorates when exposed to ambient air. Recently, *Liang* et al. used ab initio DFT methods to examine the potential energy surfaces between sliding $MoS_2(001)|MoS_2(001)$, $MoS_2(001)|MoO_3(001)$, and $MoO_3(001)|MoO_3(001)$ interface systems in an effort to understand the deterioration in lubricity due to oxidation [30.202]. The potential energy surfaces give information on the minimum energy path along particular sliding directions from which lateral forces needed to slide the interface can be calculated. In this work a normal force of 500 MPa was applied to all three interfacial systems before the energy surface calculations were performed. It was found that the minimum energy path for self-mated $MoS_2(001)|MoS_2(001)$ was a zigzag path that avoided direct overlap of sulfur atoms in the topmost layers of the top and bottom surfaces. From this PE surface a lateral frictional force of 0.058 nN/atom was predicted for this interface. The

lowest frictional force of 0.011 nN/atom predicted in these calculations was for the $MoS_2(001)|MoO_3(001)$ interface along the channel direction formed by S atoms

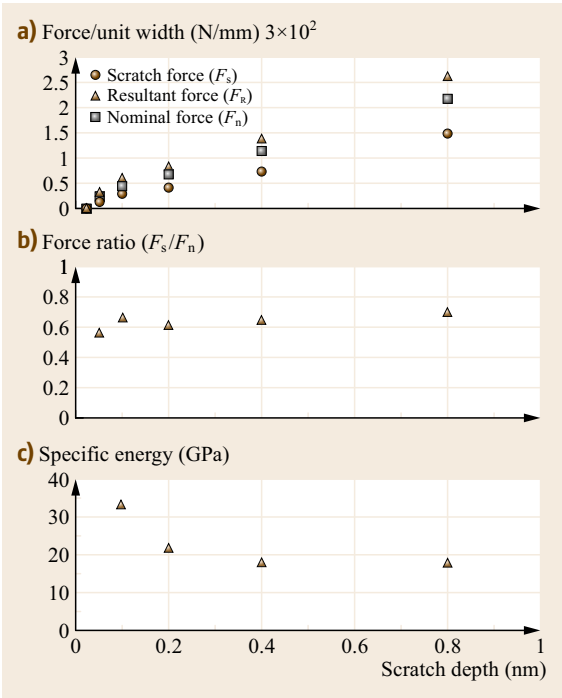


Fig. 30.35 Variation in **(a)** the scratching force, the normal force, and the resultant force, and **(b)** the friction coefficient, and **(c)** the specific energy during scratch processes similar to those shown in Fig. 30.34 at scratch depths ranging from 0.8 to almost 0 nm (after [30.189], © APS 2000)

at the sliding surface. Although this doesn't explain the mechanism by which degradation occurs experimentally, it is in general agreement with experimental results that show that this interface can produce lower frictional coefficients than pure MoS_2. As suggested by the authors, it may be a surface defect driven process that is currently inaccessible to first principles methods employed in these calculations. The last interface of $MoO_3(001)|MoO_3(001)$ was found to have the largest frictional force of 0.352 nN/atom.

Crystalline ceramics differ from layered ceramics in that they are held together by relatively strong covalent or ionic bonds. In the case of ionic systems, *Shluger* et al. [30.203] used a mixture of atomistic and macroscopic modeling methods to study the interaction of a MgO tip and a LiF surface. In particular, the tip–surface interaction was treated atomistically and the cantilever deflection was treated with a macroscopic approach. The results, shown in Fig. 30.37, show that if the tip is charged and in hard contact with the surface, tip and surface distortions are possible that can lead to motion of the surface ions within the surface plane and the transfer of some of the ions onto the tip.

In the case of covalently bound ceramics, there is extensive literature related to friction of diamond [30.204, 205] because, while it is the hardest material known, it also exhibits relatively low friction. The *ratchet mechanism* has been proposed for energy dissipation during friction on the macroscale in diamond, where energy is released by the transfer of normal force from one surface asperity to another. The elastic mechanism is another mechanism that has been proposed, where the released energy comes from

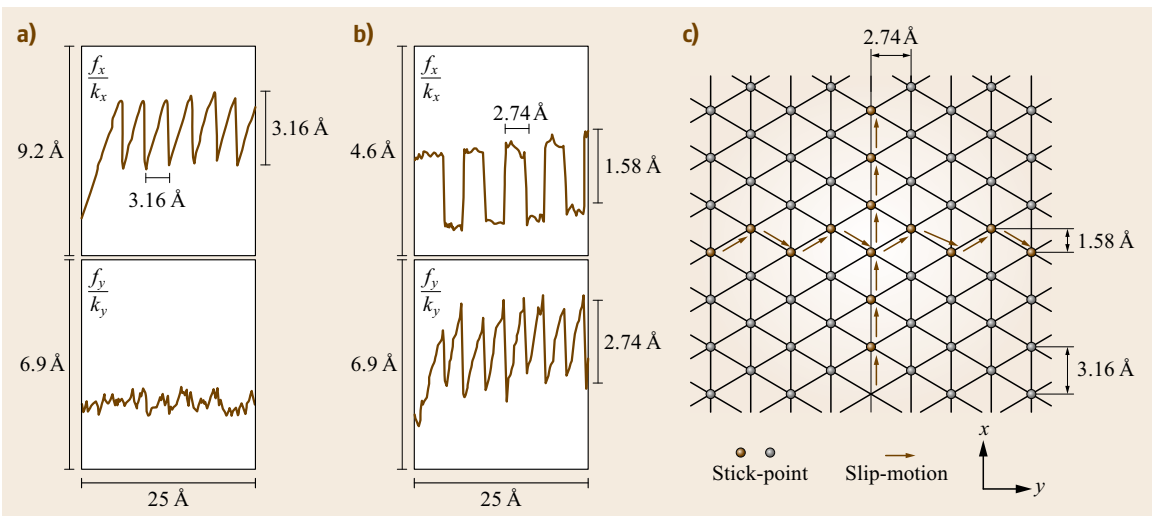


Fig. 30.36a–c Displacement data from a scan across a $MoS_2(001)$ surface. The data in **(a)** and **(b)** are form scans along the x -and y-directions, respectively, on the surface shown in **(c)** (after [30.198], © APS 1996)

elastic strain in an asperity. Atomic-scale friction has been measured experimentally [30.20] for diamond tips with near atomic-scale radii sliding over hydrogen-terminated diamond surfaces. These experiments are sensitive enough to detect the 2×1 reconstruction on the diamond (100) surface. Furthermore, the average friction coefficient determined with an AFM on H-terminated diamond (111) surfaces is about two orders of magnitude smaller than the value measured on bare, 2×1 diamond (111) surfaces, indicating greater adhesion in the latter case [30.206]. More recently, the friction between a tungsten carbide tip and hydrogen-terminated diamond (111) was examined with AFM in UHV by *Enachescu* et al. [30.207]. The friction between these two hard surfaces was shown to obey Derjaguin–Muller–Toporov or DMT [30.208] contact mechanics and the shear strength of the interface was determined to be 246 MPa.

Extensive MD simulations have been carried out by *Harrison* and coworkers that examine the friction between hydrogen-terminated diamond (111) surfaces [30.173, 209] and diamond (100) surfaces [30.210] in sliding contact and its temperature dependence [30.211]. The simulations of sliding between the diamond (111) surfaces reveal that the potential energy, load, and friction are all periodic functions of the sliding distance (Fig. 30.38). Maxima in these quantities occur when the hydrogen atoms on opposing surfaces interact strongly. Recent ab initio studies by *Neitola* and *Pakanen* of the friction between hydrogen-terminated diamond (111) surfaces also show that the potential energy is periodic with sliding distance (Fig. 30.39) [30.212]. Because the results of the ab initio studies and the MD simulations are in good agreement, Neitola and Park conclude that the potential model used in the MD studies is accurate.

As mentioned previously, the maxima in the load and the friction values during sliding are caused by the interactions of hydrogen atoms on opposing surfaces.

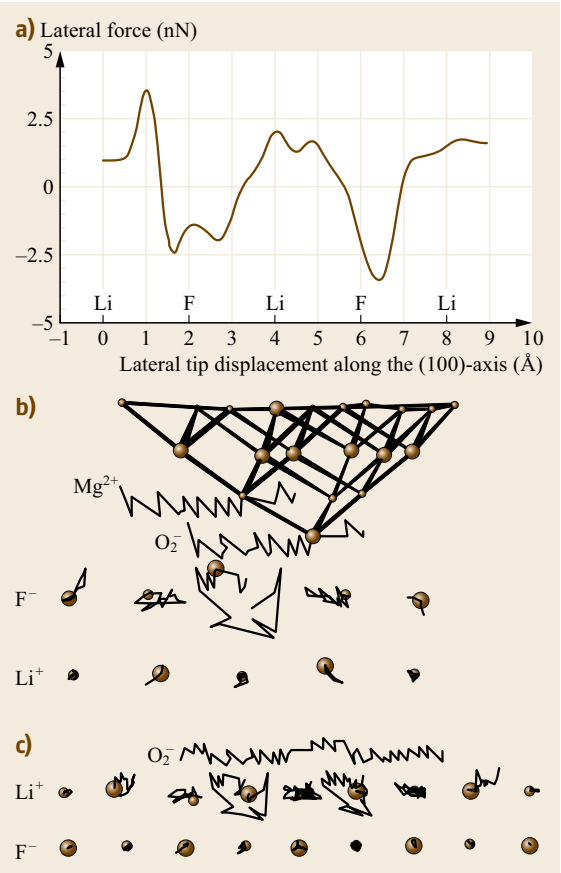


Fig. 30.37 **(a)** The lateral force calculated for a MgO tip scanning in the ⟨001⟩-direction on LiF(001). **(b,c)** A view of the side of the surface plane along the scan direction. The surface Li^+ and F^- atoms are seen to relax to relieve the frictional energy and this relaxation motion is indicated in the figure by the *lines*. **(b)** An ion on the surface can be moved into an interstitial site by the tip and then returns to its original position. **(c)** Reversible relaxation of the surface atoms (after [30.203], © Elsevier 1995)

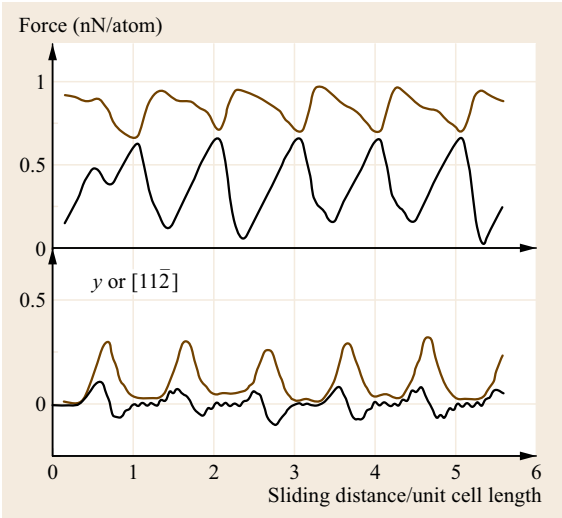


Fig. 30.38 Calculated frictional force (*lower lines*) and normal force (*upper lines*) felt by a hydrogen-terminated (111) surface as it slides against another hydrogen-terminated diamond (111) surface in a MD simulation. The sliding direction is given in the legend. The sliding speed is 1 Å/ps and the simulation temperature is 300 K. The two plots show how the simulated stick–slip motion changes as a function of the applied load. The load is high and low in the *upper* and *lower panels*, respectively (after [30.210], © ACS 1995)

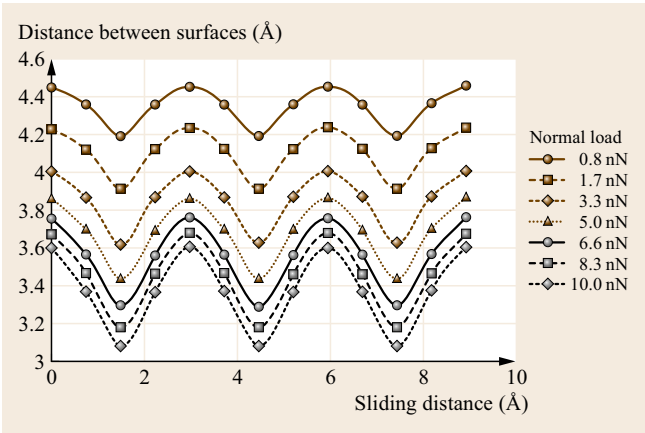


Fig. 30.39 Distance between hydrogen-terminated (111) crystals as a function of sliding distance (after [30.212], © ACS 2001)

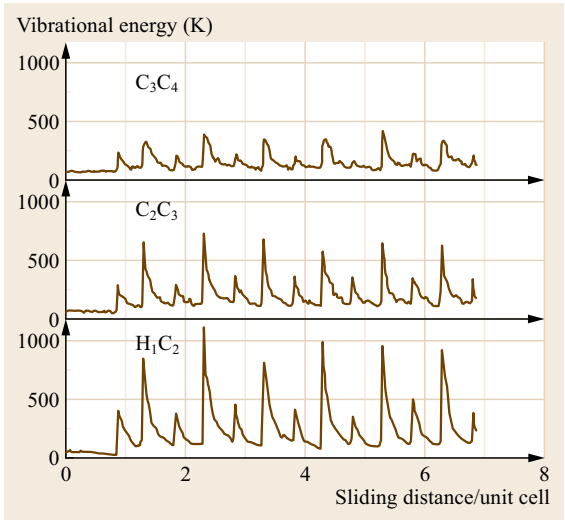


Fig. 30.40 Average vibrational energy of oscillators between diamond layers as a function of sliding distance. These energies are derived from a molecular dynamics simulation of the sliding of a hydrogen-terminated diamond (111) surface over another hydrogen-terminated diamond (111) surface. The vibrational energy between the first and second layers of the lower diamond surface is shown in the *lower panel*, between the second and third layers in the *middle panel*, and between the third and fourth layers in the *upper panel* (after [30.213], © Elsevier 1995)

When sliding in the $[11\bar{2}]$ direction, the H atoms *revolve* around one another, thus decreasing the repulsive interaction between the sliding surfaces because the hydrogen atoms are not forced to pass directly over one another [30.173]. Increasing the load causes increased stress at the interface. The opposing hydrogen atoms become *stuck*. Once the stress at the interface becomes large enough to overcome the hydrogen–hydrogen interaction between opposing surfaces, the hydrogen atoms *slip* past one another with the same *revolving* motion observed at low loads. This phenomenon is known as atomic-scale stick–slip and has the periodicity of the diamond lattice. It should be noted that due to the alignment of the opposing surfaces, the hydrogen atoms are directly in line with each other when sliding in the $[11\bar{2}]$ direction. However, the hydrogen atoms are not *aligned* with each other when sliding in the $[1\bar{1}0]$, so the friction in this direction is lower than in the $[11\bar{2}]$ direction. It should be noted, however, that experimentally all initial alignments are likely to be probed.

Harrison and coworkers have further shown that the peaks in the frictional force are correlated with peaks in the temperature of the atoms at the interface when two hydrogen-terminated diamond (111) surfaces are in sliding contact [30.209]. Figure 30.40 shows the vibrational energy (or temperature) between diamond layers as a function of sliding distance. These data clearly show that layers close to the sliding interface can be vibrationally excited during sliding. When the hydrogen atoms are *stuck* or interacting with each other strongly, the stress and friction force at the interface build up. When the hydrogen atoms *slip* past one another, the stress at the interface is relieved and the energy is transferred to the diamond in the form of vibration or heat. Thus, the peaks in the temperature occur slightly after the peaks in the frictional force.

It should be noted that atomic-scale stick slip is observed in other systems. *Perry* and *Harrison* used MD simulations to demonstrate that two hydrogen-terminated diamond (100) (2×1) surfaces in sliding contact also exhibit stick–slip [30.210]. In addition, it was shown that the shape of the friction versus sliding distance curves is influenced slightly by the speed of the sliding, with features in the curves becoming more pronounced at slower speeds. Stick–slip behavior was also observed in AFM studies of diamond (100) (2×1) surfaces [30.206]. However, in this case the stick–slip was over a much longer length scale and may be due to the fact that the surfaces were not hydrogen-terminated.

Mulliah et al. [30.214] used MD simulations with bond-order potentials [30.215] to model interactions between indenter atoms, EAM potentials [30.216] to model interactions between substrate atoms, and the Ziegler–Biersack–Littmack potential [30.217] to model interactions between indenter and substrate atoms to study the atomic-scale stick–slip phenomenon of a pyramidal diamond tip interacting with a silver surface at several sliding rates and vertical support displacements. These simulations showed that dislocations are related to the stick events emitting a dislocation in

the substrate near the tip. The scratch in the substrate is discrete due to the tip jumping over the surface in the case of small vertical displacements. In contrast, large displacements of 15 Å or more result in a continuous scratch. These simulations also showed how the dynamic friction coefficient and the static friction coefficient increase with increasing tip depth. The tip moves continuously through a stick and slip motion at large depths, whereas it comes to a halt in the case of shallow indents. Although the sliding rate can change the exact points of stick and slip, the range of sliding rates over the range of values considered in this study (1.0–5.0 m/s) has no influence on the damage to the substrate, the atomistic stick–slip mechanisms, or the calculated friction coefficients.

The effect of the way in which the tip is rastered across the surface in MD simulations was considered by *Cai* and *Wang* [30.218, 219] using bond-order potentials. In particular, they dragged silicon tips across several silicon surfaces, as illustrated in Fig. 30.41, in two different ways. In the first, they moved the tip every MD step while in the other they advanced the tip every 1000 steps. In both cases, the overall sliding rate is the same and equals 1.67 m/s. In both cases, wear of the tip such as is illustrated in Fig. 30.41 occurs. However, the mechanisms by which the wear occurs are found to depend on the approach used, and the latter approach is found to be in better agreement with experimental data.

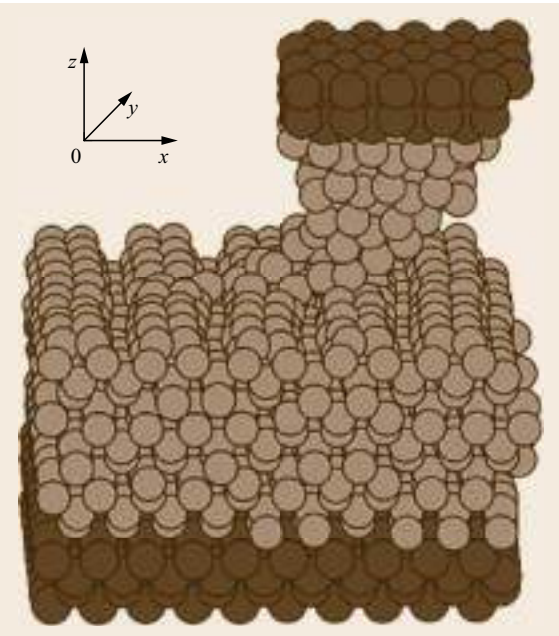


Fig. 30.41 Snapshots of a Si(111) tip interacting with a Si(001) 2 × 1 surface. The tip is rastering along the surface in the *x*-direction and starts off at a distance of 9 Å from the surface (after [30.218], © CCLRC 2002)

In many studies, diamond tips or diamond-decorated tips are used in friction measurements. Diamond is an attractive material for an FFM tip because of its high mechanical strength and the belief that such tips are wear-resistant. However, diamond tips that were used to scratch diamond and silicon surfaces and then imaged showed significant wear that increased with the increasing hardness of the tested material [30.220, 221]. This wear altered the shape of the tip and hence influenced the contact area that is used to determine friction coefficients.

In summary, MD simulations provide insight into dry sliding friction and the sliding of metal tips across clean metal, crystalline ceramics, and layered ceramics surfaces. Stick–slip friction or wear can occur depending on the sliding conditions. The good lubricating properties of layered ceramics are observed in the simulations along with localized fluctuations in atomic-scale friction. Crystalline ceramics, such as diamond, exhibit relatively low friction and the simulations show how stick–slip atomic-scale motion changes with the conditions of sliding and the way in which the simulation is performed.

30.3.2 Decorated Surfaces

While dry sliding friction in vacuum assumes that ambient gas particles have no direct effect on the results, MD simulations show that free particles between two surfaces in sliding contact influence friction to a surprisingly large degree. These so-called third-body molecules have been studied extensively by *Perry* and *Harrison* [30.180, 222, 223] using MD simulations with bond-order and LJ potentials. These simulations focus on the effect of trapped small hydrocarbon molecules on the atomic-scale friction of two (111) crystal faces of diamond with hydrogen termination. These molecules might represent hydrocarbon contamination trapped between contacting surfaces prior to a sliding experiment in dry friction, or hydrocarbon debris formed during sliding.

In particular, the effects on friction of methane (CH_4), ethane (C_2H_6), and isobutane ($(CH_3)_3CH$ trapped between diamond (111) surfaces in sliding contact were examined in separate studies (Fig. 30.42). The frictional force for all these systems generally increases as the load increases, as illustrated in Fig. 30.43. The simulations predict that the third-body molecules markedly reduce the average frictional force compared to the results for pristine hydrogen-terminated surfaces. This is particularly true at high loads, where the third-body molecules act as a boundary layer between the two

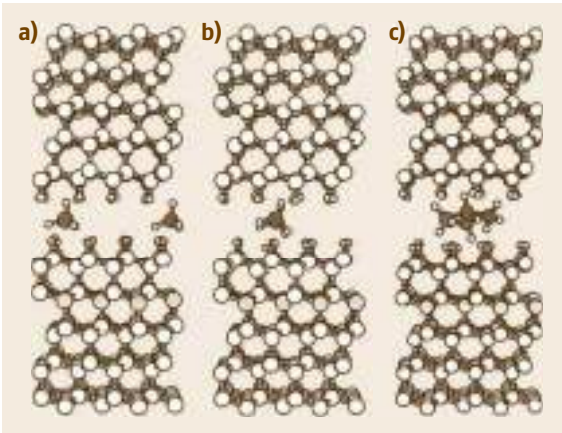


Fig. 30.42a–c Initial configuration at low load for the diamond plus third-body molecule systems. These systems are composed of two diamond surfaces, viewed along the [$\bar{1}$10] direction, and two methane molecules in **(a)**, one ethane molecule in **(b)**, and one isobutane molecule in **(c)**. Large *white* and *dark gray spheres* represent carbon atoms of the diamond surfaces and the third-body molecules, respectively. *Small gray spheres* represent hydrogen atoms of the lower diamond surface. Hydrogen atoms of the upper diamond surface and the third-body molecules are both represented by *small white spheres*. Sliding is achieved by moving the rigid layers of the upper surface from *left to right* in the figure (after [30.180], © ACS 1997)

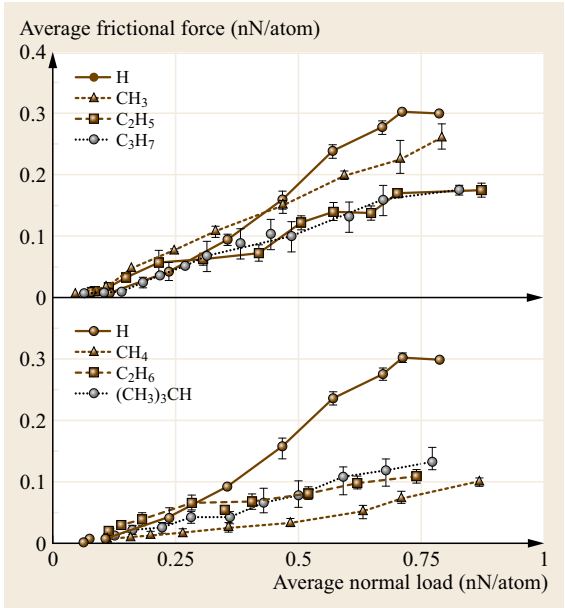


Fig. 30.43 Average frictional force per rigid-layer atom as a function of average normal load per rigid-layer atom for sliding the upper diamond surface in the [11$\bar{2}$] crystallographic direction. Data for the methane (CH_4) system (*brown triangles*), the ethane (C_2H_6) system (*brown squares*), the isobutane $(CH_3)_3CH$ system (*gray circles*), and diamond surfaces in the absence of third-body molecules (*brown circles*) are shown in the lower panel. Data for the methyl-terminated $-CH_3$ system (*brown triangles*), the ethyl-terminated ($-C_2H_5$) system (*brown squares*), the *n*-propyl-terminated ($-C_3H_7$) system (*gray circles*), and diamond surfaces in the absence of third-body molecules (*brown circles*) are shown in the *upper panel*. *Lines* have been drawn to aid the eye (after [30.180], © ACS 1997)

diamond surfaces. That is, the third-body molecules reduce the interaction of hydrogen atoms on opposing surfaces [30.223]. This is demonstrated by examining the vibrational energy excited in the diamond lattice during the sliding (Fig. 30.44). Significant vibrational excitation of the diamond outer layer (C−H) occurs in the absence of the methane molecules. Thus, the friction is ≈ 3.5 times larger when methane is not present. The application of load to the diamond surfaces causes the normal mode vibrations of the trapped methane molecules to change. Power spectra calculated from MD simulations [30.222, 223] show that even under low loads, the peaks in the power spectra are significantly broadened. Peaks in the low-energy region of the spectrum almost disappear with the additional application of load.

The size of the methane molecules allows them to be *pushed* in-between hydrogen atoms on the diamond surfaces while sliding [30.223]. However, steric considerations cause the larger ethane and isobutane molecules to change orientation during sliding. Conformations that lead to increased interactions with the diamond surfaces increase the average frictional force. Thus, despite the fact that the two diamond surfaces are farther apart when ethane and isobutane are present compared to when methane is present, the friction is larger because these molecules do not *fit* nicely into potential energy valleys between hydrogen atoms when sliding.

When similar hydrocarbon molecules (methyl, ethyl, and *n*-propyl groups) are chemisorbed to one of the sliding diamond surfaces, instead of trapped between the surfaces, different behavior is observed by *Harrison* et al. [30.174, 175, 209, 224]. Simulations show that methyl-termination does not decrease friction significantly but results in frictional forces that are nearly the same as they are for hydrogen-terminated diamond surfaces [30.213]. While the methane third-body molecules decrease the frictional force to a greater extent than the chemisorbed methyl groups, friction as a function of load is comparable for the ethyl-terminated and ethane systems, with the former giving

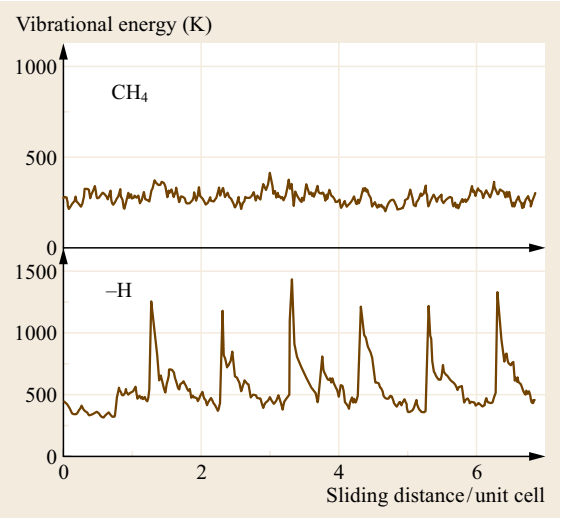


Fig. 30.44 Average vibrational energy between the (C−H) bonds of the upper diamond (111) surface versus sliding distance for hydrogen-terminated diamond (111) surfaces, with (CH_4) and without methane (H), trapped between them. The average normal load is approximately the same in both simulations and is in the range 0.8−0.85 nN/atom. The average frictional force on the upper surface is about 3.5 times smaller in the presence of the methane third-body molecules (after [30.223], © Elsevier 1996)

slightly higher frictional forces. Attaching the hydrocarbon groups to the diamond surfaces causes them to have less freedom to move between hydrogen atoms on opposing diamond surfaces during sliding. This generally increases their repulsive interaction with the diamond counterface.

MD simulations can also provide insight into the rich, nonequilibrium tribochemistry that occurs between surfaces in sliding contact. *Harrison* and *Brenner* examined the tribochemistry that occurs when ethane molecules are trapped between diamond surfaces in sliding contact, as illustrated in Fig. 30.45 [30.225]. This simulation was the first to show the atomic-scale mechanisms for the degradation of lubricant molecules due to friction. The type of debris formed during the sliding simulation is similar to the types of debris molecules that were observed in macroscopic experiments that examined the friction between diamond surfaces [30.226].

In the case of sliding metal surfaces, impurity molecules or atoms (both adsorbed and absorbed) on thin metal films can be expected to affect the film's properties. For example, calculations have shown that resistivity changes in the metal are strongly dependent on the nature of the adsorption bond [30.227]. When this result is used to interpret atomic-scale friction results obtained with the QCM, the sliding of adsorbate structures on metal surfaces are shown to be a combination of electron excitation and lattice vibrations. Additionally, other interesting quantum effects can come into play when the adsorbate is very different chemically from the surface on which it is sliding. For instance, the electronic frictional forces acting on small, inert atoms and molecules, such as C_2H_6 and Xe, sliding on metal surfaces have been calculated by *Persson* and *Volokitin* [30.228], where the metal surface was approximated by a electron gas (jellium) model. The calculations showed that the Pauli repulsive and attractive van der Waals forces are of similar magnitudes. In addition, the calculated electronic friction contributions agree well with the values derived from surface resistivity by *Grabhorn* et al. [30.229] and QCM measurements. These studies showed that parallel friction is mainly due to electronic effects while perpendicular friction is phononic in nature in this system.

In summary, MD simulations show that the average frictional force decreases significantly in systems with third-body molecules, especially at high loads. Simulations also provide information about the details of tribochemical interactions that can occur between lubricants and sliding surfaces. Additionally, the effect of the presence of small molecules on thin metal films can influence film properties, such as resistivity.

30.3.3 Thin Films

As discussed at the beginning of this section, the conversion of the work of sliding into some other less ordered form is responsible for friction at sliding solid interfaces. In the case of adhering systems, the work of sliding may be converted into damage within the bulk (plastic deformation), while in the case of weakly adhesive forces, friction can occur through the conversion of work to heat at the interface that causes no permanent damage to the surface (wearless friction). The latter case, when it is achieved through the presence of lubricating thin films, is the topic of this section.

There are several types of lubricating thin films, the simplest of which consist of small molecules that are analogous to wear debris that can *roll* between the sliding surfaces or that represent very short-chain bonded lubricants. These thin films were discussed in the previous subsection. The rest of this subsection will, therefore, focus on the effects of liquids, larger nanoparticles, self-assembled monolayers and solid thin films on lubrication and friction.

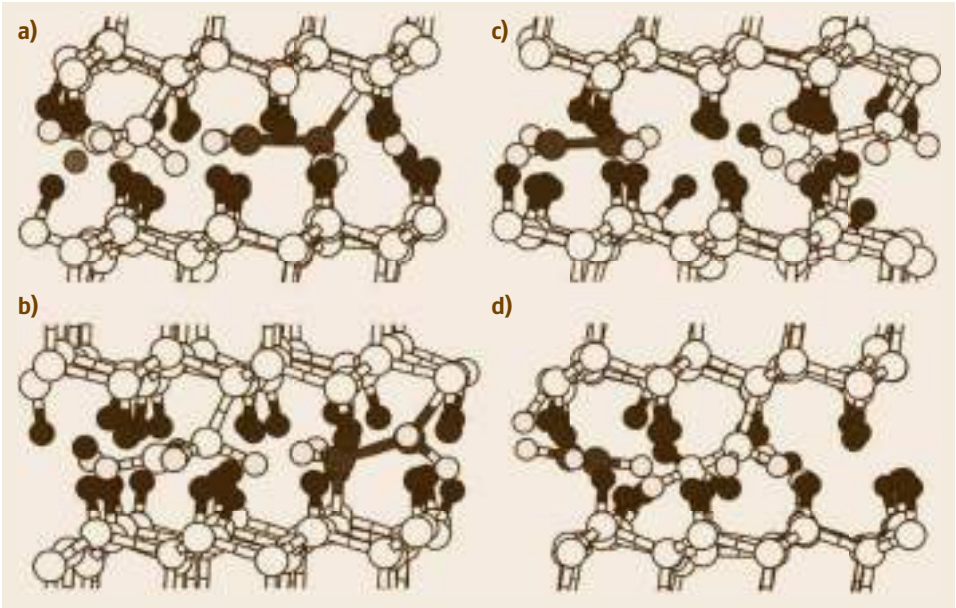


Fig. 30.45a–d Snapshots from a molecular dynamics simulation of the sliding of two hydrogen-terminated diamond (111) surfaces against one another in the [11$\bar{2}$] direction. The upper surface has two ethyl fragments chemisorbed to it. The simulation shows how sliding can induce chemistry at the interface. **(a)** Initial structure of the sliding surfaces with ethyl fragments chemisorbed to the top surface, **(b)** the ethyl fragments have started to react with atoms on the bottom surface, **(c)** continued sliding modifies the reactions, and **(d)** continued reactions are occurring between the ethyl fragments and the bottom surface (after [30.225], © ACS 1994)

Liquids

Liquids are common lubricants and so they have been studied in great depth at the macroscale. At the nanoscale, the tribological response of spherical liquid molecules has been well-characterized experimentally using the SFA and computationally with MD simulations by *Berman* et al. [30.230]. The SFA experiments considered one to three liquid layers and the stick–slip motion at the interface is found to increase in a quantized fashion as the number of lubricant layers decrease. When no external forces are applied to the system, the sliding stops and the solid–lubricant interactions are strong enough to force the liquid molecules to form a close-packed structure that is ordered. The transformation of the liquid into this solidlike structure causes the two surfaces to effectively bond to each other through the lubricant. When the surfaces start to slide again, lateral shear forces are introduced that steadily increase, which causes the molecules in the liquid to undergo small lateral displacements that change the film thickness. If these shear forces become greater than a critical value, the film disorders in a manner that is analogous to melting. This allows the surfaces to slide easily past each other in a manner that is still quantized. This sequence of events is nicely illustrated in Fig. 30.46 and can be reproduced multiple times for the same system.

Persson [30.231] used MD simulations with pairwise potentials similar to those in (30.6) to examine the mechanism by which this sharp transition occurs. They find that in the case of sliding on insulating crystal surfaces, the solid-state lubricant may be in a *superlubric* state where the friction is negligible. It is clear from the simulations, however, that any surface defects, even in low concentrations, will disrupt this state and transform the lubricant back into a fluid. In addition, when sliding occurs on metallic surfaces above cryogenic temperatures, the electronic contributions to friction are no longer zero and no superlubric state is possible.

High applied pressures can force the fluid molecules out from between the two confining surfaces [30.232]. The fact that liquid molecules close to a stiff surface are strongly layered in the direction perpendicular to the surface explains the experimental observation of a $(n \rightarrow n-1)$ layer transition, where n is number of monolayers, that is observed as the normal load increases [30.233]. Nucleation theory is used to calculate the critical pressure and determine the spreading dynamics of the $(n-1)$ island.

The reactivity of the liquid molecules are also critically important to boundary layer friction. MD studies by *Persson* [30.234] show that inert molecules interact

weakly with sliding surfaces. Consequently, as the rate of sliding increases, the molecular conversion from the solid state to the liquid state occurs in an abrupt manner. However, when the molecules interact strongly with the surfaces, they undergo a more gradual transition from the solid to the liquid state. *Persson* et al. [30.34] also considered systems where the molecules are attached to one of the surfaces, which causes the transitions to be abrupt. This is especially true if there are large separations between the chains.

While the studies discussed so far have focused on spherical liquids, most widely used liquid lubricants consist of long-chain hydrocarbons. Nonspherical liquid molecules have more difficulty aligning and solidifying. This is borne out in MD simulations by *Thompson* and *Robbins* [30.235] that show that spherical molecules have higher critical velocities than branched molecules. In particular, the simulations show that when the molecules are branched, the amount of time various parts of the system spend in the sticking and sliding modes changes with sliding rate. The critical velocity can also depend on the number of liquid layers in the film, the structure and relative orientation of the two sliding surfaces, the applied load and the stiffness of the surfaces.

Additional studies by *Landman* et al. [30.236] used MD simulations with bond-order and EAM potentials coupled with pair-wise potentials similar to (30.6) to study the sliding of two gold surfaces with pyramidal asperities that have straight chain $C_{16}H_{34}$ lubricant molecules trapped between them, as illustrated in Fig. 30.47. An important aspect of this study is that the sliding rate in the simulations is about 10 m/s, which is the same order of magnitude as the scanning speed in a computer disk. As the asperities approach each other, the hydrocarbon molecules begin to form layers. This is reflected in the oscillations in the frictional force shown in Fig. 30.48. When the asperities overlap in height and approach each other laterally, the pressure of the lubricant molecules increases to about 4 GPa which leads to the deformation of the gold asperities.

Glosli and *McClelland* [30.176] modeled the sliding of two ordered monolayers of alkane chains that are attached to two rigid substrates. This system is shown schematically in Fig. 30.49. The simulations predicted that energy dissipation occurs by a discontinuous plucking mechanism (sudden release of shear strain) or a viscous mechanism (continuous collisions of atoms of opposite films). The specific mechanism that occurs depends on the interfacial interaction strength. In particular, the *pluck* occurs when mechanical energy stored as strain is converted into thermal energy that leads to low friction forces at low temperatures. On the other hand, at higher temperatures some of the energy of sliding is dissipated through phonon excitations, which results in higher frictional forces. Interestingly, this trend reverses again at the highest temperatures considered when the molecules move so much that they slide easily over the surfaces, which decreases the frictional force. These results are summarized in Figs. 30.50 and 30.51.

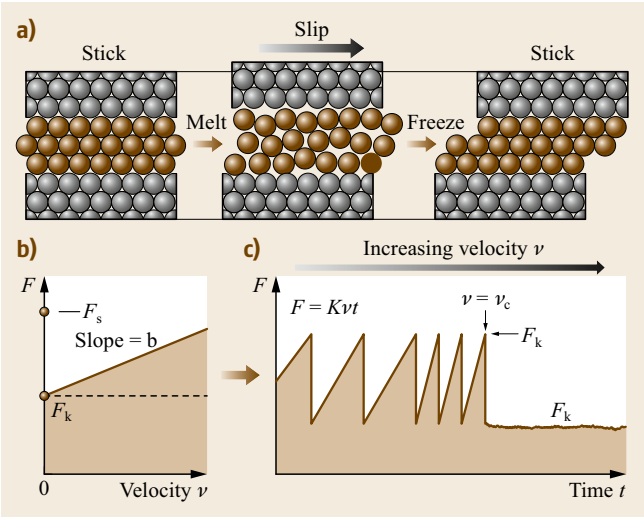


Fig. 30.46a–c The stick–slip transition that occurs for thin films of liquid between two sliding solid surfaces. F is the intrinsic friction and F_s is the friction where the liquid is in the rigid state; F_k is the friction where the liquid is in the liquidlike state. **(a)** Snapshots that indicate how the liquid melting and freezing is related to the stick–slip process. **(b)** Indicates the intrinsic friction relative to the friction values where the liquid is liquidlike and solidlike. **(c)** Illustrates how the intrinsic friction varies with time as the velocity increases (after [30.230], © ACS 1996)

Other studies of sliding surfaces with attached organic chains include MD simulations with LJ potentials by *Müser* and coworkers, [30.41, 237, 238] which considered friction between polymer *brushes* in sliding contact with one another. In particular, they considered the effect of sliding rate on the tilting of polymers and the effect of steady-state sliding versus nonsteady-state (*transient*) sliding. The simulations find that shear forces are lower for chains that tilt in a direction that is parallel to the shear direction. This tilting effect is significant for grafted polymers, as illustrated in Fig. 30.52, and less significant for absorbed polymers. This is due to the decrease in the differential frictional coefficient for the grafted polymers as well as the increase in the friction coefficient for absorbed polymers under shear. The tilting is also affected by the rate of sliding and is much larger at high sliding rates than small rates, as indicated in Fig. 30.52. The simulations further show that the inclination an-

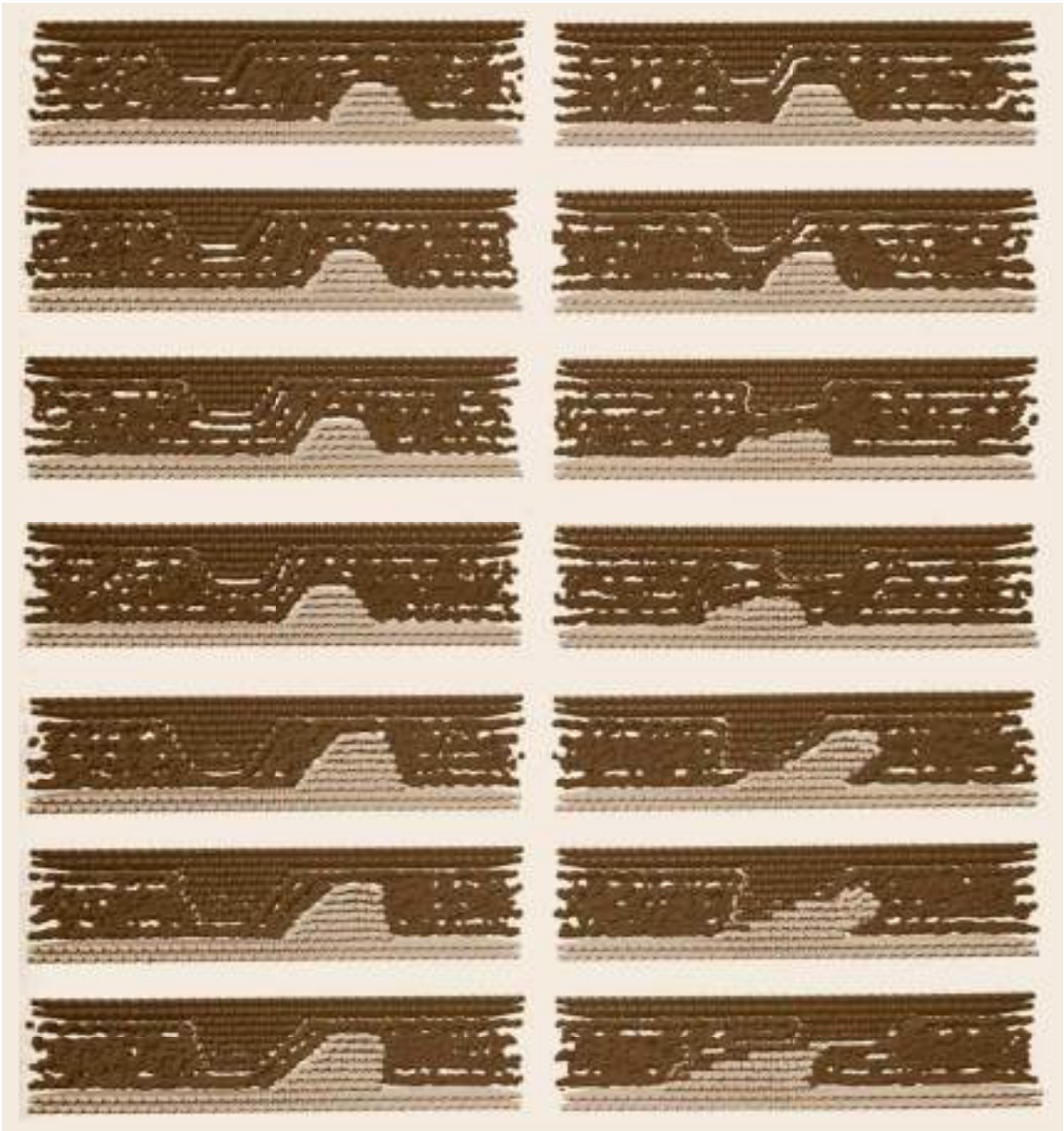


Fig. 30.47 Stills from a molecular dynamics simulation where Au(111) surfaces with surface roughness slide over one another while separated by hexadecane molecules. The scanning velocity is 10 m/s. Layering of the lubricant and asperity deformation occurs as the sliding continues. The *top three rows* show the results when the asperity heights are separated by 4.6 Å. The *bottom three rows* show the results when the asperity heights are separated by −6.7 Å (after [30.236], © ACS 1996)

gle of the chains decreases much more slowly than the shear stress, and the shear stress maximum is more pronounced if there is hysteresis in the chain orientations.

Typical friction loops for tips that are functionalized and sliding against surfaces that are functionalized in the same manner as illustrated in Fig. 30.25 are shown in Fig. 30.53. The friction force between the OH/OH pairs is significantly larger than the friction force between the CH_3/CH_3 pairs. This is due to the formation and breaking of hydrogen bonds during the shearing for the OH/OH pairs. The mean forces versus load forces

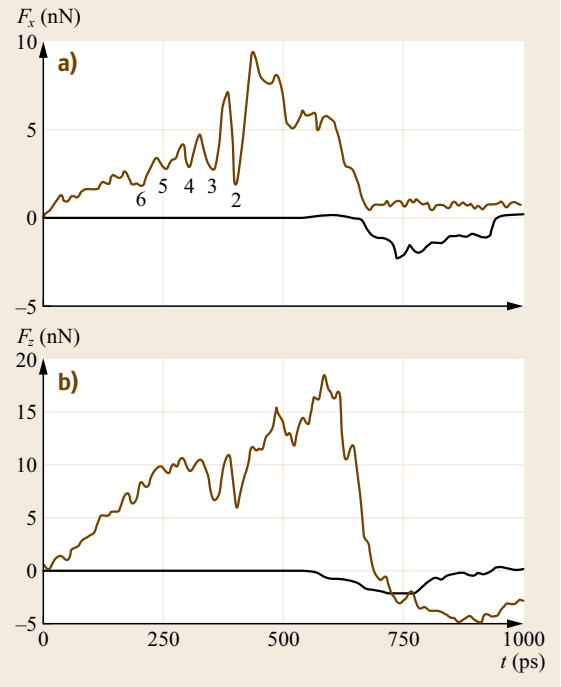


Fig. 30.48 **(a)** The lateral force (F_x) and **(b)** normal force (F_z) from the molecular dynamics simulations shown in Fig. 30.35 as a function of time. The forces between the two metal surfaces are shown by the *dashed line*. The force oscillations correspond to the structural changes of the lubricant in Fig. 30.35 (after [30.236], © ACS 1996)

for the OH/OH and CH_3/CH_3 pairs given in Fig. 30.54 are reduced by the tip radius.

MD simulations by *Manias* et al. considered the shearing of entangled oligomer chains that are attached to sliding surfaces, as illustrated in Fig. 30.55 [30.239]. They find that slip takes place within the film and that this occurs through changes in the chain conformations. Increased viscosity is predicted at the film–surface interface compared to the middle of the film, which results in a range of viscosities across the film as one moves away from the points of sliding contact.

To summarize this section, experiments and MD simulations show similar stick/slip transitions that occur for thin films of liquid between two sliding solid surfaces. Frictional properties are found to depend to a significant degree on molecular shape, whether the molecules are grafted on the surfaces or merely absorbed on them, and the degree of tilting in the case of molecular chains. In the case of long-chain molecules, temperature is found to affect the frictional force because the mechanical energy stored in long-chains can be converted into thermal energy by friction.

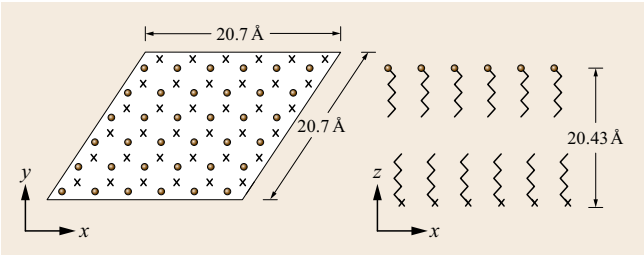


Fig. 30.49 Top and side views of the alkane chains attached to surfaces that are sliding against each other (after [30.176], © APS 1993)

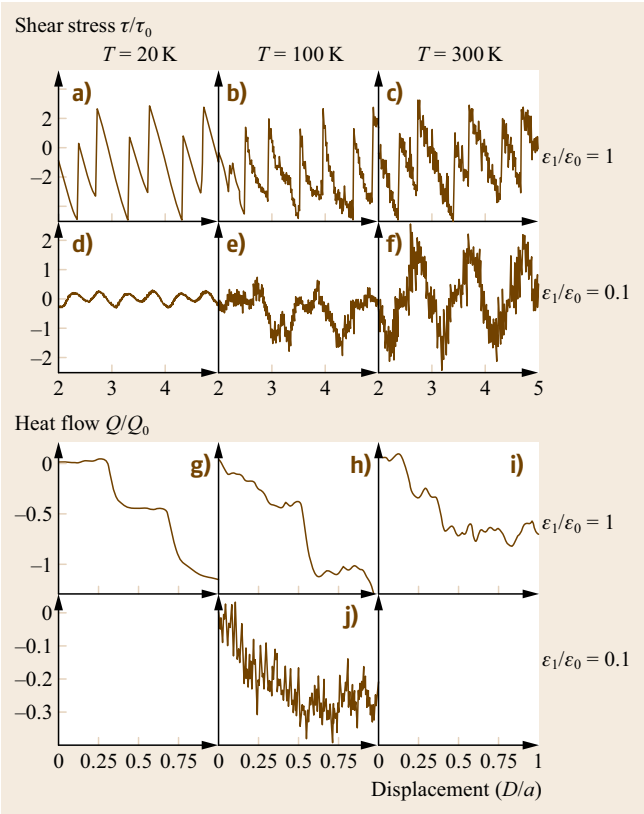


Fig. 30.50a–j Data from molecular dynamics simulations of the sliding of the surfaces shown in Fig. 30.32. **(a–f)** The shear stress and **(g–j)** the heat flow as a function of sliding for normal and reduced interfacial strengths. The *plots* show how the calculated values change with system temperature (after [30.176], © APS 1993)

Self-Assembled and Polymer Thin-Film Structures

There have been numerous experimental studies of friction on SAMs on solid surfaces with AFM and FFM. The experimental results reveal relationships among elastic compliance, topography and friction on thin

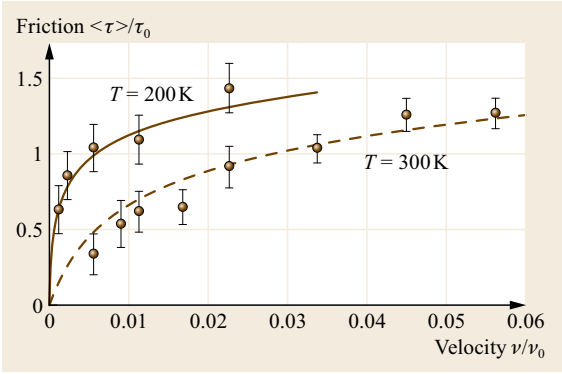


Fig. 30.51 A plot of calculated values of the average interfacial shear stress as a function of the velocity of sliding of the two surfaces shown in Fig. 30.33 (after [30.176], © APS 1993)

L–B films [30.240]. For example, they have detected differences in the adhesive interactions between the microscope tips and CH_3 and CF_3 end groups [30.109]. Fluorocarbon domains generally exhibit higher friction than the hydrocarbon films, which the authors attribute to the lower elasticity modulus of the fluorocarbon films that results in a larger contact area between the tip and the sample [30.240–242]. *Perry* and coworkers examined the friction of alkanethiols terminated with $-CF_3$ and $-CH_3$ [30.243]. The lattice constants for both films are similar and the films are well-ordered. The friction of the SAMs with chains that are terminated with fluorine end groups is larger than the friction of the SAMs with chains that are terminated with hydrogen end groups. However, the pull-off force is similar in both systems, which implies that these end groups have similar contact areas. The authors speculate that the larger $-CF_3$ groups interact more strongly with adjacent chains than the $-CH_3$-terminated chains. Therefore, the fluorinated chains have more modes of energy dissipation within the plane of the monolayer and, thus, have larger friction.

Molecular disorder of the alkyl chains at the surface can also affect the frictional properties of self-assembled films if the layers are not packed too closely together [30.244]. Indentation can induce disorder in the chains that then compress as the tip continues to press against them. If the tip presses hard enough, the film hardens as a result of the repulsive forces between the chains. However, if the chains are tilted, they bend or deform when the tip pushes on them in a mostly elastic fashion that produces long lubrication lifetimes. At low contact loads of about 10^{-8} N, wear usually occurs at defect sites, such as steps. Wear can also occur if there are strong adhesive forces between the film and the surface [30.245].

The friction of model SAMs composed of alkane chains was examined using MD simulations with bond-order and LJ potentials by *Mikulski* and *Harrison* [30.246, 247]. These simulations show that periodicities observed in a number of system quantities are the result of the synchronized motion of the chains when they are in sliding contact with the diamond counterface. The tight packing of the monolayer and commensurability of the counterface are both needed to achieve synchronized motion when sliding in the direction of chain tilt. The tightly packed monolayer is composed of alkane chains attached to diamond (111) in the (2×2) arrangement and the loosely packed system has $\approx 30\%$ fewer chains. The average friction at low loads is similar in both the tightly and loosely packed systems at low loads. Increasing the load, however, causes the tightly packed monolayer to have significantly lower friction than the loosely packed monolayer (Fig. 30.56). While the movement of chains is somewhat restricted in both systems, the tightly packed monolayer under high loads is more constrained with respect to the movement of individual chains than the loosely packed monolayer,

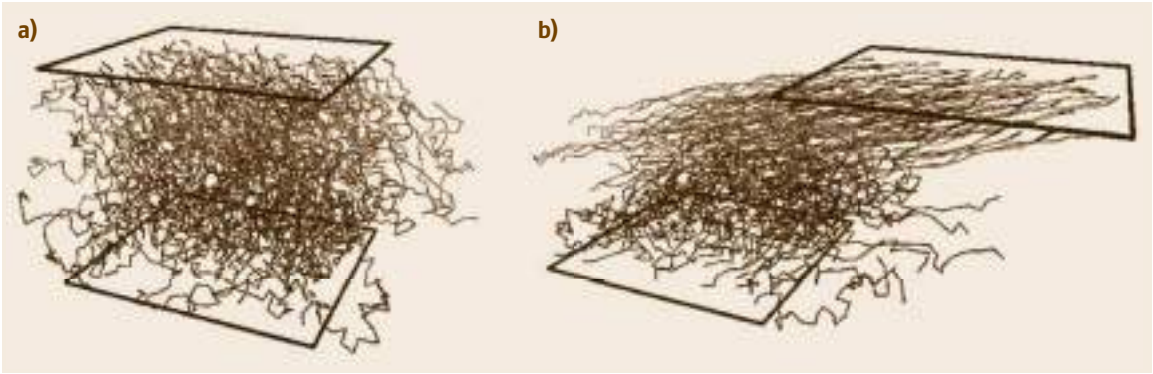


Fig. 30.52a,b Snapshots of sliding walls with attached polymers in a solvent. **(a)** illustrates the sliding process at low sliding rates while **(b)** illustrates the sliding process at high sliding rates (after [30.41], © CCLR 2002)

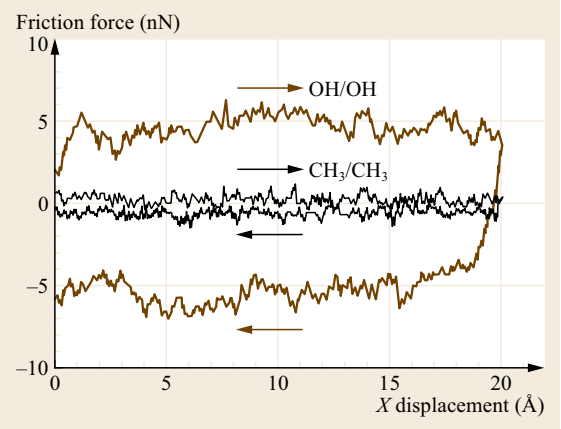


Fig. 30.53 Typical friction loops for the systems shown in Fig. 30.25 for CH_3/CH_3 and OH/OH pairs under a contact load of 0.2 nN (after [30.163], © ACS 2002)

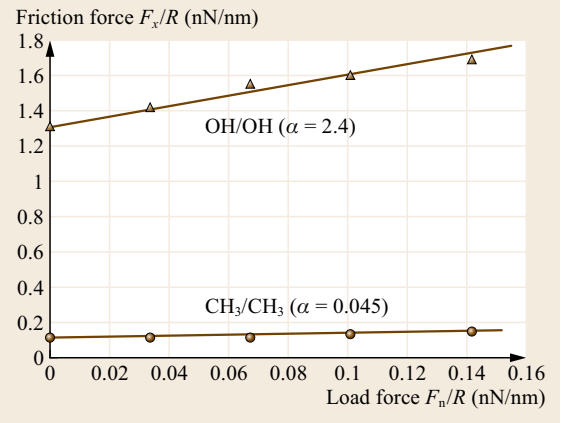


Fig. 30.54 Friction force versus contact load from the systems shown in Fig. 30.25 for CH_3/CH_3 and OH/OH (after [30.163], © ACS 2002)

as illustrated in Fig. 30.57. Therefore, sliding initiates larger bond-length fluctuations in the loosely packed system, which ultimately lead to more energy dissipation via vibration and, thus, higher friction. Thus, the efficient packing of the chains is responsible for the lower friction observed for tightly packed monolayers under high loads.

Several AFM experiments have examined the friction of SAMs composed of chains of mixed lengths [30.248]. For example, the friction of SAMs composed of spiroalkanedithiols was examined by *Perry* and coworkers [30.249]. The effects of crystalline order at the sliding interface were examined by systematically shortening some of the chains. The resulting increase in disorder at the sliding interface causes an increase in friction.

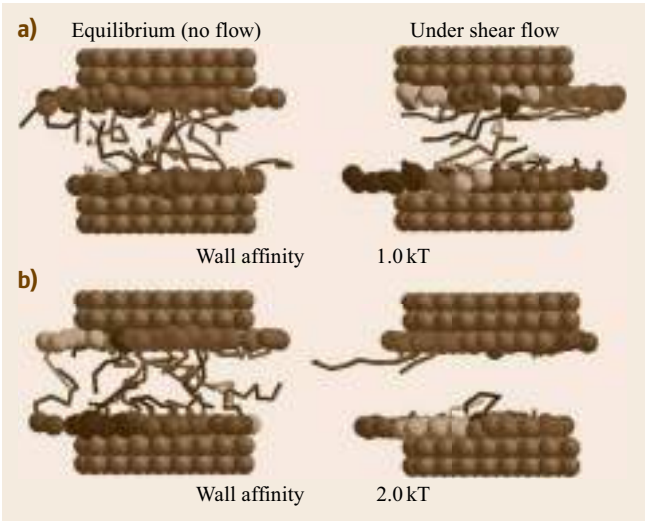


Fig. 30.55a,b Changes in the conformation of adsorbed hydrocarbon chains on weakly **(a)** and strongly **(b)** physisorbing surfaces at equilibrium and under shear (after [30.239], © ACS 1996)

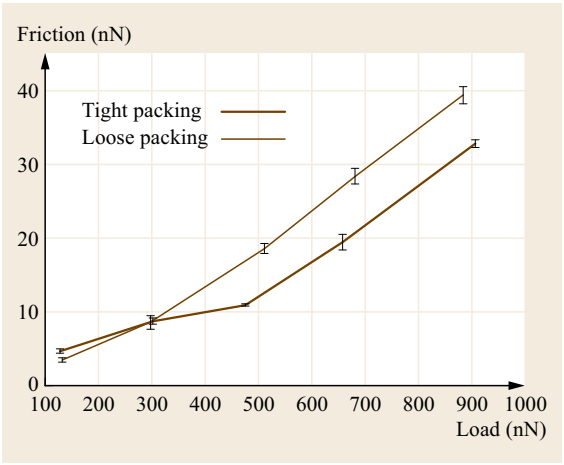


Fig. 30.56 Friction as a function of load when a hydrogen-terminated counterface is in sliding contact with C_{18} alkane monolayers (after [30.247], © ACS 2001)

The link between friction and disorder in monolayers composed of *n*-alkane chains was recently examined using MD simulations by *Harrison* and coworkers [30.250]. The tribological behavior of monolayers of 14 carbon atom-containing alkane chains, or pure monolayers, was compared to monolayers that randomly combine equal amounts of 12 and 16 carbon-atom chains, or mixed monolayers. Pure monolayers consistently show lower friction than mixed monolayers when sliding under repulsive (positive) loads in the direction of chain tilt. These MD simulations reproduce trends observed in AFM experiments of mixed-

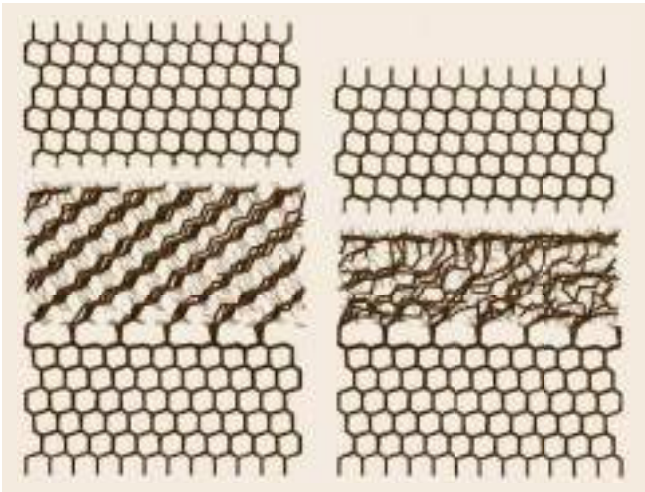


Fig. 30.57 Snapshots of tightly packed C_{18} alkane monolayers on the *left*, and loosely packed monolayers on the *right* under a load of about 500 nN. The chains in both systems are arranged in a (2×2) arrangement on diamond (111). The loosely packed system has 30% of the chains randomly removed. The sliding direction is from *left to right* (after [30.247], © ACS 2001)

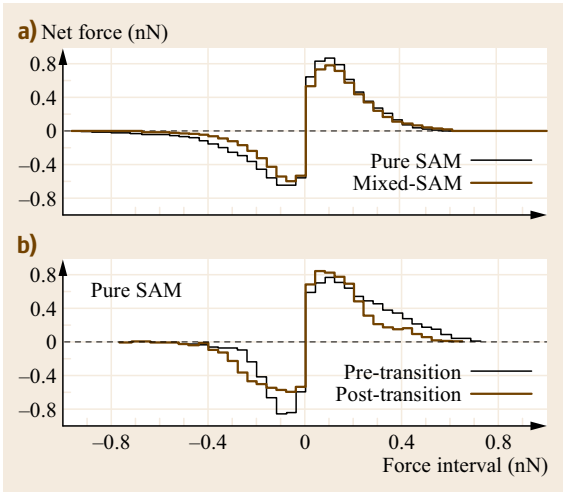


Fig. 30.58a,b The distribution of contact forces along the sliding direction (friction force). In **(a)**, the forces for the mixed and pure system sliding in the direction of chain tilt are shown. The forces for the pure system sliding in the transverse direction to the chain tilt are shown in **(b)**. Positive force intervals correspond to chain groups that resist tip motion while negative intervals correspond to chain groups that *push* the tip in the sliding direction. Forces from four runs with independent starting configurations are binned for all sets of data

length alkanethiols [30.248] and spiroalkanedithiols on Au [30.251]. *Harrison* and coworkers [30.252] have also examined the *odd-even* effect noted in experiment [30.253], where friction is found to be larger for SAMs differing by one methylene group. The MD simulations demonstrated that the effect was due to conformational differences in the chains of different length and became more pronounced at higher loads.

Because the force on individual atoms is known as a function of time in MD simulations, it is possible to calculate the contact forces between individual monolayer chain groups and the tip, where contact force is defined as the force between the tip and a $-CH_3$ or a $-CH_2$-group in the alkane chains. The distribution of contact forces between individual monolayer chain groups and the tip are shown in Fig. 30.58. It is clear from these contact force data that the magnitude, or scale of the forces, is similar in both the pure and the mixed monolayers. In addition, it is also apparent that the pure and mixed monolayers resist tip motion in the same way. That is, the shape of the histograms in the positive force intervals is similar. In contrast, the contact forces *pushing* the tip along differ in the two monolayers. The pure monolayers exhibit a high level of symmetry between resisting and pushing forces. Because the net friction is the sum of the resisting and pushing forces, the symmetry in these distributions of the pure monolayers results in a lower net friction than the mixed monolayers. Thus, the ordered, densely packed nature of the pure monolayers allows the energy stored when the monolayer is resisting tip motion (positive forces) to be regained efficiently when the monolayer *pushes* on the tip (negative forces). The distribution of negative contact forces in the mixed monolayers is different from the distribution of the positive forces. For this reason, mechanical energy is not efficiently channeled back into the mixed monolayer as the tip passes over the chains and, as a result, the friction is higher. The range of motion of the chains is monitored by computing the deviation in a chain group's position compared to its starting position, as illustrated in Fig. 30.59. It is clear from analyzing these data that the increased range of motion is linked to large contact forces. The increased range of motion of the protruding tails in the mixed system prevents the efficient recovery of energy during sliding (negative contact force distribution) and allows for the dissipation of energy.

The pure monolayers exhibit marked friction anisotropy. The contact force distribution changes dramatically as a result of the change in sliding direction, resulting in an increase in friction (Fig. 30.58). Sliding in the direction perpendicular to chain tilt can cause both types of monolayers to transition to a state where the chains are primarily tilted along the sliding direction. This transition is accompanied by a large change

in the distribution of contact forces and a reduction in friction.

Recently, the response of monolayers composed of alkyne chains, which contain diacetylene moieties, to compression and shear [30.254] was examined using MD simulations. These are the only simulations to date that show that compression and shear can result in cross-linking, or polymerization, between chains. The vertical positioning of the diacetylene moieties within the alkyne chains (spacer length) and the sliding direction both have an influence on the pattern of cross-linking and friction. Compression and shear cause irregular polymerization patterns to be formed among the carbon backbones, as illustrated in Fig. 30.60. When diacetylene moieties are located at the ends of the chains closest to the tip, chemical reactions between the chains of the monolayer and the amorphous carbon tip occur causing the friction to increase 100 times, as indicated in Fig. 30.61. The friction between the amorphous carbon tip and all of the diacetylene-containing chains is larger than the friction between a hydrogen-terminated diamond counterface and tightly packed monolayers composed of *n*-alkane chains [30.247]. This is attributed to the disorder at the interface caused by the irregular counterface.

Zhang and *Jiang* [30.255] used MD simulations to study the effect of confined water between alkyl monolayers terminated with $-CH_3$ (hydrophobic) and $-OH$ (hydrophilic) groups on Si(111), as illustrated in Fig. 30.62. For the hydrophobic molecules, the friction coefficient is almost constant independent of the number of water molecules. For the hydrophilic molecules, the friction coefficient decreases rapidly with an increase in the number of water molecules, as shown in Fig. 30.63. These results are in good agreement with surface force microscopy (SFM) experimental results. *Zhang* et al. [30.256] also studied the friction of alkanethiol SAMs on gold using hybrid molecular simulations at the same time scales as are used in AFM and FFM experiments. Various quantities were varied in the simulations, including chain length, terminal group, scan direction and scan velocity. The simulations showed that the frictional force decreases as the chain length increases and is smallest when scanned along the tilt direction. They also predicted a maximum friction coefficient for hydrophobic $-CH_3$-terminated SAMs and low friction coefficients for hydrophilic $-OH$-terminated SAMs as the scan velocity increases. The simulations further predicted a saturated constant value at high scan rates for both surfaces. These results are summarized in Figs. 30.64 and 30.65.

The work of *Chandross* et al. [30.257, 258] illustrates the effects of chain length on friction and stick–slip behavior between two ordered SAMs consisting of

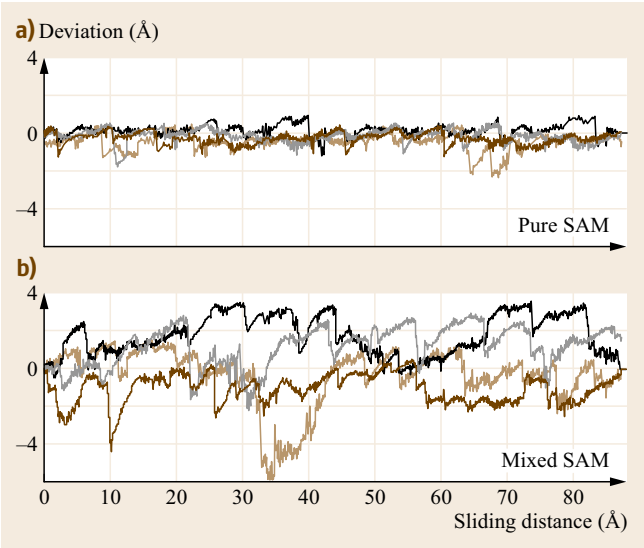


Fig. 30.59a,b Trajectories of individual chain groups that generate the largest contact forces when sliding in the direction of chain tilt for both the pure **(a)** and mixed monolayer systems **(b)**. The deviation is defined as the change in position along the sliding direction relative to the chain group's starting position. (The positions are averaged over 2000 simulation steps)

alkylsilane chains over a range of shear rates at various separation distances or pressure, as illustrated in Fig. 30.66. The adhesion forces between the two SAMs at the same separation distance decrease as the chain length increases from 6 to 18 carbon atoms. However, the friction forces are independent of the chain length and the shear velocity. The system size is shown to have an effect on the sharpness of the slip transitions but not on the dynamical events, as shown in Fig. 30.67. In a later paper, *Chandross* et al. used SiO_2 tips with radius of curvatures in the range of 3–30 nm to interact with fully physisorbed, fully chemisorbed, and a mixture of chemisorbed and physisorbed alkylsilane SAMs on amorphous Si [30.259]. This tip-based geometry allows for the exploration of actual contact area as a function of load, which was found to be proportional to the square root of the load.

SAMs have been very successful in the lubrication of surfaces that infrequently come into sliding or normal contact. Their inability to lubricate a reciprocating contact is due, in part, to the inability to replenish the coating in situ. One proposed solution is to use a chemically bound SAM in conjunction with a physisorbed mobile molecule in a *bound + mobile* lubricant scheme, which is similar to the lubricants used to mitigate head crashes in hard drives.

Irving and *Brenner* used molecular dynamics to study the interfacial structure, self-diffusion, and abil-

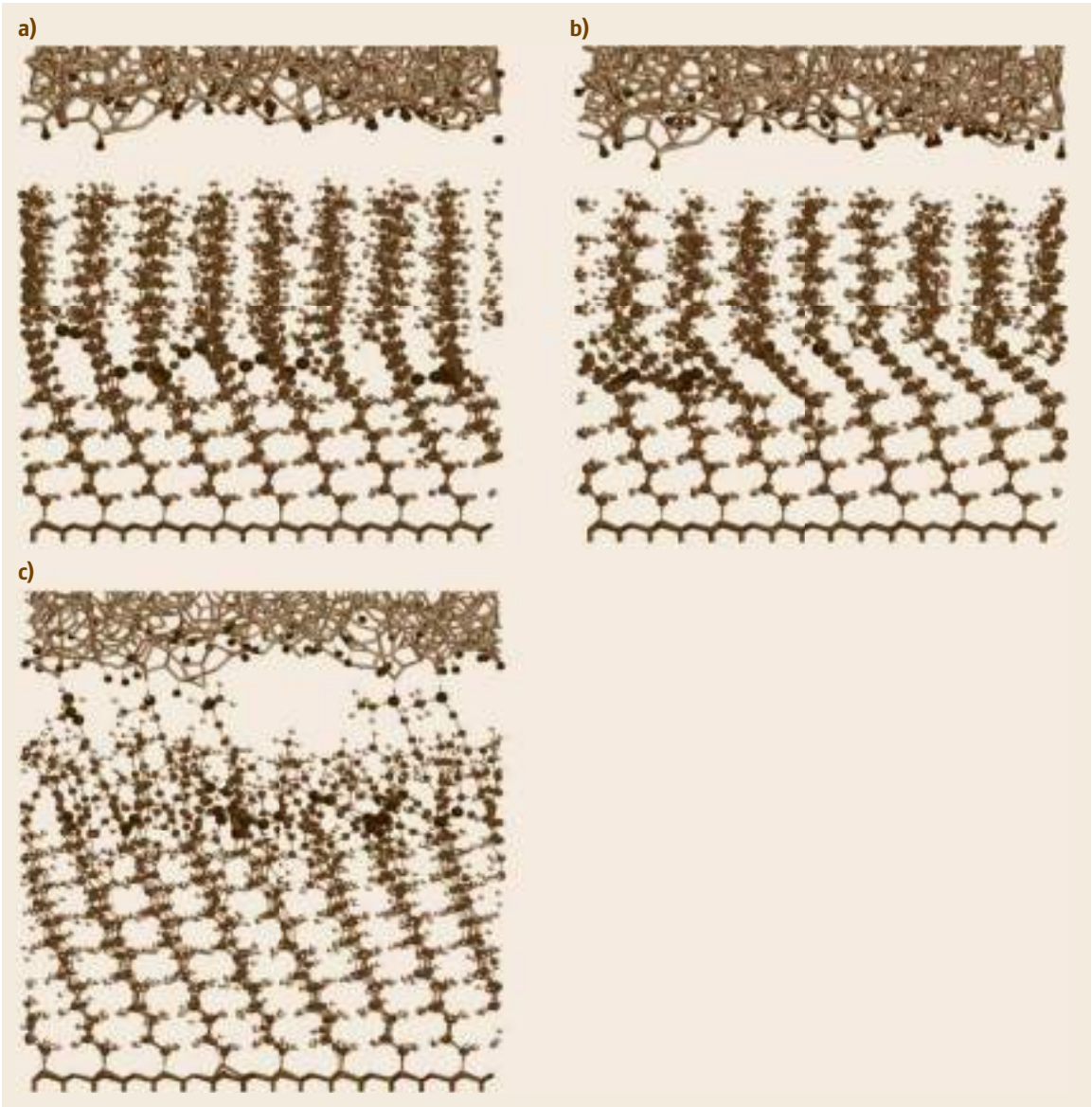


Fig. 30.60 **(a)** Perpendicular-chain, **(b)** tilted-chain, and **(c)** end-chain monolayer systems after compression to 200 nN and pull-back of the hydrogen-terminated tip. *Large, dark spheres* in the hydrocarbon monolayers represent cross-linked atoms with sp^2 hybridization. *Dark, small spheres* represent hydrogen atoms that are initially on the hydrogen-terminated amorphous carbon tip (after [30.254], © ACS 2004)

ity of the mobile phase to incorporate into defected sites [30.260]. A potential bound + mobile lubricant combination of a chemically bound octadecyltrichlorosilane (ODTS) SAM together with mobile tricresyl phosphate (TCP) molecules was studied. The simulations showed that the TCP did not incorporate into the interior of the close packed defect free ODTS SAM. The TCP molecules on the surface of the SAM were also not tightly bound to a particular surface site but instead were found to readily diffuse across the surface in a random walk fashion. An estimated diffusion barrier of 0.0937 eV with an Arrhenius prefactor of $26.47 \times 10^{-4}\,cm^2/s$ was calculated for single-molecule diffusion. It was also found that the TCP molecules would only localize in the vicinity of methylene groups ($-CH_2-$) along the backbone of the ODTS chain, which were exposed when a cylindrical defect was created in the SAM. To get to these local-

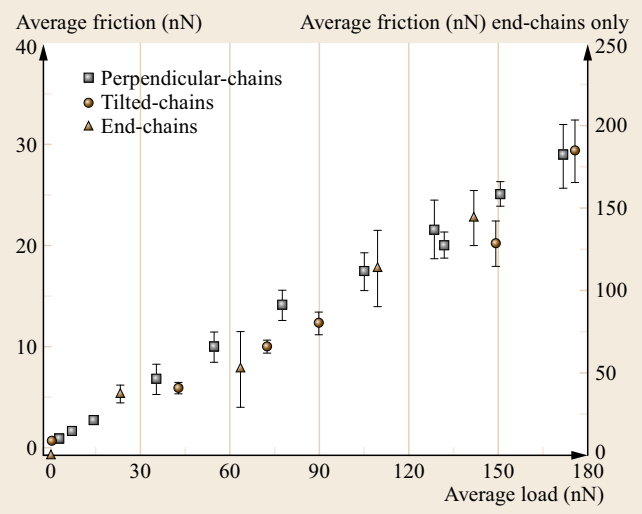


Fig. 30.61 Average friction on the tip as a function of load for the monolayer systems shown in Fig. 30.60. The *scale* for the average friction in the end-chain system is shown on the *right-hand side* of the figure (after [30.254], © ACS 2004)

izing sites, however, first required the TCP molecules to overcome an anisotropic energy barrier for inclusion into the cylindrical defect. This anisotropic barrier was found to depend on the direction of ODTS chain tilt and the direction the TCP molecules approached the cylindrical defect in the SAM. A later study used the diffusion information and multiscale methods to analyze the conditions under which this scheme would be successful [30.261].

Polymer thin films are also a widely studied for their lubricating properties. An example is polytetrafluoroethylene (PTFE), which has been used in a wide range of applications from satellites to frying pans. In a joint computational and experimental work *Jang* et al. examined the molecular origins of friction using classical molecular dynamics as well as an AFM and microtribometer [30.262]. The simulations predicted an anisotropic behavior of the friction coefficient depending on whether the sliding direction was parallel or perpendicular to the PTFE chains lying on the surface. Sliding directions parallel produced lower friction coefficients and wear while sliding perpendicular to alignment produced higher coefficients and wear. The microtribometer results were in agreement with these findings. Also of interest in the AFM work was that transfer films were always produced parallel to the sliding direction.

Similar experimental findings for anisotropic tribological behavior of polyethylene (PE) in the literature motivated *Heo* et al. to examine this system using classical molecular dynamics to study crystalline PE interfaces [30.263]. Like the findings for PTFE it was found that friction and wear had an anisotropic behavior that depended on molecular orientation and sliding direction. Unlike the findings for PTFE, the PE system exhibited a stick–slip motion as the interfaces passed by one another. The reason for the differences between the two systems was attributed to increased bond scission seen in PTFE as compared to PE under sliding conditions. This scission allows collections of molecules in PTFE to move at the interface, which does not occur as readily in the PE system.

In short, atomic-scale simulations show the relationship between elastic properties, degree of molecular disorder and friction of self-assembled thin films that illuminates the origin of the properties that are measured experimentally.

Nanoparticles

Nanoparticles are being considered for a wide variety of applications, including as fillers for nanocomposite materials, novel catalysts or catalytic supports, and components for nanometer-scale electronic devices [30.264]. They have also generated considerable interest as possible new lubricant materials that have the potential to function as *nanoballbearings* with exceptionally low friction coefficients. The nanoparticles of most interest for tribological applications include C_{60} [30.265–277], carbon nanotubes [30.278–285], and MoO_3 nanoparticles [30.286, 287], among others [30.288].

The experiments report wide variations in frictional coefficients (for instance, values of 0.06 to 0.9 have been measured for C_{60}) that may be caused by differences in the experimental methods used, the thickness of the nanoparticle layer or island, the atmosphere (argon versus air, levels of humidity) used, and the transfer of nanoparticles to the FFM tips. As a result, there is much that remains to be clarified about the tribological behavior of nanoparticle films.

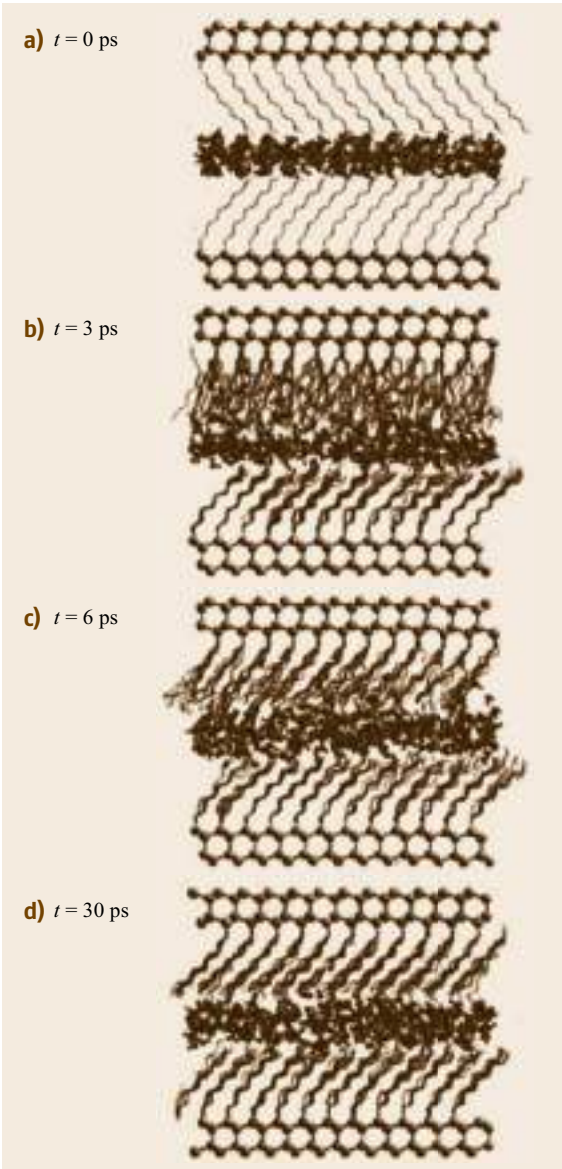


Fig. 30.62a–d Snapshots of hydrophilic monolayers and confined water molecules from MD simulations at 300 K. The tilt direction of monolayers on the top plate changed after $t = 10.0$ ps. **(a–d)** illustrates how the tilting of the monolayers change as a function of time, and the way in which the water becomes increasingly less confined (after [30.255], © AIP 2005)

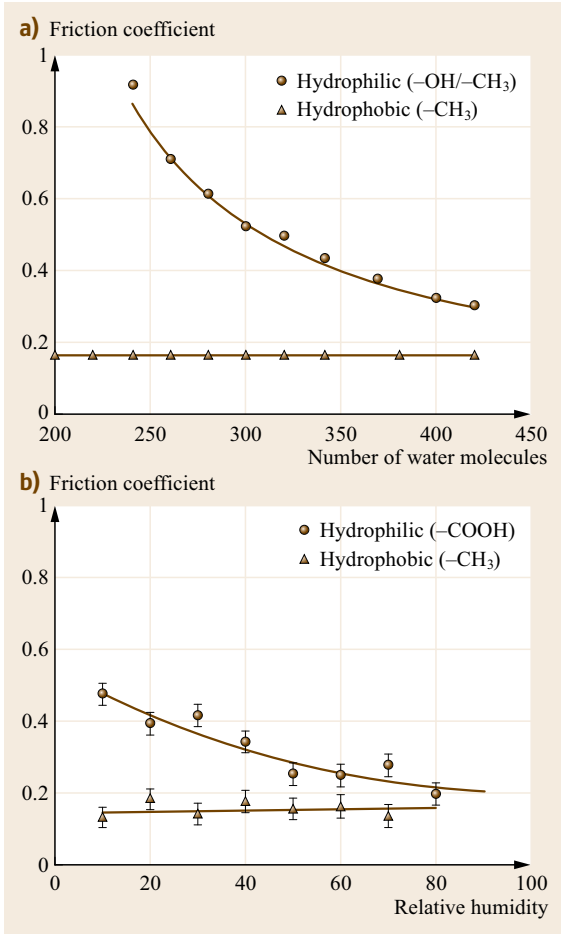


Fig. 30.63 (a) Friction coefficients for hydrophobic ($-CH_3$) and hydrophilic (50% mixed $-CH_3/-OH$) monolayers as a function of water molecules from MD simulations at 300 K ($H = 6.0$ Å), and **(b)** scanning force microscopy measurements of frictional forces of difference surfaces under various relative humidities (after [30.255], © AIP 2005)

In the case of C_{60}, the mechanistic response to applied shear forces has not been definitively determined. For example, some experimental studies show evidence of C_{60} molecules rolling against the substrate, each other, or the sliding surfaces [30.265, 270, 272, 275, 277] while others hypothesize that the low friction of C_{60} films is due in part to blunting of the tip by transfer of fullerene molecules to the tip apex. Fullerene films are found experimentally to have dissipation energies and shear strengths that are a full order of magnitude lower than the values that are typical for boundary lubricants [30.289]. Experimental testing of the frictional properties of fullerenes reveal low mechanical stability accompanied by progressive wear and transfer of fullerene materials when they are only physisorbed on a solid surface [30.290]. Furthermore, measurements with a FFM show that under certain conditions, adsorbed fullerene films deteriorate

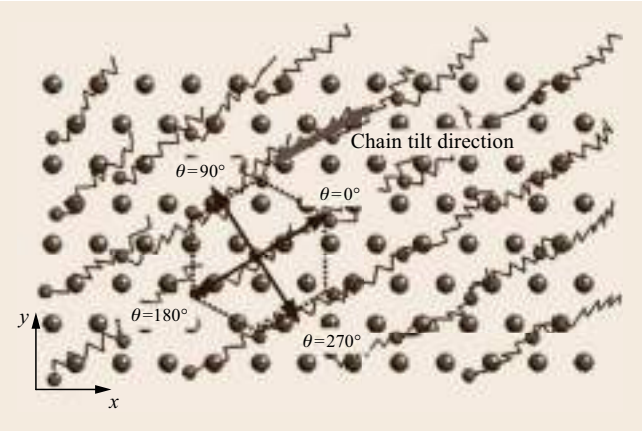


Fig. 30.64 Schematic illustration of the chain tilt and scan directions on alkanethiol SAMs/Au(111) in hybrid molecular simulations; θ is the angle between the tip moving direction and the chain tilt direction. The *larger spheres* represent substrate Au atoms, smaller spheres sulfur atoms in molecular chains, and *zigzag lines* molecular chains (after [30.256], © ACS 2003)

at pressures as low as about 0.1 GPa [30.291]. The challenge is therefore to obtain mechanically stable, ordered molecular films of fullerenes firmly attached to a solid substrate.

There have been several MD simulation studies to investigate the tribological properties of fullerenes. A representative study by *Legoas* et al. [30.292] investigated the experimentally observed low-friction system of C_{60} molecules positioned on highly oriented pyrolytic graphite. The results show that decreasing the van der Waals interaction between a C_{60} monolayer and graphite sheets, and the characteristic movements of graphite flakes over C_{60} monolayers, explains the measured ultralow friction of C_{60} molecules and graphite sheets.

Several MD simulation studies have also been carried out on the tribological properties of carbon nanotubes. For example, simulations by *Buldum* and *Lu* [30.278] and *Schall* and *Brenner* [30.281] indicate that single-wall carbon nanotubes roll when their honeycomb lattice is *in registry* with the honeycomb lattice of the graphite. If this registry is not present, the carbon nanotubes respond to applied forces from an AFM by sliding. This behavior is nicely summarized in Fig. 30.68. These MD simulation findings were simultaneously confirmed in experimental studies by *Falvo* et al. [30.280]. Experimental studies of multiwall carbon nanotubes on graphite [30.284] show similar evidence of nanotube rolling when the outer tube is pushed.

The tribological properties of nanotube bundles are important, as it is well-known that carbon nanotubes agglomerate together very readily to form bundles and are often grown in bundle form [30.264]. An experimental study by *Miura* et al. [30.285] of carbon nanotube bundles being pushed around on a KCl surface with an AFM tip indicates that bundles of single-wall carbon

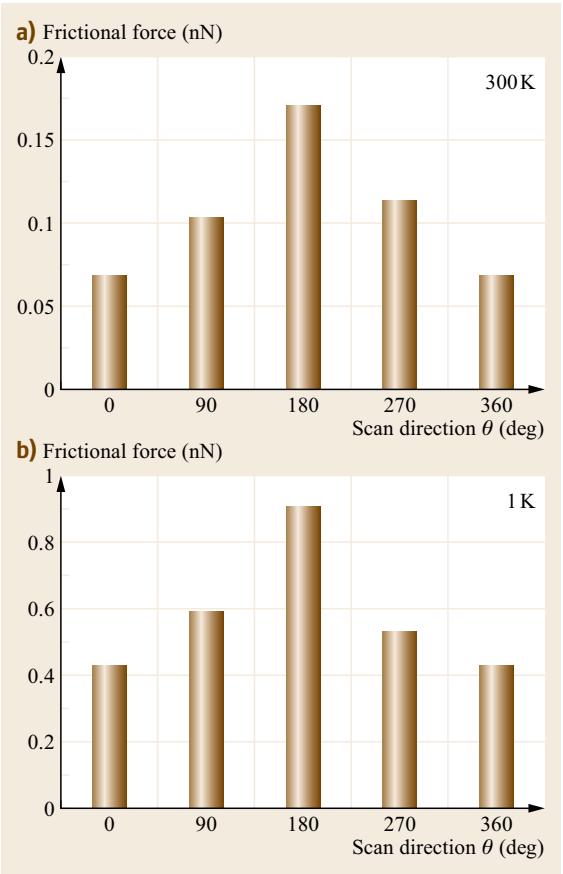


Fig. 30.65a,b Frictional force as a function of scan direction from hybrid simulations for $C_{11}CH_3$ SAMs on Au(111) at **(a)** 300 K and **(b)** 1.0 K. Frictional force is the smallest when scanned along the tilt direction, the largest when scanned against the tilt direction, and between when scanned perpendicular to the tilt direction at both temperatures (after [30.256], © ACS 2003)

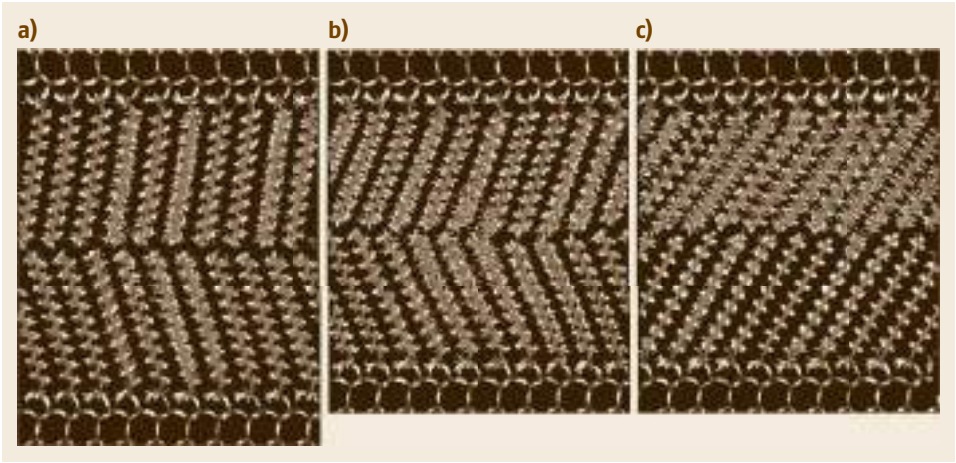


Fig. 30.66a–c Wireframe images of $n = 18$ SAMs at fixed separations of **(a)** $d = -5.2$ Å (low pressure, under compression only) **(b)** $d = -10.2$ Å (high pressure, under compression only) and **(c)** $d = -10.2$ Å (high pressure, under shear) (after [30.258], © ACS 2005)

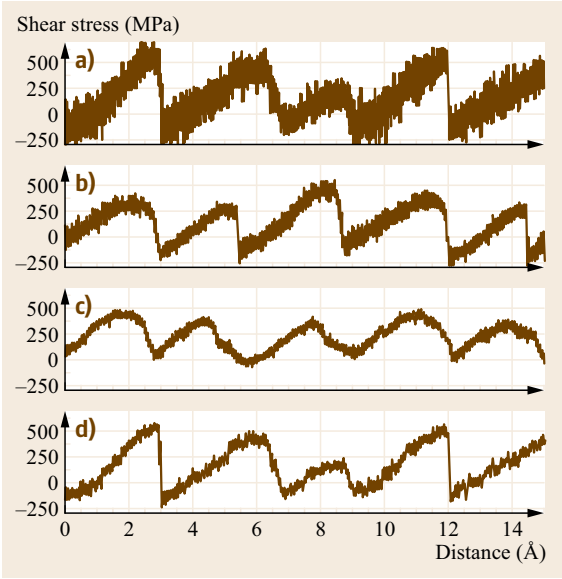


Fig. 30.67a–d Shear stress σ_s as a function of system size for $n = 6$ SAMs corresponding to a pressure of 200 MPa at $v = 1.0$ m/s: **(a)** 100 chains per surface, **(b)** 400 chains per surface, **(c)** 1600 chains per surface, and **(d)** 16 point box average of system with 100 chains per surface (after [30.258], © ACS 2002)

nanotubes can be induced to roll in a manner that is similar to the rolling observed for multiwall nanotubes.

MD simulations by *Ni* and *Sinnott* [30.282, 283] considered the responses of horizontally and vertically aligned single-wall carbon nanotubes between two hydrogen-terminated diamond surfaces, where the top surface is slid relative to the bottom surface. The movement of the carbon nanotubes in response to the shear forces was predicted to be simple sliding for both orientations. Interestingly, the simulations do not predict rolling of the horizontally arranged carbon nanotubes even when they are aligned with each other in two-layer and three-layer structures. Instead, at low compressive forces, illustrated in Fig. 30.69, the nanotube bundles slide as a single unit, and at high compressive forces, also illustrated in Fig. 30.69, the deformed carbon nanotubes closest to the topmost moving diamond surface start to slide in a motion reminiscent of the movement of a tank or bulldozer wheel belt. However, when these moving carbon nanotube atoms would have turned the first corner at the top of the ellipse, they encounter the neighboring nanotube and cannot slide past it. This causes them to deform even further, form cross-links with one another, and, in some cases, move in the reverse direction to the sliding motion of the diamond surface. This causes the large oscillations in the normal and lateral forces plotted in Fig. 30.69.

A later study by *Heo* and *Sinnott* examined the frictional properties of single wall, double wall, and filled carbon nanotubes contained between two hydrogen terminated diamondlike carbon layers [30.293]. It was shown that over a wide range of loads the simulations predicted a friction coefficient of 0.13 for filled as well as unfilled nanotubes. This friction coefficient was found to be constant for pressures below 5 GPa, which should include most practical applications. At pressures above 5 GPa the friction coefficient was found to increase to 0.2. Unlike the work of Ni and Sinnott it was demonstrated that the nanotubes would roll in response to the applied shear. This difference was attributed to the finite size of the nanotubes used in this

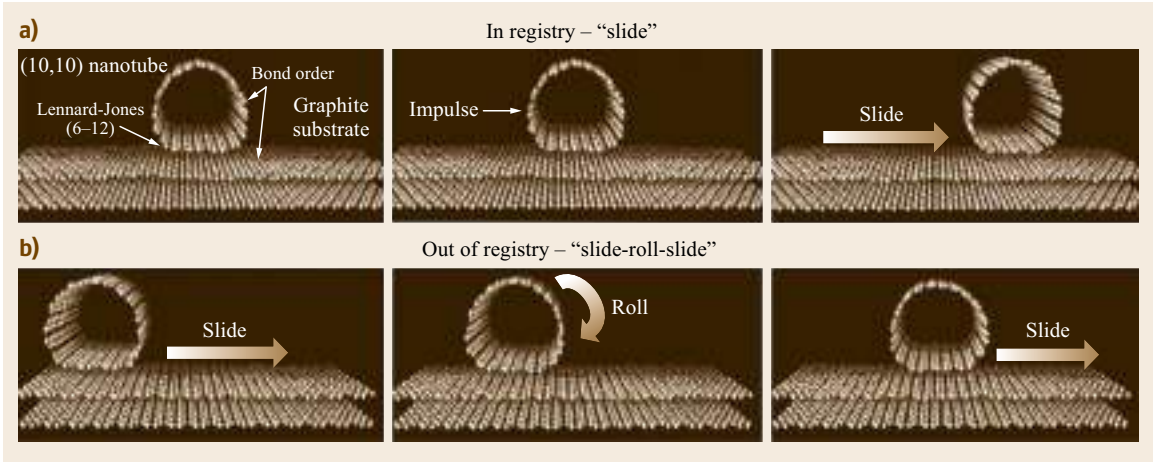


Fig. 30.68a,b Dynamics of a nanotube on a graphite surface. When the nanotube and graphite plane are out of registry **(a)**, the nanotube slides as it slows down from an initial impulse (*upper right panel*). When the nanotube is oriented such that it is in registry **(b)** with the graphite, it slows by a combination of rolling and sliding

simulation as compared to the infinite tubes used by Ni and Sinnott. Although the frictional properties were similar for all systems considered, it was found that filled nanotubes, whether they are double-walled nanotubes or nano-peapods, were more adept at sustaining higher load then the unfilled single-walled nanotubes. Surprisingly, it was also found that the addition of lubricating benzene molecules did little to alter the friction in the system. Rather, it was found that the addition of benzene altered the mechanism by which the system responded to the applied shear stress.

The responses of the horizontally arranged carbon nanotubes are substantially different from the responses of the vertically arranged nanotubes at high compression, as can be seen by comparing Figs. 30.69 and 30.70. The vertical, capped carbon nanotubes are quite flexible and bend and buckle in response to applied forces. As the buckle is forming, the normal force decreases then stabilizes in the buckled structure, as illustrated in Fig. 30.70. As the topmost diamond surface slides, the buckled nanotubes swing around the buckle *neck* which helps dissipate the applied stresses. For this reason, the magnitudes of the lateral forces are not significantly different for the vertical nanotubes at low and high compression, as indicated in Fig. 30.70.

When the ratio of the frictional (lateral) force to the normal force is taken to calculate friction coefficients for these systems, high, nonintuitive values were obtained. As outlined by *Ni* and *Sinnott* [30.282], this is because the actual contact area of the nanotubes is not proportional to the sliding force. In the case of the horizontal nanotube bundles, the tubes are able to deform and significantly change their contact area with the sliding surface with minimal change in the normal force, as shown in Fig. 30.69. In the case of the vertical nanotubes, the contact area remains approximately the

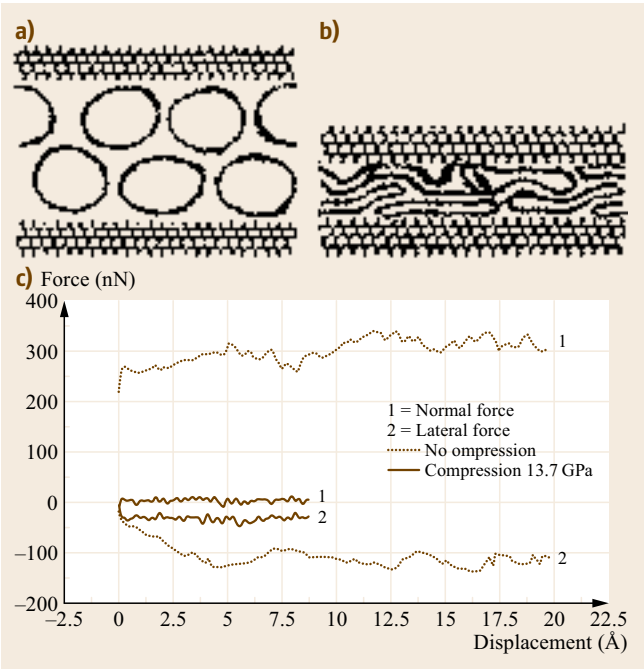


Fig. 30.69 (a,b) Snapshots from simulations that examine the sliding of the topmost diamond surface on horizontally arranged nanotubes at different compressions; **(a)** is at a pressure of ≈ 0 GPa; **(b)** is with a pressure of 13.7 GPa. **(c)** Plots of the normal and lateral components of force during sliding of the top diamond's surface on horizontally arranged nanotubes as a function of the displacement of the top diamond surface with respect to the diamond surface on the bottom

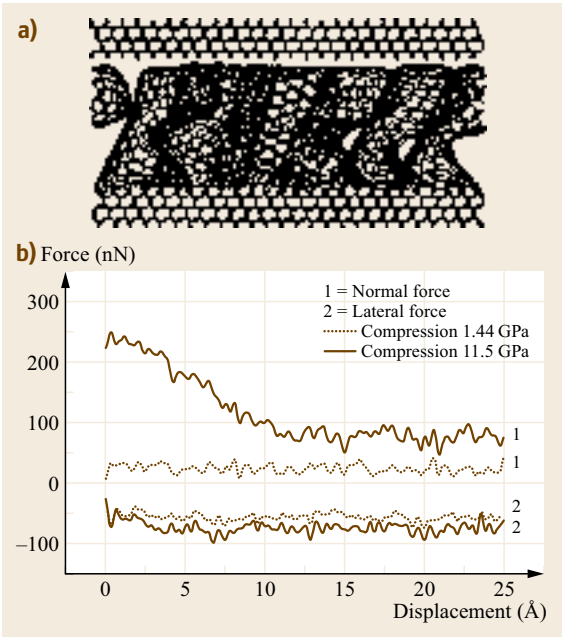


Fig. 30.70 **(a)** Snapshots from simulations that examine the sliding of the topmost diamond surface on vertically arranged nanotubes with one set of capped ends compressed at a pressure of 11.5 GPa. **(b)** Plots of the normal and lateral components of force during sliding of the top diamond's surface on vertically arranged nanotubes as a function of the displacement of the top diamond surface with respect to the diamond surface on the bottom

same regardless of the initial loading force because of the flexibility of the nanotubes. This causes the lateral forces to change only slightly with significant changes in the normal force, as shown in Fig. 30.70. This analysis indicates that care must be taken in calculating friction coefficients for nanotube systems. Recent experiments by *Dickrell* et al. [30.294] show good agreement with these predictions, as shown in Fig. 30.71.

To summarize, this section shows that nanoparticles show some promise as lubricating materials due to their exceptionally low friction coefficients in experiments and simulations. Some nanopaticles show lattice-directed sliding on substrates due to their unique atomic structures. However, there is much that remains to be done before the nanometer-scale friction of these materials is well understood.

Solid State

Surfaces are able to slide over each other at high loads with a minimum of resistance from friction in the presence of liquid lubricants. Some solid thin films can also fulfill these functions and, when they do, are termed solid lubricants. Solid lubricants are generally defined

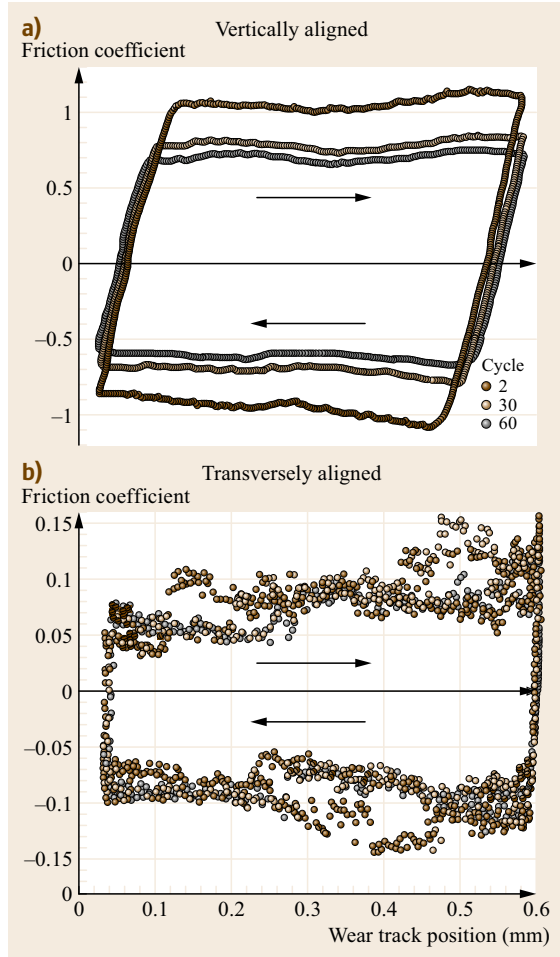


Fig. 30.71a,b Coefficient of friction data versus track position collected for one full cycle of reciprocating sliding for nanotubes that are vertically **(a)** and transversely aligned **(b)** (after [30.294])

as having friction coefficients of 0.3 or less and low wear.

Bowden and *Tabor* showed how thin solid films can reduce friction as follows [30.296]. The total friction force F_f is given as

$$F_f = AF_S + F_p \,, \tag{30.11}$$

where F_p is the plowing term, A is the area of contact and F_S is the shear strength of the interface. If the surfaces are soft, F_S will be reduced while the other parameters will increase. However, if the surfaces under the solid film are very stiff, A and F_p will decrease thereby decreasing friction. The properties specific to the film will also have an effect on friction. For instance, if the films are less than 1 μm thick, the surface asperities will be able to break through the film to eventually

cause wear between the surfaces under normal circumstances. On the other hand, if the lubricant film is too thick, there will be increased plowing and wear that causes the frictional forces to increase. It is important that the lubricant not delaminate in response to the frictional forces, so strong bonds between the lubricant and the surface are required for a solid state lubricant to be effective.

The most common materials used as solid lubricants have layered structures like graphite or MoS_2, that, as discussed above, experience low friction. It is not necessary for the lubricant film to have a layered structure to give low friction. For example, diamondlike carbon has some of the lowest coefficients of friction measured and yet does not have a layered structure. Similarly, not all layered structures are lubricants. For instance, mica gives a relatively high coefficient of friction (> 1).

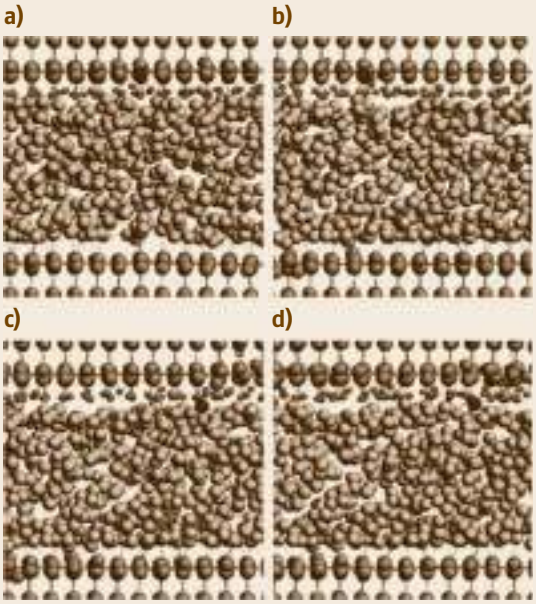


Fig. 30.72a–d A series of chemical reactions induced by sliding of the counterface over the thin film under an average load of 300 nN. **(a)** The sliding causes the rupture of a carbon–hydrogen bond in the counterface. **(b)** The hydrogen atom is incorporated into the film and forms a bond to a carbon atom in the film. **(c)** A bond is formed between the unsaturated carbon atoms in the film and the carbon that suffered the bond rupture in the counterface, and continued sliding causes this carbon to be transferred into the film. **(d)** The transferred carbon forms a bond with another carbon in the counterface. The counterface has slid 0.0 Å **(a)**, 15.9 Å **(b)**, 26.1 Å **(c)**, and 30.5 Å **(d)** (after [30.295], © ACS 2002)

The atomic-scale tribological behavior that occurs when a hydrogen-terminated diamond (111) counterface is in sliding contact with amorphous, hydrogen-free, DLC films was examined using MD simulations by *Gao* et al. [30.295]. Two films, with approximately the same ratio of sp^3–sp^2 carbon but different thicknesses, were examined. Similar average friction was obtained from both films in the load range examined. A series of tribochemical reactions occur above a critical load that result in a significant restructuring of the film, which is analogous to the *run-in* observed in macroscopic friction experiments, and reduces the friction. The contribution of adhesion between the counter-

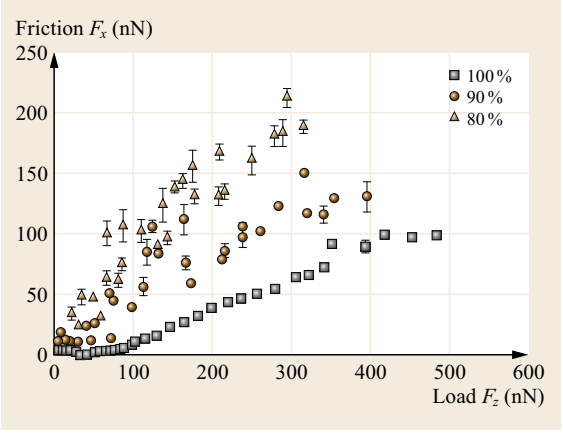


Fig. 30.73 Friction curves for the thin film system with a counterface that is 100% hydrogen-terminated (*gray squares*), 90% hydrogen-terminated (*brown squares*), and 80% hydrogen-terminated (*brown triangles*) (after [30.295], © ACS 2002)

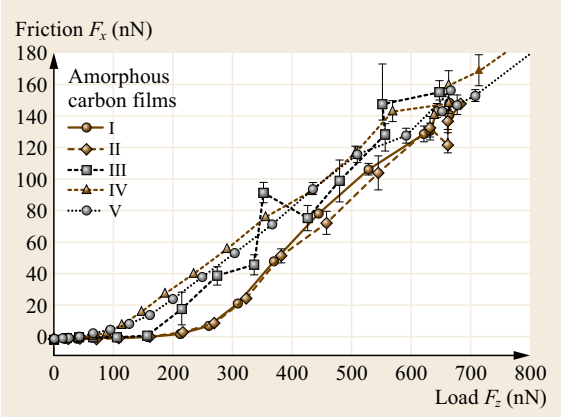


Fig. 30.74 Average friction versus load for five amorphous carbon films. Films I–III are hydrogen-free and contain various ratios of sp^2-to-sp^3 carbon. Films IV and V are both over 90% sp^3 carbon and have surface hydrogenation

face and the sample to friction is examined by varying the saturation of the counterface. The friction increases when the degree of saturation of the diamond counterface is reduced by randomly removing hydrogen atoms. Lastly, two potential energy functions that differ only in their long-range forces are used to examine the contribution of long-range interactions to friction in the same system (as illustrated in Figs. 30.72 and 30.73).

MD simulations were also recently used by *Gao* et al. [30.297] to examine the effects of the sp^2–sp^3 carbon ratio and surface hydrogen on the mechanical and tribological properties of amorphous carbon films. This work showed that, in addition to the sp^2–sp^3 ratio of carbon, the three-dimensional structures of the films are important when determining the mechanical properties of the films. For example, it is possible to have high sp^2-carbon content, which is normally associated with softer films, and large elastic constants. This occurs when sp^2-ringlike structures are oriented perpendicular to the compression direction. The layered nature of the amorphous films examined leads to novel mechanical behavior that influences the shape of the friction versus load data, as illustrated in Fig. 30.74. When load is applied to the films, the film layer closest to the interface is compressed. This results in the very low friction of films I and II up to ≈ 300 nN and the response of films IV and V up to 100 nN. Once the outer film layers have been compressed, additional application of load causes an almost linear increase in friction for films I and II as well as IV and V. Film III has an erratic friction versus load response due to the early onset of tribochemical reactions between the tip and the film.

30.4 Conclusions

This chapter provides a wide-ranging discussion of the background of MD and related simulation methods, their role in the study of nanometer-scale indentation and friction, and their contributions to these fields. Specific, illustrative examples are presented that show how these approaches are providing new and exciting insights into mechanisms responsible for nanoindentation, atomic-scale friction, wear, and related atomic-scale and molecular scale processes. The examples also illustrate how the results from MD and related simulations are complementary to experimental studies, serve to guide experimental work, and assist in the interpretation of experimental data. The ability of these simulations and experimental techniques such as the surface force apparatus and proximal probe microscopes to study nanometer-scale indentation and friction at approximately the same scale is revolutionizing our understanding of the origin of friction at its most fundamental atomic level.

Acknowledgments. S.B.S. and S.-J.H. acknowledge support from the Air Force through grant FA9550-04-1-0367 and from the National Science Foundation supported Network for Computational Nanotechnology (EEC-0228390). D.W.B. and D.L.I. acknowledge support from the Office of Naval Research through grant N00014-04-2006, the National Science Foundation through grant DMR-0304299, the Army Research Office, and the Air Force Office of Scientific Research. J.A.H. acknowledges support from the Air Force Office of Scientific Research under contracts F1ATA08018G001 (Extreme Friction MURI) and F1ATA07351G001. J.A.H. also acknowledges support from The Office of Naval Research under contract N0001408WR20106.

References

30.1 D. Dowson: *History of Tribology* (Longman, London 1979)
30.2 K.L. Johnson, K. Kendell, A.D. Roberts: Surface energy and the contact of elastic solids, Proc. R. Soc. A **324**, 301–313 (1971)
30.3 M. Gad-el-Hak (Ed.): *The MEMS Handbook*, Mech. Eng. Handbook (CRC, Boca Raton 2002)
30.4 J. Krim: Friction at the atomic scale, Sci. Am. **275**, 74–80 (1996)
30.5 J. Krim: Atomic-scale origins of friction, Langmuir **12**, 4564–4566 (1996)
30.6 J. Krim: Progress in nanotribology: Experimental probes of atomic-scale friction, Comments Condens. Matter Phys. **17**, 263–280 (1995)
30.7 A.P. Sutton: Deformation mechanisms, electronic conductance and friction of metallic nanocontacts, Curr. Opin. Solid State Mater. Sci. **1**, 827–833 (1996)
30.8 C.M. Mate: Force microscopy studies of the molecular origins of friction and lubrication, IBM J. Res. Dev. **39**, 617–627 (1995)
30.9 A.M. Stoneham, M.M.D. Ramos, A.P. Sutton: How do they stick together – The statics and dynamics

of interfaces, Philos. Mag. A **67**, 797–811 (1993)
30.10 I.L. Singer: Friction and energy dissipation at the atomic scale: A review, J. Vac. Sci. Technol. A **12**, 2605–2616 (1994)
30.11 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology – Friction, wear and lubrication at the atomic scale, Nature **374**, 607–616 (1995)
30.12 J.A. Harrison, D.W. Brenner: Atomic-scale simulation of tribological and related phenomena. In: *Handbook of Micro/Nanotechnology*, ed. by B. Bhushan (CRC, Boca Raton 1995) pp. 397–439
30.13 J.B. Sokoloff: Theory of atomic level sliding friction between ideal crystal interfaces, J. Appl. Phys. **72**, 1262–1270 (1992)
30.14 W. Zhong, G. Overney, D. Tomanek: Theory of atomic force microscopy on elastic surfaces. In: *The Structure of Surfaces III: Proc. 3rd Int. Conf. Struct. Surf.*, Vol. 24, ed. by S.Y. Tong, M.A.V. Hove, X. Xide, K. Takayanagi (Springer, Berlin, Heidelberg 1991) p. 243
30.15 J.N. Israelachvili: Adhesion, friction and lubrication of molecularly smooth surfaces. In: *Fundamentals of Friction: Macroscopic and Microscopic Processes*, ed. by I.L. Singer, H.M. Pollock (Kluwer, Dordrecht 1992) pp. 351–385
30.16 S.B. Sinnott: Theory of atomic-scale friction. In: *Handbook of Nanostructured Materials and Nanotechnology*, Vol. 2, ed. by H. Nalwa (Academic, San Diego 2000) pp. 571–618
30.17 S.-J. Heo, S.B. Sinnott, D.W. Brenner, J.A. Harrison: Computational modeling of nanometer-scale tribology. In: *Nanotribology and Nanomechanics*, ed. by B. Bhushan (Springer, Berlin, Heidelberg 2005)
30.18 G. Binnig, C.F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56**, 930–933 (1986)
30.19 C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59**, 1942–1945 (1987)
30.20 G.J. Germann, S.R. Cohen, G. Neubauer, G.M. McClelland, H. Seki, D. Coulman: Atomic-scale friction of a diamond tip on diamond (100) surface and (111) surface, J. Appl. Phys. **73**, 163–167 (1993)
30.21 R.W. Carpick, M. Salmeron: Scratching the surface: Fundamental investigations of tribology with atomic force microscopy, Chem. Rev. **97**, 1163–1194 (1997)
30.22 J.N. Israelachvili: *Intermolecular and Surface Forces: With Applications to Colloidal and Biological Systems* (Academic, London 1992)
30.23 J. Krim, D.H. Solina, R. Chiarello: Nanotribology of a Kr monolayer – A quartz crystal microbalance study of atomic-scale friction, Phys. Rev. Lett. **66**, 181–184 (1991)
30.24 G.A. Tomlinson: A molecular theory of friction, Philos. Mag. **7**, 905–939 (1929)
30.25 F.C. Frenkel, T. Kontorova: On the theory of plastic deformation and twinning, Zh. Eksp. Teor. Fiz. **8**, 1340 (1938)
30.26 G.M. McClelland, J.N. Glosli: Friction at the atomic scale. In: *Fundamentals of Friction: Macroscopic and Microscopic Processes*, ed. by I.L. Singer, H.M. Pollock (Kluwer, Dordrecht 1992) pp. 405–422
30.27 J.B. Sokoloff: Theory of dynamical friction between idealized sliding surfaces, Surf. Sci. **144**, 267–272 (1984)
30.28 J.B. Sokoloff: Theory of energy dissipation in sliding crystal surfaces, Phys. Rev. B **42**, 760–765 (1990)
30.29 J.B. Sokoloff: Possible nearly frictionless sliding for mesoscopic solids, Phys. Rev. Lett. **71**, 3450–3453 (1993)
30.30 J.B. Sokoloff: Microscopic mechanisms for kinetic friction: Nearly frictionless sliding for small solids, Phys. Rev. B **52**, 7205–7214 (1995)
30.31 J.B. Sokoloff: Theory of electron and phonon contributions to sliding friction. In: *Physics of Sliding Friction*, ed. by B.N.J. Persson, E. Tosatti (Kluwer, Dordrecht 1996) pp. 217–229
30.32 J.B. Sokoloff: Static friction between elastic solids due to random asperities, Phys. Rev. Lett. **86**, 3312–3315 (2001)
30.33 J.B. Sokoloff: Possible microscopic explanation of the virtually universal occurrence of static friction, Phys. Rev. B **65**, 115415 (2002)
30.34 B.N.J. Persson, D. Schumacher, A. Otto: Surface resistivity and vibrational damping in adsorbed layers, Chem. Phys. Lett. **178**, 204–212 (1991)
30.35 A.I. Volokitin, B.N.J. Persson: Resonant photon tunneling enhancement of the van der Waals friction, Phys. Rev. Lett. **91**, 106101 (2003)
30.36 A.I. Volokitin, B.N.J. Persson: Noncontact friction between nanostructures, Phys. Rev. B **68**, 155420 (2003)
30.37 A.I. Volokitin, B.N.J. Persson: Adsorbate-induced enhancement of electrostatic noncontact friction, Phys. Rev. Lett. **94**, 086104 (2005)
30.38 J.S. Helman, W. Baltensperger, J.A. Holyst: Simple model for dry friction, Phys. Rev. B **49**, 3831–3838 (1994)
30.39 T. Kawaguchi, H. Matsukawa: Dynamical frictional phenomena in an incommensurate two-chain model, Phys. Rev. B **56**, 13932–13942 (1997)
30.40 M.H. Müser: Nature of mechanical instabilities and their effect on kinetic friction, Phys. Rev. Lett. **89**, 224301 (2002)
30.41 M.H. Müser: Towards an atomistic understanding of solid friction by computer simulations, Comput. Phys. Commun. **146**, 54–62 (2002)
30.42 P. Reimann, M. Evstigneev: Nonmonotonic velocity dependence of atomic friction, Phys. Rev. Lett. **93**, 230802 (2004)
30.43 C. Ritter, M. Heyde, B. Stegemann, K. Rademann, U.D. Schwarz: Contact area dependence of frictional forces: Moving adsorbed antimony nanoparticles, Phys. Rev. B **71**, 085405 (2005)
30.44 C. Fusco, A. Fasolino: Velocity dependence of atomic-scale friction: A comparative study of the one- and two-dimensional Tomlinson model, Phys. Rev. B **71**, 045413 (2005)
30.45 C.W. Gear: *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs 1971)

30.46 W.G. Hoover: *Molecular Dynamics* (Springer, Berlin, Heidelberg 1986)
30.47 D.W. Heermann: *Computer Simulation Methods in Theoretical Physics* (Springer, Berlin, Heidelberg 1986)
30.48 M.P. Allen, D.J. Tildesley: *Computer Simulation of Liquids* (Clarendon, Oxford 1987)
30.49 J.M. Haile: *Molecular Dynamics Simulation: Elementary Methods* (Wiley, New York 1992)
30.50 M. Finnis: *Interatomic Forces in Condensed Matter* (Oxford Univ. Press, Oxford 2003)
30.51 D.W. Brenner: Relationship between the embedded-atom method and Tersoff potentials, Phys. Rev. Lett. **63**, 1022–1022 (1989)
30.52 D.W. Brenner: The art and science of an analytic potential, Phys. Status Solidi (b) **217**, 23–40 (2000)
30.53 R.G. Parr, W. Yang: *Density Functional Theory of Atoms and Molecules* (Oxford Univ. Press, New York 1989)
30.54 R. Car, M. Parrinello: Unified approach for molecular dynamics and density functional theory, Phys. Rev. Lett. **55**, 2471–2474 (1985)
30.55 C. Cramer: *Essentials of Computational Chemistry, Theories and Models* (Wiley, Chichester 2004)
30.56 A.P. Sutton: *Electronic Structure of Materials* (Clarendon, Oxford 1993)
30.57 K. Kadau, T.C. Germann, P.S. Lomdahl: Large-scale molecular dynamics simulation of 19 billion particles, Int. J. Mod. Phys. C **15**, 193–201 (2004)
30.58 B.J. Thijsse: Relationship between the modified embedded-atom method and Stillinger–Weber potentials in calculating the structure of silicon, Phys. Rev. B **65**, 195207 (2002)
30.59 M.I. Baskes, J.S. Nelson, A.F. Wright: Semiempirical modified embedded atom potentials for silicon and germanium, Phys. Rev. B **40**, 6085–6100 (1989)
30.60 T. Ohira, Y. Inoue, K. Murata, J. Murayama: Magnetite scale cluster adhesion on metal oxide surfaces: Atomistic simulation study, Appl. Surf. Sci. **171**, 175–188 (2001)
30.61 F.H. Streitz, J.W. Mintmire: Electrostatic potentials for metal oxide surfaces and interfaces, Phys. Rev. B **50**, 11996–12003 (1994)
30.62 A. Yasukawa: Using an extended Tersoff interatomic potential to analyze the static fatigue strength of SiO_2 under atmospheric influence, JSME Int. J. A **39**, 313–320 (1996)
30.63 T. Iwasaki, H. Miura: Molecular dynamics analysis of adhesion strength of interfaces between thin films, J. Mater. Res. **16**, 1789–1794 (2001)
30.64 B.-J. Lee, M.I. Baskes: Second nearest-neighbor modified embedded-atom method potential, Phys. Rev. B **62**, 8564–8567 (2000)
30.65 G.C. Abell: Empirical chemical pseudopotential theory of molecular and metallic bonding, Phys. Rev. B **31**, 6184–6196 (1985)
30.66 J. Tersoff: New empirical approach for the structure and energy of covalent systems, Phys. Rev. B **37**, 6991–7000 (1988)
30.67 J. Tersoff: Modeling solid-state chemistry: Interatomic potentials for multicomponent systems, Phys. Rev. B **39**, 5566–5569 (1989)
30.68 D.W. Brenner: Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films, Phys. Rev. B **42**, 9458–9471 (1990)
30.69 D.W. Brenner, O.A. Shenderova, J.A. Harrison, S.J. Stuart, B. Ni, S.B. Sinnott: Second generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons, J. Phys. C **14**, 783–802 (2002)
30.70 A.J. Dyson, P.V. Smith: Extension of the Brenner empirical interactomic potential to C-Si-H, Surf. Sci. **355**, 140–150 (1996)
30.71 B. Ni, K.-H. Lee, S.B. Sinnott: Development of a reactive empirical bond order potential for hydrocarbon-oxygen interactions, J. Phys. C **16**, 7261–7275 (2004)
30.72 J. Tanaka, C.F. Abrams, D.B. Graves: New C-F interatomic potential for molecular dynamics simulation of fluorocarbon film formation, Nucl. Instrum. Methods B **18**, 938–945 (2000)
30.73 I. Jang, S.B. Sinnott: Molecular dynamics simulations of the chemical modification of polystyrene through $C_xF_y^+$ beam deposition, J. Phys. Chem. B **108**, 9656–9664 (2004)
30.74 J.D. Schall, G. Gao, J.A. Harrison: Elastic constants of silicon materials calculated as a function of temperature using a parametrization of the second-generation reactive empirical bond-order potential, Phys. Rev. B **77**(11), 115209 (2008)
30.75 Y. Hu: Personal communication (2008)
30.76 S.B. Sinnott, O.A. Shenderova, C.T. White, D.W. Brenner: Mechanical properties of nanotubule fibers and composites determined from theoretical calculations and simulations, Carbon **36**, 1–9 (1998)
30.77 S.J. Stuart, A.B. Tutein, J.A. Harrison: A reactive potential for hydrocarbons with intermolecular interactions, J. Chem. Phys. **112**, 6472–6486 (2000)
30.78 F.H. Stillinger, T.A. Weber: Computer simulation of local order in condensed phases of silicon, Phys. Rev. B **31**, 5262–5271 (1985)
30.79 S.M. Foiles: Application of the embedded-atom method to liquid transition metals, Phys. Rev. B **32**, 3409–3415 (1985)
30.80 M.S. Daw, M.I. Baskes: Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals, Phys. Rev. Lett. **50**, 1285–1288 (1983)
30.81 T.J. Raeker, A.E. Depristo: Theory of chemical bonding based on the atom-homogeneous electron gas system, Int. Rev. Phys. Chem. **10**, 1–54 (1991)
30.82 R.W. Smith, G.S. Was: Application of molecular dynamics to the study of hydrogen embrittlement in Ni-Cr-Fe alloys, Phys. Rev. B **40**, 10322–10336 (1989)
30.83 R. Pasianot, D. Farkas, E.J. Savino: Empirical many-body interatomic potential for bcc transition metals, Phys. Rev. B **43**, 6952–6961 (1991)
30.84 R. Pasianot, E.J. Savino: Embedded-atom method interatomic potentials for hcp metals, Phys. Rev. B

45, 12704–12710 (1992)
30.85 M.I. Baskes, J.S. Nelson, A.F. Wright: Semiempirical modified embedded-atom potentials for silicon and germanium, Phys. Rev. B **40**, 6085–6100 (1989)
30.86 M.I. Baskes: Modified embedded-atom potentials for cubic materials and impurities, Phys. Rev. B **46**, 2727–2742 (1992)
30.87 K. Ohno, K. Esfarjani, Y. Kawazoe: *Computational Materials Science from ab initio to Monte Carlo Methods* (Springer, Berlin, Heidelberg 1999)
30.88 A.K. Rappe, W.A. Goddard III: Charge equilibration for molecular dynamics simulations, J. Phys. Chem. **95**, 3358–3363 (1991)
30.89 D. Frenkel, B. Smit: *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego 1996)
30.90 L.V. Woodcock: Isothermal molecular dynamics calculations for liquid salts, Chem. Phys. Lett. **10**, 257–261 (1971)
30.91 T. Schneider, E. Stoll: Molecular dynamics study of a three-dimensional one-component model for distortive phase transitions, Phys. Rev. B **17**, 1302–1322 (1978)
30.92 K. Kremer, G.S. Grest: Dynamics of entangled linear polymer melts – A molecular dynamics simulation, J. Chem. Phys. **92**, 5057–5086 (1990)
30.93 S.A. Adelman, J.D. Doll: Generalized Langevin equation approach for atom-solid-surface scattering – General formulation for classical scattering off harmonic solids, J. Chem. Phys. **64**, 2375–2388 (1976)
30.94 S.A. Adelman: Generalized Langevin equations and many-body problems in chemical dynamics, Adv. Chem. Phys. **44**, 143–253 (1980)
30.95 J.C. Tully: Dynamics of gas-surface interactions – 3-D generalized Langevin model applied to fcc and bcc surfaces, J. Chem. Phys. **73**, 1975–1985 (1980)
30.96 S. Nosé: A unified formulation of the constant temperature molecular dynamics methods, J. Chem. Phys. **81**, 511–519 (1984)
30.97 S. Nosé: A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. **52**, 255–268 (1984)
30.98 G.J. Martyna, M.L. Klein, M. Tuckerman: Nosé–Hoover chains – The canonical ensemble via continuous dynamics, J. Chem. Phys. **97**, 2635–2643 (1992)
30.99 M. D'Alessandro, M. D'Abramo, G. Brancato, A. Di Nola, A. Amadei: Statistical mechanics and thermodynamics of simulated ionic solutions, J. Phys. Chem. B **106**, 11843–11848 (2002)
30.100 J.D. Schall, C.W. Padgett, D.W. Brenner: Ad hoc continuum-atomistic thermostat for modeling heat flow in molecular dynamics simulations, Mol. Simul. **31**, 283–288 (2005)
30.101 C.W. Padgett, D.W. Brenner: A continuum-atomistic method for incorporating Joule heating into classical molecular dynamics simulations, Mol. Simul. **31**(11), 749–757 (2005)
30.102 M. Schoen, C.L. Rhykerd, D.J. Diestler, J.H. Cushman: Shear forces in molecularly thin films, Science **245**, 1223–1225 (1989)
30.103 J.E. Curry, F.S. Zhang, J.H. Cushman, M. Schoen, D.J. Diestler: Transient coexisting nanophases in ultrathin films confined between corrugated walls, J. Chem. Phys. **101**, 10824–10832 (1994)
30.104 D.J. Adams: Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid, Mol. Phys. **29**, 307–311 (1975)
30.105 R.J. Colton, N.A. Burnham: Force microscopy. In: *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications*, ed. by D.A. Bonnell (VCH, New York 1993) pp. 191–249
30.106 E. Meyer: *Nanoscience: Friction and Rheology on the Nanometer Scale* (World Scientific, Hackensack 1998)
30.107 G.E. Totten, H. Liang: *Mechanical Tribology: Materials Characterization and Applications* (Marcel Dekker, New York 2004)
30.108 N.A. Burnham, R.J. Colton: Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope, J. Vac. Sci. Technol. A **7**, 2906–2913 (1989)
30.109 N.A. Burnham, D.D. Dominguez, R.L. Mowery, R.J. Colton: Probing the surface forces of monolayer films with an atomic force microscope, Phys. Rev. Lett. **64**, 1931–1934 (1990)
30.110 E. Meyer, R. Overney, D. Brodbeck, L. Howald, R. Luthi, J. Frommer, H.J. Guntherodt: Friction and wear of Langmuir–Blodgett films observed by friction force microscopy, Phys. Rev. Lett. **69**, 1777–1780 (1992)
30.111 A.P. Sutton, J.B. Pethica, H. Rafii-Tabar, J.A. Nieminen: Mechanical properties of metals at the nanometer scale. In: *Electron Theory in Alloy Design*, ed. by D.G. Pettifor, A.H. Cottrell (Institute of Materials, London 1992) pp. 191–233
30.112 H. Raffi-Tabar, A.P. Sutton: Long-range Finnis–Sinclair potentials for fcc metallic alloys, Philos. Mag. Lett. **63**, 217–224 (1991)
30.113 U. Landman, W.D. Luedtke, E.M. Ringer: Atomistic mechanisms of adhesive contact formation and interfacial processes, Wear **153**, 3–30 (1992)
30.114 U. Landman, W.D. Luedtke, N.A. Burnham, R.J. Colton: Atomistic mechanisms and dynamics of adhesion, nanoindentation and fracture, Science **248**, 454–461 (1990)
30.115 O. Tomagnini, F. Ercolessi, E. Tosatti: Microscopic interaction between a gold tip and a Pb(110) surface, Surf. Sci. **287/288**, 1041–1045 (1991)
30.116 N. Ohmae: Field ion microscopy of microdeformation induced by metallic contacts, Philos. Mag. A **74**, 1319–1327 (1996)
30.117 N.A. Burnham, R.J. Colton, H.M. Pollock: Interpretation of force curves in force microscopy, Nanotechnology **4**, 64–80 (1993)
30.118 N. Agrait, G. Rubio, S. Vieira: Plastic deformation in nanometer-scale contacts, Langmuir **12**, 4505–4509 (1996)

30.119 U. Landman, W.D. Luedtke, A. Nitzan: Dynamics of tip–substrate interactions in atomic force microscopy, Surf. Sci. **210**, L177–L182 (1989)
30.120 U. Landman, W.D. Luedtke: Nanomechanics and dynamics of tip substrate interactions, J. Vac. Sci. Technol. B **9**, 414–423 (1991)
30.121 U. Landman, W.D. Luedtke, J. Ouyang, T.K. Xia: Nanotribology and the stability of nanostructures, Jpn. J. Appl. Phys. **32**, 1444–1462 (1993)
30.122 J.W.M. Frenken, H.M. Vanpinxteren, L. Kuipers: New views on surface melting obtained with STM and ion scattering, Surf. Sci. **283**, 283–289 (1993)
30.123 O. Tomagnini, F. Ercolessi, E. Tosatti: Microscopic interaction between a gold tip and a Pb(110) surface, Surf. Sci. **287**, 1041–1045 (1993)
30.124 K. Komvopoulos, W. Yan: Molecular dynamics simulation of single and repeated indentation, J. Appl. Phys. **82**, 4823–4830 (1997)
30.125 T. Yokohata, K. Kato: Mechanism of nanoscale indentation, Wear **168**, 109–114 (1993)
30.126 J. Belak, I.F. Stowers: A molecular dynamics model of the orthogonal cutting process. In: *Proc. Am. Soc. Precis. Eng. Annu. Conf.* (1990) pp. 76–79
30.127 M. Fournel, E. Lacaze, M. Schott: Tip–surface interactions in STM experiments on Au(111): Atomic-scale metal friction, Europhys. Lett. **34**, 489–494 (1996)
30.128 J.L. Costakramer, N. Garcia, P. Garciamochales, P.A. Serena: Nanowire formation in macroscopic metallic contacts – Quantum-mechanical conductance tapping a table top, Surf. Sci. **342**, L1144–L1149 (1995)
30.129 A.I. Yanson, J.M. van Ruitenbeek, I.K. Yanson: Shell effects in alkali metal nanowires, Low Temp. Phys. **27**, 807–820 (2001)
30.130 A.I. Yanson, I.K. Yanson, J.M. van Ruitenbeek: Crossover from electronic to atomic shell structure in alkali metal nanowires, Phys. Rev. Lett. **8721**, 216805 (2001)
30.131 C.L. Kelchner, S.J. Plimpton, J.C. Hamilton: Dislocation nucleation and defect structure during surface indentation, Phys. Rev. B **58**, 11085–11088 (1998)
30.132 E.T. Lilleodden, J.A. Zimmerman, S.M. Foiles, W.D. Nix: Atomistic simulations of elastic deformation and dislocation nucleation during nanoindentation, J. Mech. Phys. Solids **51**, 901–920 (2003)
30.133 A. Hasnaoui, P.M. Derlet, H.V. Swygenhoven: Interaction between dislocations, grain boundaries under an indenter – A molecular dynamics simulation, Acta Mater. **52**, 2251–2258 (2004)
30.134 O.R. de la Fuente, J.A. Zimmerman, M.A. Gonzalez, J. de la Figuera, J.C. Hamilton, W.W. Pai, J.M. Rojo: Dislocation emission around nanoindentations on a (001) fcc metal surface studied by scanning tunneling microscopy and atomistic simulations, Phys. Rev. Lett. **88**, 036101 (2002)
30.135 O.A. Shenderova, J.P. Mewkill, D.W. Brenner: Nanoindentation as a probe of nanoscale residual stresses, Mol. Simul. **25**, 81–92 (2000)
30.136 S. Kokubo: On the change in hardness of a plate caused by bending, Sci. Rep. Tohoku Imp. Univ. **21**, 256–267 (1932)
30.137 G. Sines, R. Calson: Hardness measurements for determination of residual stresses, ASTM Bulletin **180**, 35–37 (1952)
30.138 G.U. Oppel: Biaxial elasto-plastic analysis of load and residual stresses, Exp. Mech. **21**, 135–140 (1964)
30.139 T.R. Simes, S.G. Mellor, D.A. Hills: A note on the influence of residual stress on measured hardness, J. Strain Anal. Eng. Des. **19**, 135–137 (1984)
30.140 T.Y. Tsui, G.M. Pharr, W.C. Oliver, C.S. Bhatia, C.T. White, S. Anders, A. Anders, I.G. Brown: Nanoindentation and nanoscratching of hard carbon coatings for magnetic disks, Mater. Res. Soc. Symp. Proc. **383**, 447–452 (1995)
30.141 A. Bolshakov, W.C. Oliver, G.M. Pharr: Influences of stress on the measurement of mechanical properties using nanoindentation. 2. Finite element simulations, J. Mater. Res. **11**, 760–768 (1996)
30.142 J.D. Schall, D.W. Brenner: Atomistic simulation of the influence of pre-existing stress on the interpretation of nanoindentation data, J. Mater. Res. **19**, 3172–3180 (2004)
30.143 U. Landman, W.D. Luedtke, M.W. Ribarsky: Structural and dynamical consequences of interactions in interfacial systems, J. Vac. Sci. Technol. A **7**, 2829–2839 (1989)
30.144 J.S. Kallman, W.G. Hoover, C.G. Hoover, A.J. Degroot, S.M. Lee, F. Wooten: Molecular-dynamics of silicon indentation, Phys. Rev. B **47**, 7705–7709 (1993)
30.145 W.C.D. Cheong, L.C. Zhang: Molecular dynamics simulation of phase transformations in silicon monocrystals due to nano-indentation, Nanotechnology **11**, 173–180 (2000)
30.146 C.F. Sanz-Navarro, S.D. Kenny, R. Smith: Atomistic simulations of structural transformations, Nanotechnology **15**, 692–697 (2004)
30.147 D.R. Clarke, M.C. Kroll, P.D. Kirchner, R.F. Cook, B.J. Hockey: Amorphization and conductivity of silicon and germanium induced by indentation, Phys. Rev. Lett. **60**, 2156–2159 (1988)
30.148 A. Kailer, K.G. Nickel, Y.G. Gogotsi: Raman microspectroscopy of nanocrystalline and amorphous phases in hardness indentations, J. Raman Spectrosc. **30**, 939–961 (1999)
30.149 P. Walsh, A. Omeltchenko, R.K. Kalia, A. Nakano, P. Vashishta, S. Saini: Nanoindentation of silicon nitride: A multimillion-atom molecular dynamics study, Appl. Phys. Lett. **82**, 118–120 (2003)
30.150 J.A. Harrison, C.T. White, R.J. Colton, D.W. Brenner: Nanoscale investigation of indentation, adhesion and fracture of diamond (111) surfaces, Surf. Sci. **271**, 57–67 (1992)
30.151 K. Minowa, K. Sumino: Stress-induced amorphization of a silicon crystal by mechanical scratching, Phys. Rev. Lett. **69**, 320–322 (1992)
30.152 G.S. Smith, E.B. Tadmor, E. Kaxiras: Multiscale simulation of loading and electrical resistance

in silicon nanoindentation, Phys. Rev. Lett. **84**, 1260–1263 (2000)
30.153 K. Cho, J.D. Joannopoulos: Mechanical hysteresis on an atomic-scale, Surf. Sci. **328**, 320–324 (1995)
30.154 J.A. Harrison, S.J. Stuart, D.H. Robertson, C.T. White: Properties of capped nanotubes when used as SPM tips, J. Phys. Chem. B **101**, 9682–9685 (1997)
30.155 J.A. Harrison, S.J. Stuart, A.B. Tutein: A new, reactive potential energy function to study indentation and friction of $C_{13}n$-alkane monolayers. In: *Interfacial Properties on the Submicron Scale*, ed. by J.E. Frommer, R. Overney (ACS, Washington 2001) pp. 216–229
30.156 A. Garg, J. Han, S.B. Sinnott: Interactions of carbon-nanotubule proximal probe tips with diamond and graphene, Phys. Rev. Lett. **81**, 2260–2263 (1998)
30.157 A. Garg, S.B. Sinnott: Molecular dynamics of carbon nanotubule proximal probe tip-surface contacts, Phys. Rev. B **60**, 13786–13791 (1999)
30.158 K.J. Tupper, D.W. Brenner: Compression-induced structural transition in a self-assembled monolayer, Langmuir **10**, 2335–2338 (1994)
30.159 K.J. Tupper, R.J. Colton, D.W. Brenner: Simulations of self-assembled monolayers under compression – Effect of surface asperities, Langmuir **10**, 2041–2043 (1994)
30.160 S.A. Joyce, R.C. Thomas, J.E. Houston, T.A. Michalske, R.M. Crooks: Mechanical relaxation of organic monolayer films measured by force microscopy, Phys. Rev. Lett. **68**, 2790–2793 (1992)
30.161 L. Zhang, Y. Leng, S. Jiang: Tip-based hybrid simulation study of frictional properties of self-assembled monolayers: Effects of chain length, terminal group, and scan direction, scan velocity, Langmuir **19**, 9742–9747 (2003)
30.162 A.B. Tutein, S.J. Stuart, J.A. Harrison: Indentation analysis of linear-chain hydrocarbon monolayers anchored to diamond, J. Phys. Chem. B **103**, 11357–11365 (1999)
30.163 Y. Leng, S. Jiang: Dynamic simulations of adhesion and friction in chemical force microscopy, J. Am. Chem. Soc. **124**, 11764–11770 (2002)
30.164 C.M. Mate: Atomic force microscope study of polymer lubricants on silicon surfaces, Phys. Rev. Lett. **68**, 3323–3326 (1992)
30.165 K. Enke, H. Dimigen, H. Hubsch: Frictional properties of diamond-like carbon layers, Appl. Phys. Lett. **36**, 291–292 (1980)
30.166 K. Enke: Some new results on the fabrication of and the mechanical, electrical, optical properties of i-carbon layers, Thin Solid Films **80**, 227–234 (1981)
30.167 S. Miyake, S. Takahashi, I. Watanabe, H. Yoshihara: Friction and wear behavior of hard carbon films, ASLE Transactions **30**, 121–127 (1987)
30.168 A. Erdemir, C. Donnet: Tribology of diamond, diamond-like carbon, and related films. In: *Modern Tribology Handbook*, Vol. II, ed. by B. Bhushan (CRC, Boca Raton 2000) pp. 871–908
30.169 S.B. Sinnott, R.J. Colton, C.T. White, O.A. Shenderova, D.W. Brenner, J.A. Harrison: Atomistic simulations of the nanometer-scale indentation of amorphous carbon thin films, J. Vac. Sci. Technol. A **15**, 936–940 (1997)
30.170 J.N. Glosli, M.R. Philpott, G.M. McClelland: Molecular dynamics simulation of mechanical deformation of ultra-thin amorphous carbon films, Mater. Res. Soc. Symp. Proc. **383**, 431–435 (1995)
30.171 I.L. Singer: A thermochemical model for analyzing low wear-rate materials, Surf. Coat. Technol. **49**, 474–481 (1991)
30.172 I.L. Singer, S. Fayeulle, P.D. Ehni: Friction and wear behavior of tin in air – The chemistry of transfer films and debris formation, Wear **149**, 375–394 (1991)
30.173 J.A. Harrison, C.T. White, R.J. Colton, D.W. Brenner: Molecular dynamics simulations of atomic-scale friction of diamond surfaces, Phys. Rev. B **46**, 9700–9708 (1992)
30.174 J.A. Harrison, R.J. Colton, C.T. White, D.W. Brenner: Effect of atomic-scale surface roughness on friction – A molecular dynamics study of diamond surfaces, Wear **168**, 127–133 (1993)
30.175 J.A. Harrison, C.T. White, R.J. Colton, D.W. Brenner: Atomistic simulations of friction at sliding diamond interfaces, MRS Bulletin **18**, 50–53 (1993)
30.176 J.N. Glosli, G.M. McClelland: Molecular dynamics study of sliding friction of ordered organic monolayers, Phys. Rev. Lett. **70**, 1960–1963 (1993)
30.177 A. Koike, M. Yoneya: Molecular dynamics simulations of sliding friction of Langmuir–Blodgett monolayers, J. Chem. Phys. **105**, 6060–6067 (1996)
30.178 J.E. Hammerberg, B.L. Holian, S.J. Zhou: Studies of sliding friction in compressed copper. In: *Conf. Am. Phys. Soc. Top. Group Shock Compress. Condens. Matter, Seattle*, ed. by S.C. Schmidt, W.C. Tao (AIP, New York 1995) p. 370
30.179 M.R. Sorensen, K.W. Jacobsen, P. Stoltze: Simulations of atomic-scale sliding friction, Phys. Rev. B **53**, 2101–2113 (1996)
30.180 M.D. Perry, J.A. Harrison: Friction between diamond surfaces in the presence of small third-body molecules, J. Phys. Chem. B **101**, 1364–1373 (1997)
30.181 A. Buldum, S. Ciraci: Atomic-scale study of dry sliding friction, Phys. Rev. B **55**, 2606–2611 (1997)
30.182 A.P. Sutton, J.B. Pithica: Inelastic flow processes in nanometre volumes of solids, J. Phys. Condens. Matter **2**, 5317–5326 (1990)
30.183 S. Akamine, R.C. Barrett, C.F. Quate: Improved atomic force microscope images using microcantilevers with sharp tips, Appl. Phys. Lett. **57**, 316–318 (1990)
30.184 J.A. Nieminen, A.P. Sutton, J.B. Pethica: Static junction growth during frictional sliding of metals, Acta Metall. Mater. **40**, 2503–2509 (1992)
30.185 J.A. Niemienen, A.P. Sutton, J.B. Pethica, K. Kaski: Mechanism of lubrication by a thin solid film on a metal surface, Model. Simul. Mater. Sci. Eng. **1**, 83–90 (1992)

30.186 V.V. Pokropivny, V.V. Skorokhod, A.V. Pokropivny: Atomistic mechanism of adhesive wear during friction of atomic sharp tungsten asperity over (114) bcc-iron surface, Mater. Lett. **31**, 49–54 (1997)
30.187 B. Li, P.C. Clapp, J.A. Rifkin, X.M. Zhang: Molecular dynamics simulation of stick-slip, J. Appl. Phys. **90**, 3090–3094 (2001)
30.188 T.-H. Fang, C.-I. Weng, J.-G. Chang: Molecular dynamics simulation of nanolithography process using atomic force microscopy, Surf. Sci. **501**, 138–147 (2002)
30.189 R. Komanduri, N. Chandrasekaran: Molecular dynamics simulation of atomic-scale friction, Phys. Rev. B **61**, 14007–14019 (2000)
30.190 A. Dayo, W. Alnasrallah, J. Krim: Superconductivity-dependent sliding friction, Phys. Rev. Lett. **80**, 1690–1693 (1998)
30.191 R. Erlandsson, G. Hadziioannou, C.M. Mate, G.M. McClelland, S. Chiang: Atomic scale friction between the muscovite mica cleavage plane and a tungsten tip, J. Chem. Phys. **89**, 5190–5193 (1988)
30.192 K.L. Johnson: *Contact Mechanics* (Cambridge Univ. Press, Cambridge 1985)
30.193 J.B. Pethica: Interatomic forces in scanning tunneling microscopy – Giant corrugations of the graphite surface – Comment, Phys. Rev. Lett. **57**, 3235–3235 (1986)
30.194 H. Tang, C. Joachim, J. Devillers: Interpretation of AFM images – The graphite surface with a diamond tip, Surf. Sci. **291**, 439–450 (1993)
30.195 S. Fujisawa, Y. Sugawara, S. Morita: Localized fluctuation of a two-dimensional atomic-scale friction, Jpn. J. Appl. Phys. **35**, 5909–5913 (1996)
30.196 S. Fujisawa, Y. Sugawara, S. Ito, S. Mishima, T. Okada, S. Morita: The two-dimensional stick-slip phenomenon with atomic resolution, Nanotechnology **4**, 138–142 (1993)
30.197 S. Fujisawa, Y. Sugawara, S. Morita, S. Ito, S. Mishima, T. Okada: Study on the stick-slip phenomenon on a cleaved surface of the muscovite mica using an atomic-force lateral force microscope, J. Vac. Sci. Technol. B **12**, 1635–1637 (1994)
30.198 S. Morita, S. Fujisawa, Y. Sugawara: Spatially quantized friction with a lattice periodicity, Surf. Sci. Rep. **23**, 1–41 (1996)
30.199 J.A. Ruan, B. Bhushan: Atomic-scale and microscale friction studies of graphite and diamond using friction force microscopy, J. Appl. Phys. **76**, 5022–5035 (1994)
30.200 R.W. Carpick, N. Agrait, D.F. Ogletree, M. Salmeron: Variation of the interfacial shear strength and adhesion of a nanometer-sized contact, Langmuir **12**, 3334–3340 (1996)
30.201 R.W. Carpick, N. Agrait, D.F. Ogletree, M. Salmeron: Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscope, J. Vac. Sci. Technol. B **14**, 1289–2772 (1996)
30.202 T. Liang, W.G. Sawyer, S.S. Perry, S.B. Sinnott, S.R. Phillpot: First-principles determination of static potential energy surfaces for atomic friction in MoS_2 and MoO_3, Phys. Rev. B **77**(10), 104105 (2008)
30.203 A.L. Shluger, R.T. Williams, A.L. Rohl: Lateral and friction forces originating during force microscope scanning of ionic surfaces, Surf. Sci. **343**, 273–287 (1995)
30.204 B. Samuels, J. Wilks: The friction of diamond sliding on diamond, J. Mater. Sci. **23**, 2846–2864 (1988)
30.205 T. Cagin, J.W. Che, M.N. Gardos, A. Fijany, W.A. Goddard: Simulation and experiments on friction, wear of diamond: A material for MEMS and NEMS application, Nanotechnology **10**, 278–284 (1999)
30.206 R.J.A. van den Oetelaar, C.F.J. Flipse: Atomic-scale friction on diamond(111) studied by ultra-high vacuum atomic force microscopy, Surf. Sci. **384**, L828–L835 (1997)
30.207 M. Enachescu, R.J.A. van den Oetelaar, R.W. Carpick, D.F. Ogletree, C.F.J. Flipse, M. Salmeron: Atomic force microscopy study of an ideally hard contact: The diamond(111) tungsten carbide interface, Phys. Rev. Lett. **81**, 1877–1880 (1998)
30.208 B.V. Derjaguin, V.M. Muller, Y. Toporov: Effect of contact deformations on adhesion of particles, J. Colloid Interf. Sci. **53**, 314–326 (1975)
30.209 J.A. Harrison, C.T. White, R.J. Colton, D.W. Brenner: Investigation of the atomic-scale friction and energy dissipation in diamond using molecular dynamics, Thin Solid Films **260**, 205–211 (1995)
30.210 M.D. Perry, J.A. Harrison: Universal aspects of the atomic-scale friction of diamond surfaces, J. Phys. Chem. B **99**, 9960–9965 (1995)
30.211 M.J. Brukman, G.G.R.J. Nemanich, J.A. Harrison: Temperature dependence of single asperity diamond-diamond friction elucidated using AFM and MD simulations, J. Phys. Chem. **112**, 9358–9369 (2008)
30.212 R. Neitola, T.A. Pakannen: Ab initio studies on the atomic-scale origin of friction between diamond (111) surfaces, J. Phys. Chem. B **105**, 1338–1343 (2001)
30.213 J.A. Harrison, R.J. Colton, C.T. White, D.W. Brenner: Atomistic simulation of the nanoindentation of diamond and graphite surfaces, Mater. Res. Soc. Symp. Proc. **239**, 573–578 (1992)
30.214 D. Mulliah, S.D. Kenny, R. Smith: Modeling of stick-slip phenomena using molecular dynamics, Phys. Rev. B **69**, 205407 (2004)
30.215 D.W. Brenner: Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films, Phys. Rev. B **42**, 9458–9471 (1990)
30.216 G.J. Ackland, G. Tichy, V. Vitek, M.W. Finnis: Simple n-body potentials for the noble metals and nickel, Philos. Mag. A **56**, 735–756 (1987)
30.217 J.P. Biersack, J. Ziegler, U. Littmack: *The Stopping and Range of Ions in Solids* (Pergamon, Oxford 1985)
30.218 J. Cai, J.-S. Wang: Friction between Si tip and (001)–2 × 1 surface: A molecular dynamics simulation, Comput. Phys. Commun. **147**, 145–148 (2002)

30.219 J. Cai, J.S. Wang: Friction between a Ge tip and the (001)–2 × 1 surface: A molecular dynamics simulation, Phys. Rev. B **64**, 113313 (2001)

30.220 A.G. Khurshudov, K. Kato, H. Koide: Nano-wear of the diamond AFM probing tip under scratching of silicon, studied by AFM, Tribol. Lett. **2**, 345–354 (1996)

30.221 A. Khurshudov, K. Kato: Volume increase phenomena in reciprocal scratching of polycarbonate studied by atomic-force microscopy, J. Vac. Sci. Technol. B **13**, 1938–1944 (1995)

30.222 M.D. Perry, J.A. Harrison: Molecular dynamics studies of the frictional properties of hydrocarbon materials, Langmuir **12**, 4552–4556 (1996)

30.223 M.D. Perry, J.A. Harrison: Molecular dynamics investigations of the effects of debris molecules on the friction and wear of diamond, Thin Solid Films **291**, 211–215 (1996)

30.224 J.A. Harrison, C.T. White, R.J. Colton, D.W. Brenner: Effects of chemically-bound, flexible hydrocarbon species on the frictional properties of diamond surfaces, J. Phys. Chem. **97**, 6573–6576 (1993)

30.225 J.A. Harrison, D.W. Brenner: Simulated tribochemistry – An atomic-scale view of the wear of diamond, J. Am. Chem. Soc. **116**, 10399–10402 (1994)

30.226 Z. Feng, J.E. Field: Friction of diamond on diamond and chemical vapor deposition diamond coatings, Surf. Coat. Technol. **47**, 631–645 (1991)

30.227 B.N.J. Persson: Applications of surface resistivity to atomic scale friction, to the migration of hot adatoms, and to electrochemistry, J. Chem. Phys. **98**, 1659–1672 (1993)

30.228 B.N.J. Persson, A.I. Volokitin: Electronic friction of physisorbed molecules, J. Chem. Phys. **103**, 8679–8683 (1995)

30.229 H. Grabhorn, A. Otto, D. Schumacher, B.N.J. Persson: Variation of the dc-resistance of smooth and atomically rough silver films during exposure to C_2H_6 and C_2H_4, Surf. Sci. **264**, 327–340 (1992)

30.230 A.D. Berman, W.A. Ducker, J.N. Israelachvili: Origin and characterization of different stick-slip friction mechanisms, Langmuir **12**, 4559–4563 (1996)

30.231 B.N.J. Persson: Theory of friction – Dynamical phase transitions in adsorbed layers, J. Chem. Phys. **103**, 3849–3860 (1995)

30.232 B.N.J. Persson, E. Tosatti: Layering transition in confined molecular thin films – Nucleation and growth, Phys. Rev. B **50**, 5590–5599 (1994)

30.233 H. Yoshizawa, J. Israelachvili: Fundamental mechanisms of interfacial friction. 2. Stick-slip friction of spherical and chain molecules, J. Phys. Chem. **97**, 11300–11313 (1993)

30.234 B.N.J. Persson: Theory of friction: Friction dynamics for boundary lubricated surfaces, Phys. Rev. B **55**, 8004–8012 (1997)

30.235 P.A. Thompson, M.O. Robbins: Origin of stick-slip motion in boundary lubrication, Science **250**, 792–794 (1990)

30.236 U. Landman, W.D. Luedtke, J.P. Gao: Atomic-scale issues in tribology: Interfacial junctions and nano-elastohydrodynamics, Langmuir **12**, 4514–4528 (1996)

30.237 T. Kreer, M.H. Müser, K. Binder, J. Klein: Frictional drag mechanisms between polymer-bearing surfaces, Langmuir **17**, 7804–7813 (2001)

30.238 T. Kreer, K. Binder, M.H. Müser: Friction between polymer brushes in good solvent conditions: Steady-state sliding versus transient behavior, Langmuir **19**, 7551–7559 (2003)

30.239 E. Manias, G. Hadziioannou, G. ten Brinke: Inhomogeneities in sheared ultrathin lubricating films, Langmuir **12**, 4587–4593 (1996)

30.240 R.M. Overney, T. Bonner, E. Meyer, M. Reutschi, R. Luthi, L. Howald, J. Frommer, H.J. Guntherodt, M. Fujihara, H. Takano: Elasticity, wear, and friction properties of thin organic films observed with atomic-force microscopy, J. Vac. Sci. Technol. B **12**, 1973–1976 (1994)

30.241 R.M. Overney, E. Meyer, J. Frommer, D. Brodbeck, R. Luthi, L. Howald, H.J. Guntherodt, M. Fujihira, H. Takano, Y. Gotoh: Friction measurements on phase-separated thin-films with a modified atomic force microscope, Nature **359**, 133–135 (1992)

30.242 R.M. Overney, E. Meyer, J. Frommer, H.J. Guntherodt, M. Fujihira, H. Takano, Y. Gotoh: Force microscopy study of friction and elastic compliance of phase-separated organic thin-films, Langmuir **10**, 1281–1286 (1994)

30.243 H.I. Kim, T. Koini, T.R. Lee, S.S. Perry: Systematic studies of the frictional properties of fluorinated monolayers with atomic force microscopy: Comparison of CF_3- and CH_3-terminated films, Langmuir **13**, 7192–7196 (1997)

30.244 M. GarciaParajo, C. Longo, J. Servat, P. Gorostiza, F. Sanz: Nanotribological properties of octadecyltrichlorosilane self-assembled ultrathin films studied by atomic force microscopy: Contact and tapping modes, Langmuir **13**, 2333–2339 (1997)

30.245 R.M. Overney, H. Takano, M. Fujihira, E. Meyer, H.J. Guntherodt: Wear, friction and sliding speed correlations on Langmuir–Blodgett films observed by atomic force microscopy, Thin Solid Films **240**, 105–109 (1994)

30.246 P.T. Mikulski, J.A. Harrison: Periodicities in the properties associated with the friction of model self-assembled monolayers, Tribol. Lett. **10**, 29–35 (2001)

30.247 P.T. Mikulski, J.A. Harrison: Packing density effects on the friction of n-alkane monolayers, J. Am. Chem. Soc. **123**, 6873–6881 (2001)

30.248 E. Barrena, C. Ocal, M. Salmeron: A comparative AFM study of the structural and frictional properties of mixed and single component films of alkanethiols on Au(111), Surf. Sci. **482**, 1216–1221 (2001)

30.249 Y.-S. Shon, S. Lee, R. Colorado, S.S. Perry, T.R. Lee: Spiroalkanedithiol-based SAMS reveal unique insight into the wettabilities and frictional properties of organic thin films, J. Am. Chem. Soc. **122**, 7556–7563 (2000)

30.250 P.T. Mikulski, G. Gao, G.M. Chateauneuf, J.A. Harrison: Contact forces at the sliding interface: Mixed versus pure model alkane monolayers, J. Chem. Phys. **122**, 024701 (2005)
30.251 S. Lee, Y.S. Shon, R. Colorado, R.L. Guenard, T.R. Lee, S.S. Perry: The influence of packing densities, surface order on the frictional properties of alkanethiol self-assembled monolayers (SAMs) on gold: A comparison of SAMs derived from normal and spiroalkanedithiols, Langmuir **16**, 2220–2224 (2000)
30.252 P.T. Mikulski, L.A. Herman, J.A. Harrison: Odd and even model self-assembled monolayers: Links between friction and structure, Langmuir **21**(26), 12197–12206 (2005)
30.253 S.S. Wong, H. Takano, M.D. Porter: Mapping orientation differences of terminal functional groups by friction force microscopy, Anal. Chem. **70**(24), 5209–5212 (1998)
30.254 G.M. Chateauneuf, P.T. Mikulski, G.T. Gao, J.A. Harrison: Compression- and shear-induced polymerization in model diacetylene-containing monolayers, J. Phys. Chem. B **108**, 16626–16635 (2004)
30.255 L. Zhang, S. Jiang: Molecular simulation study of nanoscale friction for alkyl monolayers on Si(111), J. Chem. Phys. **117**, 1804–1811 (2002)
30.256 L.Z. Zhang, Y.S. Leng, S.Y. Jiang: Tip-based hybrid simulation study of frictional properties of self-assembled monolayers: Effects of chain length, terminal group, scan direction, and scan velocity, Langmuir **19**, 9742–9747 (2003)
30.257 M. Chandross, E.B. Webb III, M.J. Stevens, G.S. Grest, S.H. Garofalini: Systematic study of the effect of disorder on nanotribology of self-assembled monolayers, Phys. Rev. Lett. **93**, 166103 (2004)
30.258 M. Chandross, G.S. Grest, M.J. Stevens: Friction between alkylsilane monolayers: Molecular simulation of ordered monolayers, Langmuir **18**, 8392–8399 (2002)
30.259 M. Chandross, C.D. Lorenz, M.J. Stevens, G.S. Grest: Simulations of nanotribology with realistic probe tip models, Langmuir **24**(4), 1240–1246 (2008)
30.260 D.L. Irving, D.W. Brenner: Diffusion on a self-assembled monolayer: Molecular modeling of a bound plus mobile lubricant, J. Phys. Chem. B **110**(31), 15426–15431 (2006)
30.261 D.W. Brenner, D.L. Irving, A.I. Kingon, J. Krim: Multiscale analysis of liquid lubrication trends from industrial machines to micro-electrical-mechanical systems, Langmuir **23**(18), 9253–9257 (2007)
30.262 I. Jang, D.L. Burris, P.L. Dickrell, P.R. Barry, C. Santos, S.S. Perry, S.R. Phillpot, S.B. Sinnott, W.G. Sawyer: Sliding orientation effects on the tribological properties of polytetrafluoroethylene, J. Appl. Phys. **102**(12), 123509 (2007)
30.263 S.J. Heo, I. Jang, P.R. Barry, S.R. Phillpot, S.S. Perry, W.G. Sawyer, S.B. Sinnott: Effect of the sliding orientation on the tribological properties of polyethylene in molecular dynamics simulations, J. Appl. Phys. **103**(8), 083502 (2008)
30.264 S.B. Sinnott, R. Andrews: Carbon nanotubes: Synthesis, properties and applications, Crit. Rev. Solid State Mater. Sci. **26**, 145–249 (2001)
30.265 B. Bhushan, B.K. Gupta, G.W. Van Cleef, C. Capp, J.V. Coe: Sublimed C_{60} films for tribology, Appl. Phys. Lett. **62**, 3253–3255 (1993)
30.266 T. Thundat, R.J. Warmack, D. Ding, R.N. Compton: Atomic force microscope investigation of C_{60} adsorbed on silicon and mica, Appl. Phys. Lett. **63**, 891–893 (1993)
30.267 C.M. Mate: Nanotribology studies of carbon surfaces by force microscopy, Wear **168**, 17–20 (1993)
30.268 R. Lüthi, E. Meyer, H. Haefke: Sled-type motion on the nanometer scale: Determination of dissipation and cohesive energies of C_{60}, Science **266**, 1979–1981 (1993)
30.269 R. Lüthi, H. Haefke, E. Meyer, L. Howald, H.-P. Lang, G. Gerth, H.J. Güntherodt: Frictional and atomic-scale study of C_{60} thin films by scanning force microscopy, Z. Phys. B **95**, 1–3 (1994)
30.270 Q.-J. Xue, X.-S. Zhang, F.-Y. Yan: Study of the structural transformations of C_{60}/C_{70} crystals during friction, Chin. Sci. Bull. **39**, 819–822 (1994)
30.271 W. Allers, U.D. Schwarz, G. Gensterblum, R. Wiesendanger: Low-load friction behavior of epitaxial C_{60} monolayers, Z. Phys. B **99**, 1–2 (1995)
30.272 U.D. Schwarz, W. Allers, G. Gensterblum, R. Wiesendanger: Low-load friction behavior of epitaxial C_{60} monolayers under Hertzian contact, Phys. Rev. B **52**, 14976–14984 (1995)
30.273 J. Ruan, B. Bhushan: Nanoindentation studies of sublimed fullerene films using atomic force microscopy, J. Mater. Res. **8**, 3019–3022 (1996)
30.274 U.D. Schwarz, O. Zworner, P. Koster, R. Wiesendanger: Quantitative analysis of the frictional properties of solid materials at low loads. I. Carbon compounds, Phys. Rev. B **56**, 6987–6996 (1997)
30.275 S. Okita, M. Ishikawa, K. Miura: Nanotribological behavior of C_{60} films at an extremely low load, Surf. Sci. **442**, L959–L963 (1999)
30.276 S. Okita, K. Miura: Molecular arrangement in C_{60} and C_{70} films on graphite and their nanotribological behavior, Nano Lett. **1**, 101–103 (2001)
30.277 K. Miura, S. Kamiya, N. Sasaki: C_{60} molecular bearings, Phys. Rev. Lett. **90**, 055509 (2003)
30.278 A. Buldum, J.P. Lu: Atomic scale sliding and rolling of carbon nanotubes, Phys. Rev. Lett. **83**, 5050–5053 (1999)
30.279 M.R. Falvo, R.M. Taylor, A. Helser, V. Chi, F.P. Brooks, S. Washburn, R. Superfine: Nanometer-scale rolling and sliding of carbon nanotubes, Nature **397**, 236–238 (1999)
30.280 M.R. Falvo, J. Steele, R.M.T. Taylor II, R. Superfine: Gearlike rolling motion mediated by commensurate contact: Carbon nanotubes on HOPG, Phys. Rev. B **62**, R10664–R10667 (2000)
30.281 J.D. Schall, D.W. Brenner: Molecular dynamics simulations of carbon nanotube rolling and slid-

ing on graphite, Mol. Simul. **25**, 73–80 (2000)
30.282 B. Ni, S.B. Sinnott: Tribological properties of carbon nanotube bundles, Surf. Sci. **487**, 87–96 (2001)
30.283 B. Ni, S.B. Sinnott: Mechanical and tribological properties of carbon nanotubes investigated with atomistic simulations, Nanotubes and related materials. In: *Nanotubes and Related Materials, MRS Proc.*, Vol. 633, ed. by A.M. Rao (Materials Research Society, Pittsburgh 2001) pp. A17.13.11–A17.13.15
30.284 K. Miura, T. Takagi, S. Kamiya, T. Sahashi, M. Yamauchi: Natural rolling of zigzag multiwalled carbon nanotubes on graphite, Nano Lett. **1**, 161–163 (2001)
30.285 K. Miura, M. Ishikawa, R. Kitanishi, M. Yoshimura, K. Ueda, Y. Tatsumi, N. Minami: Bundle structure and sliding of single-walled carbon nanotubes observed by friction-force microscopy, Appl. Phys. Lett. **78**, 832–834 (2001)
30.286 P.E. Sheehan, C.M. Lieber: Nanotribology and nanofabrication of MoO_3 structures by atomic force microscopy, Science **272**, 1158–1161 (1996)
30.287 J. Wang, K.C. Rose, C.M. Lieber: Load-independent friction: MoO_3 nanocrystal lubricants, J. Phys. Chem. B **103**, 8405–8408 (1999)
30.288 Q. Ouyang, K. Okada: Nanoballbearing effect of ultra-fine particles of cluster diamond, Appl. Surf. Sci. **78**, 309–313 (1994)
30.289 R. Luthi, E. Meyer, H. Haefke, L. Howald, W. Gutmannsbauer, H.J. Guntherodt: Sled-type motion on the nanometer-scale – Determination of dissipation and cohesive energies of C_{60}, Science **266**, 1979–1981 (1994)
30.290 B. Bhushan, B.K. Gupta, G.W. Vancleef, C. Capp, J.V. Coe: Fullerene (C_{60}) films for solid lubrication, Tribol. Trans. **36**, 573–580 (1993)
30.291 U.D. Schwarz, W. Allers, G. Gensterblum, R. Wiesendanger: Low-load friction behavior of epitaxial C_{60} monolayers under Hertzian contact, Phys. Rev. B **52**, 14976–14984 (1995)
30.292 S.B. Legoas, R. Giro, D.S. Galvao: Molecular dynamics simulations of C_{60} nanobearings, Chem. Phys. Lett. **386**, 425–429 (2004)
30.293 S. Heo, S.B. Sinnott: Effect of molecular interactions on carbon nanotube friction, J. Appl. Phys. **102**(6), 064307 (2007)
30.294 P.L. Dickrell, S.B. Sinnott, D.W. Hahn, N.R. Raravikar, L.S. Schadler, P.M. Ajayan, W.G. Sawyer: Frictional anisotropy of oriented carbon nanotube surfaces, Tribol. Lett. **18**, 59–62 (2005)
30.295 G.T. Gao, P.T. Mikulski, J.A. Harrison: Molecular-scale tribology of amorphous carbon coatings: Effects of film thickness, adhesion, and long-range interactions, J. Am. Chem. Soc. **124**, 7202–7209 (2002)
30.296 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids, Part 2* (Clarendon, Oxford 1964)
30.297 G.T. Gao, P.T. Mikulski, G.M. Chateauneuf, J.A. Harrison: The effects of film structure and surface hydrogen on the properties of amorphous carbon films, J. Phys. Chem. B **107**, 11082–11090 (2003)

31. Cellular Nanomechanics

Roger D. Kamm, Jan Lammerding, Mohammad R. K. Mofrad

Numerousapplications of nanotechnology have been developed to understand mechanobiology of living cells and probe their unique mechanical properties in health and disease. In addition, since biological materials exhibit such a wide spectrum of properties, they offer new concepts for nonbiological biomimetic applications. In this chapter, the biomechanical properties of a cell and its subcellular compartments are described. First, a qualitative picture is presented of the relevant building blocks: the cytoskeleton, cell membrane, nucleus, adhesive complexes, and motor proteins. Next, the various methods used to probe cellular and subcellular mechanics are described, and some of the quantitative results presented. These measurements are then discussed in the context of several theories and computational methods that have been proposed to help interpret the measurements and provide nanomechanical insight into their origin. Finally, current understanding is summarized in the context of directions for future research.

31.1 **Overview** 1069
31.1.1 The Importance of Cell Mechanics in Biology and Medicine 1069
31.1.2 Examples Drawn from Biology and Pathophysiology 1070
31.2 **Structural Components of a Cell** 1071
31.2.1 Membranes 1071
31.2.2 Cytoskeleton 1072
31.2.3 Nucleus 1075
31.2.4 Cell Contractility and Motor Proteins 1076
31.2.5 Adhesion Complexes 1077
31.3 **Experimental Methods** 1077
31.3.1 Methods of Force Application 1077
31.3.2 Rheological Properties 1081
31.3.3 Active Force Generation 1081
31.3.4 Biological Responses 1082
31.3.5 Nonlinear Effects 1082
31.3.6 Homogeneity and Anisotropy 1082
31.4 **Theoretical and Computational Descriptions** 1082
31.4.1 Continuum Models 1083
31.4.2 Biopolymer Models 1084
31.4.3 Cellular Solids 1085
31.4.4 Tensegrity 1086
31.5 **Mechanics of Subcellular Structures** 1086
31.5.1 Cell–Cell and Cell–Matrix Adhesions 1086
31.5.2 Cell Membranes 1088
31.5.3 Cell Nuclei 1090
31.5.4 Mechanosensing Proteins 1091
31.6 **Current Understanding and Future Needs** 1095
References 1096

31.1 Overview

31.1.1 The Importance of Cell Mechanics in Biology and Medicine

All living things, despite their profound diversity, share a common architectural building block: the cell. Cells are the basic functional units of life, yet they are themselves composed of numerous components with distinct mechanical characteristics. To perform their various functions, cells undergo or control a large range of intra- and extracellular events, many of which involve mechanical phenomena or may be guided by the forces experienced by the cell. These micro- and nanomechanical phenomena range from macroscopic events such as the maintenance of cell shape, motility, adhesion, and deformation, to microscopic events such as how cells sense mechanical signals and transduce them into a cascade of biochemical signals, ultimately leading to a host of essential biological responses.

Cell mechanics plays a major role in biology and physiology. The ability of a cell to perform its function often depends on its shape, and shape is maintained through structural stiffness. In the blood circulation system, red blood cells, or erythrocytes, exist in the form of biconcave disks that are easily deformed to help

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_31

facilitate their flow through the microcirculation and have a relatively large surface-to-area ratio to enhance gas exchange. White cells, or leukocytes, are spherical, enabling them to roll along the vascular endothelium before adhering and migrating into the tissue. Because their diameter is larger than some of the capillaries they pass through, leukocytes maintain excess membrane in the form of ruffles or microvilli so that they can elongate at nearly constant volume and not obstruct the microcirculation. In other organs, cells also serve specialized functions; For example, in the lung, airway epithelial cells are covered with a bed of cilia, finger-like cell extensions that propel mucus along the pulmonary airways. Lastly, the cytoskeletal structure in skeletal or cardiac muscle cells is specifically organized to actively generate forces and to sustain large strains. In each of these examples, the internal structure of the cell along with the cell membrane provide the structural integrity to maintain the particular shape needed by the cell to accomplish its function, although the specific components of the structure are highly variable and diverse.

Cell mechanics also plays an important role in cell migration. Migration is critical during early development, but also in the fully differentiated organism, e.g., in wound repair when cells from the surrounding, undamaged tissue migrate into the wound, in angiogenesis (i. e., the generation of new blood vessels), or in combating infection when cells of the immune system transmigrate from the vascular system across the vessel wall and into the infected tissues. In cancer, cell migration and invasion is a crucial step of metastasis, i. e., the spreading of cells from the primary tumor to distant tissues and organs. Regardless of the specific context, cell migration occurs in several stages and cellular mechanics and dynamics are critical modulators at each of these steps.

The importance of cell mechanics in biology is most apparent in mechanotransduction, i. e., the ability of the cell to sense and respond to externally applied forces, and these phenomena form the basis for the field that is now recognized as *mechanobiology*. Most cells are able to sense when a physical force is applied to the cell, and respond through a variety of biological pathways leading to such diverse effects as changes in membrane channel activity, up- or down-regulation of gene expression, alterations in protein synthesis, and altered cell morphology. The signaling cascades that become activated as a consequence of mechanical stress have generally been well characterized. However, the initiating processes by which cells convert the applied force into a biochemical signal are less well understood, and researchers have just begun to unravel some of these fundamental mechanisms. Stretch-activated ion channels or strain-induced activation of G proteins are two recognized means of mechanotransduction. Similarly, disruption of microtubules or conformational changes of cytoskeletal proteins that alter their binding affinities have been demonstrated as cellular mechanosensors. Others have focused on the role of the glycocalyx, a layer of carbohydrate-rich proteins on the cell surface, in the response of endothelial cells to fluid shear stress. Another potential mechanism is that forced deformations within the nucleus could directly alter transcription or transcription factor accessibility. Constrained autocrine signaling is yet another mechanism whereby the strength of autocrine signaling is regulated by changes in the volume of extracellular compartments into which the receptor ligands are shed. Changing this volume by mechanical deformation of the tissues can increase the level of autocrine signaling. Finally, others have demonstrated that conformational changes in intracellular proteins in the force transmission pathway connecting the extracellular matrix with the cytoskeleton through focal adhesions (FAs) are a key mechanotransduction mechanism. Understanding of these fundamental cellular functions has led to a general appreciation of how signals can be transmitted via parallel chemical and mechanical pathways, and that communication between these two is bidirectional. For reviews of this topic, see [31.1–6].

31.1.2 Examples Drawn from Biology and Pathophysiology

Many studies during the past two decades have shed light on a wide range of cellular responses to mechanical stimulation. It is now widely accepted that stresses experienced in vivo are instrumental in a wide spectrum of pathologies (reviewed in [31.5]). One of the first diseases found to be linked to cellular stress was atherosclerosis, where it was demonstrated that hemodynamic shear influences endothelial function, and that conditions of low or oscillatory shear stress are conducive to the formation and growth of atherosclerotic lesions. Even before then, the role of mechanical stress in bone growth and healing was widely recognized, and since then, many other stress-influenced cell functions have been identified. Hair cells in the inner ear translate acoustic vibrations into biochemical signals and thereby enable hearing; conversely, leftward fluid flow generated by oscillating cilia (hair-like extensions) of epithelial cells during gestation is responsible for the normal left–right patterning in early mammalian embryos, and mutations in molecular motors that abolish cilia motion and directed fluid flow result in Kartagener's syndrome, in which left–right patterning becomes random. Mechanotransduction of muscle cells

is pivotal to exercise-induced muscle growth (hypertrophy), and defects – caused for example by mutations in mechanosensitive proteins – can result in muscular dystrophies. Asthma is another particularly salient example where the epithelial cells lining the airways are subjected to stresses as the pulmonary airways constrict during breathing, and airway wall remodeling. Further knowledge of the mechanisms by which cells respond to such forces could enhance our understanding of these diseases.

31.2 Structural Components of a Cell

Aside from providing the outer boundary of a cell and nucleus, membranes enclose numerous intracellular structures as well, and the discussion in this section pertains to all of these.

31.2.1 Membranes

The term *membrane* generally refers to the phospholipid bilayer and the proteins associated with it. The phospholipids contained in the membrane are arranged in two layers, or *leaflets*, with their hydrophobic tails pointing inward and their hydrophilic heads outward. Together, they constitute a bilayer $\approx$6 nm in thickness (Fig. 31.1). In addition to the phospholipids, the membrane contains glycolipids and cholesterol. While the amount of glycolipids is small, constituting only $\approx$2% of the total lipid content, cholesterol is a major plasma membrane constituent, $\approx$20% by weight, a value that remains quite constant among different cell types. The specific membrane composition is critical for determining membrane structural integrity; for example, both the bending stiffness and the viscosity of the lipid bilayer are strongly dependent on the cholesterol content. Membrane composition, and thus mechanical and chemical properties, differ between the plasma membrane surrounding the cell and the membranes encapsulating intracellular organelles, such as the nucleus. Thus, it is possible to select experimental conditions (e.g., mild detergents such as digitonin) that disrupt the plasma membrane while leaving intracellular membranes intact. Cellular membranes also contain many membrane-associated proteins, which account for $\approx$50% of the membrane by weight, but because of their relatively large molecular weight, only $\approx$1–2% of the number of membrane molecules. Membrane proteins serve a variety of functions, ranging from signaling to transport of ions and other molecules across the membrane to cell–cell and cell–matrix adhesion.

The plasma membrane has associated macromolecular structures on both intra- and extracellular sides, giving rise to a three-layer composite construction. On the intracellular side, the membrane is physically attached to a cortex or the cytoskeleton. The cortex is a dense, filamentous structure that lends stiffness to the membrane, and can also interact with various transmembrane proteins, often impeding their free diffusion either by steric interactions or direct chemical bonding. In some cells, the cortex is simply a region of dense cytoskeletal matrix in the vicinity of the bilayer. In others, it exhibits a distinctly different structure or composition; for example, erythrocytes possess a cortex composed of a network of spectrin tetramers linked by actin filaments. This network is attached to the membrane by ankyrin and the integral membrane protein band 3. This spectrin network accounts for much of the bending stiffness exhibited by the red cell membrane.

Most cells are coated by a *glycocalyx*, which has been shown to extend as far as 0.5 μm from the surface of endothelial cells, where it forms a compressible barrier separating circulating erythrocytes and leukocytes from the endothelial membrane. The glycocalyx is composed of short oligosaccharide chains, glycoproteins, glycolipids, and high-molecular-weight proteoglycans, all organized into an interconnected network with an overall negative charge. Although its function is not

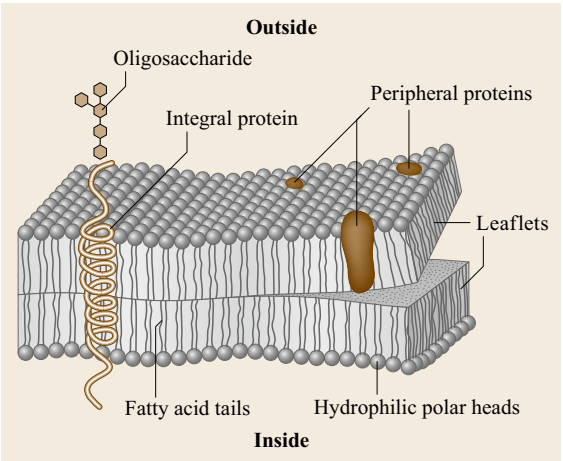


Fig. 31.1 A model of the lipid bilayer, showing the hydrophilic heads (polar head groups) on the exterior surfaces and the hydrophobic tails (fatty acid tails) pointing in. Also shown are examples of transmembrane or integral membrane proteins. The total thickness of the bilayer is $\approx$ 6 nm (after [31.7])

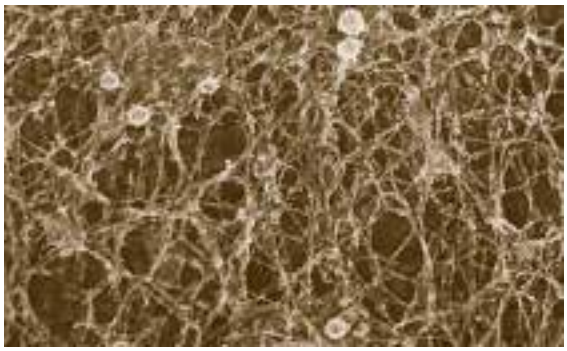


Fig. 31.2 The cytoskeleton of a macrophage lamellipodium as seen by electron microscopy. The fibrous structure is mainly composed of actin filaments (© John Hartwick, Harvard Medical School)

completely understood, it apparently plays a role in macromolecular transport across the endothelium and is an important factor in the interaction between bloodborne cells and the endothelium. Studies have demonstrated that the glycocalyx in a capillary is readily compressed by a passing leukocyte, yet is sufficiently rigid to prevent flowing erythrocytes from approaching the endothelial surface. Recent studies demonstrate that steric interference of the glycocalyx, which is substantially taller than transmembrane cell surface receptors, can also modulate cell adhesion and cell surface receptor clustering [31.3, 8]. These findings are particularly relevant in cancer invasion and metastasis, since malignant cancer cells often have altered expression of glycans and glycoproteins that results in a bulkier glycocalyx, which promotes integrin clustering and focal adhesion assembly [31.3].

31.2.2 Cytoskeleton

The cytoskeleton is the network of biopolymers that permeates the cell and largely accounts for its structural integrity. At high magnification, this network appears to be composed of several distinct types of intertwined filaments with a variety of interconnections, as can be seen in the micrograph in Fig. 31.2. The apparent stiffness of the network, as that of other fibrous materials, depends fundamentally upon the elastic properties of the constituent fibers, but also on the level of prestress generated by actomyosin contractility.

The cytoskeletal matrix is primarily composed of three constituents: actin microfilaments ($\approx 5-9$ nm in diameter), microtubules (≈ 24 nm), and intermediate filaments (≈ 10 nm) (Table 31.1). These form a complex interconnected network that exists in a state of constant flux, especially when the cell is dividing, migrating or undergoing other dynamic processes. All are polymers built from protein subunits, held together by noncovalent bonds. They also share the common feature that they cross-link, often with the aid of other proteins, to form bundles and lattice networks. The higher-order structures formed by these polymers, as we will see, are critical determinants of cytoskeletal elasticity and can vary substantially between different cells.

Actin Microfilaments

Actin filaments play an essential role in virtually all types of motility. Actin–myosin interactions are of obvious importance in muscle, but are also instrumental in the migration and movement of nonmuscle cells. Actin polymerization in combination with cytosolic pressure is one of the critical factors initiating cell migration through the formation of filopodia or lamellipodia [31.5, 10].

Actin is one of the most prevalent proteins found in the cell, ranging in concentration from 1–10% by weight of total cell protein in nonmuscle cells, to 10–20% in muscle. Molecular actin is composed of 375 amino acids (molecular weight 43 kDa) and is found in at least six isoforms that differ from each other only slightly. Of these, four are found in muscle and the other two in nonmuscle cells. Actin exists in either globular form (G-actin monomers) or filamentous form (F-actin polymers), with the balance between the two being a highly dynamic process that is finely regulated by a variety of different factors. F-actin is a long, flexible filament, $\approx 5-9$ nm in diameter. Subunits (monomers) are organized into a double-stranded helix having structural and functional polarity (pointed or negative, and

Table 31.1 The main constituents of the cytoskeleton and their mechanical properties. Note that the effective Young's modulus, estimated using the diameter and bending stiffness values for actin filaments and microtubules, is approximately the value that would be predicted on the basis of van der Waals attraction between two surfaces [31.9]. Recall that the persistence length and bending stiffness are related through the expression $l_p = K_b/k_B T$, and that $K_b = EI = \pi/4a^4E$ for a rod of circular cross-section with radius a

	Diameter $2a$ (nm)	Persistence length l_p (μm)	Bending stiffness K_b (N m^2)	Young's modulus E (Pa)
Actin filaments	5–9	17	7×10^{-26}	$1.3–2.5\times10^9$
Microtubules	25	1000–5000	2.6×10^{-23}	1.9×10^9
Intermediate filaments	10	1–3	$4–12\times10^{-27}$	$1–3\times10^9$

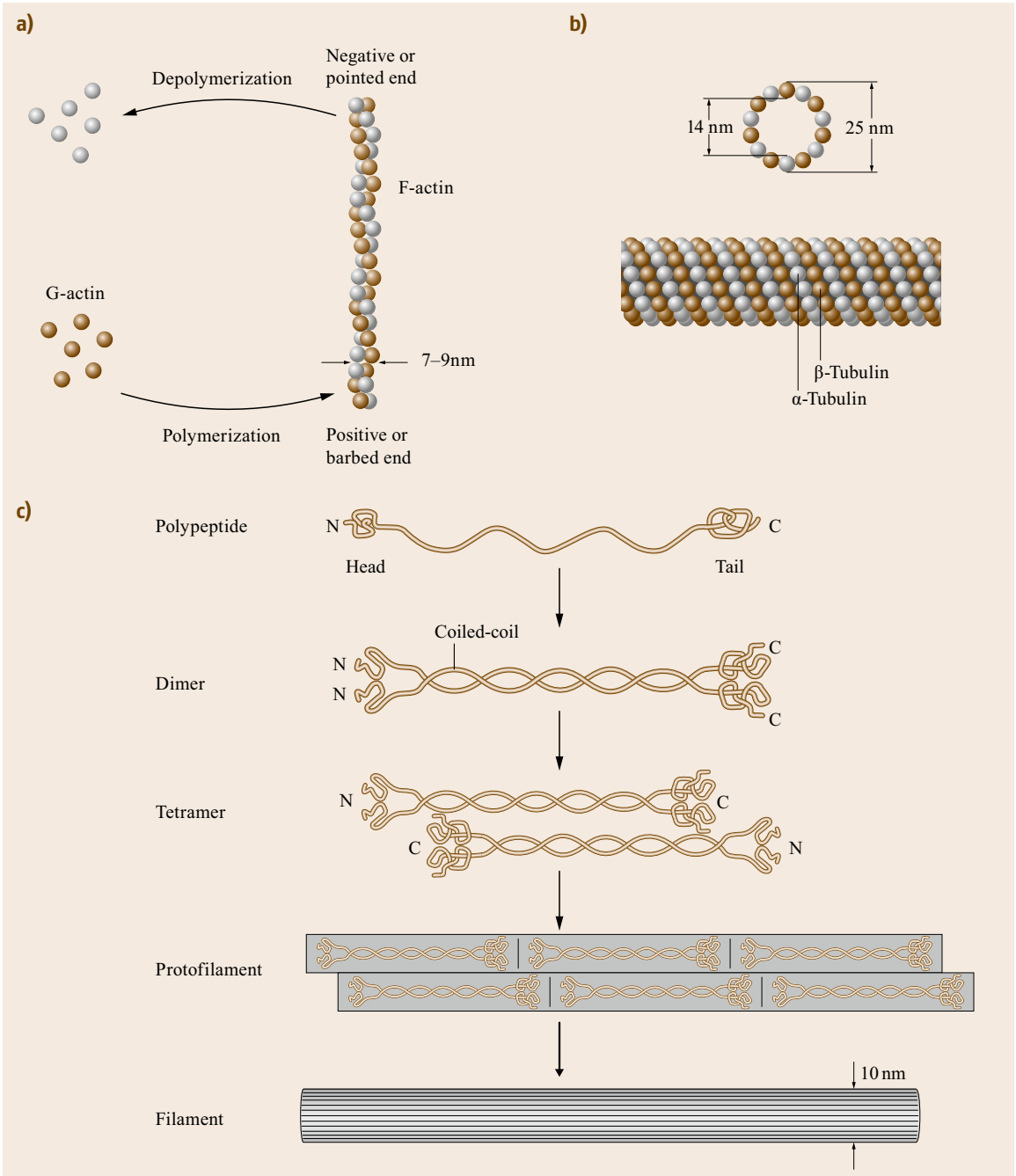


Fig. 31.3 **(a)** Schematic showing the polymerization of G-actin monomers to form F-actin. **(b)** α- and β-tubulin organized into microtubules. **(c)** Organizational structure of a cytoplasmic intermediate filament

barbed or positive ends) and a half-pitch of $\approx 37\,\text{nm}$ (Fig. 31.3a). At the barbed end, an adenosine triphosphate (ATP) binding cleft is exposed, allowing for binding of monomer and linear growth of the filament. ATP binding and hydrolysis play a critical role in regulating actin dynamics and controlling the length of the actin filament. In addition, variations in the local levels of specific actin binding proteins such as Arp2/3, formin, or gelsolin can selectively promote actin polymerization, induce actin filament branching, sever actin

filaments, or cap actin filaments at the pointed or barbed end.

The F-actin filaments can further organize into quaternary structures such as bundles or a lattice network with the aid of *actin binding proteins* (ABPs). The bundles, also referred to as *stress fibers*, are closely packed parallel arrays of filaments, connected to each other by several members of the ABP family (e.g., α-actinin, fascin, and scruin) and molecular motors of the nonmuscle myosin II family, which give the stress fibers their contractile properties. Stress fibers can vary in size, but are typically several hundred nanometers in diameter. Actin stress fibers tend to form when the cell requires additional strength, such as in endothelial cells, in response to an elevated shear stress, or in migrating fibroblasts. These fibers often concentrate around and attach to focal adhesion sites, and are therefore critical to cell adhesion. Actin networks consist of an interconnected matrix of F-actin filaments, the junctions of which are often seen to be nearly orthogonal. At least two distinct types of network are observed – cortical (membrane-associated and more planar in nature) and non-membrane-associated, which possess a more isotropic three-dimensional structure. The formation of bundles and networks is facilitated by a variety of cross-linking proteins such as filamin, which forms a V-shaped polymer that connects two actin filaments nearly at right angles.

The elastic properties of actin filaments have been measured in a variety of ways: by axial stretch [31.11], twisting [31.12, 13], and bending [31.14]. By all methods, single actin filaments were found to exhibit a Young's modulus in the range of $(1.3–2.5) \times 10^9\,\mathrm{N/m^2}$. This range compares favorably with that measured for silk and collagen [31.15] and is also roughly consistent with predictions based on van der Waals bonding between surfaces [31.9].

Intermediate Filaments

Intermediate filaments, which form ≈ 10 nm-diameter fibers in vitro and in vivo, are much less studied than actin and not as well characterized. Perhaps this is due to the fact that there is no single molecular constituent that composes the intermediate filaments (IFs); instead, there are more than 50 different IF genes that have been identified. Expression of specific IFs is very cell type and tissue specific, such as expression of vimentin in fibroblasts and many mesenchymal cells, or desmin in striated muscle cells.

Intermediate filaments constitute ≈ 1% of total protein in most cells, but can account for up to 85% in cells such as epidermal keratinocytes and neurons [31.16]. While they come in many varieties, they share a common structural organization. All have a large central α-helical rod domain flanked by amino- and carboxy-terminal domains. Assembly occurs by the formation of dimers into a coiled coil structure. For cytoplasmic IFs, the dimers assemble in a staggered antiparallel array to form tetramers that connect end to end to form apolar protofilaments. These protofilaments assemble into a rope-like structure containing ≈ 8 subunits each (Fig. 31.3c) [31.16, 17]. Although intermediate filaments are more stable than microfilaments, they can be modified by phosphorylation. Intermediate filaments exhibit a lower bending stiffness than either microfilaments or microtubules, as evidenced by their persistence length of only 1–3 μm. Linker proteins such as bullous pemphigoid antigen 1 (BPAG1) and plectin contain both actin and IF binding domains, providing a means by which these networks can be linked. Evidence also exists for plectin binding to microtubules.

Intermediate filaments are often found surrounding the nucleus and extending outward to the plasma membrane. In epithelial cells, keratin filaments connect to the plasma membrane at desmosomes and hemidesmosomes and help them to withstand mechanical stress. In addition to these cytoplasmic IFs, multicellular organisms also express nuclear IFs, lamins, which form a dense protein network underneath the inner nuclear membrane and are described in more detail in Sect. 31.2.3.

Microtubules

Microtubules are important in determining cell shape, and they play a critical role in separating chromosomes during mitosis. Microtubules are central to the motion of cilia and flagella. Compared with either microfilaments or intermediate filaments, microtubules are rigid structures, but exist in a dynamic equilibrium, much as do microfilaments. Microtubules take the form of hollow cylinders with ≈ 25 nm outer diameter and 14 nm inner diameter. The tubular structures are composed of tubulin, a globular dimer consisting of two 55 kDa polypeptides, α- and β-tubulin. The dimers polymerize to form microtubules that consist of 13 linear protofilaments forming a hollow-cored cylinder (Fig. 31.3b). The filaments are polar, having a rapidly growing end and a slowly growing end, mediated by hydrolysis of guanosine triphosphate (GTP) after polymerization. If hydrolysis occurs too quickly, before new GTP-bound tubulin can bind to the end, the microtubule might disassemble. Depending on the rate of hydrolysis and rate of GTP-bound tubulin addition to the end, the microtubule can either grow or shrink [31.18]. In fact, free ends tend to alternate between periods of steady growth and disassembly in a stochastic manner.

During interphase, microtubules tend to be anchored at their negative ends to the centrosome, located

near the nucleus. From there they extend to all parts of the cell, suggesting a strong role in maintaining the structural integrity of the cell and providing a cellular *highway* system for cargo transported by motor proteins such as the kinesins and dyneins. In many of these situations, the microtubule is thought to play a structural role that relies on its having a high bending stiffness, which was found to be of the order of $2.6\times10^{-23}\,\mathrm{N\,m^2}$ [31.15].

31.2.3 Nucleus

The nucleus is the distinguishing feature of eukaryotic cells and directs and controls deoxyribonucleic acid (DNA) replication, ribonucleic acid (RNA) transcription and processing, and ribosome assembly [31.19]. Most eukaryotic cells contain a single nucleus. However, some specialized cells (e.g., red blood cells) become anucleate during their maturation, while other cells such as skeletal and cardiac muscle cells can become multinucleated due to cell fusion. With a diameter in the range of $\approx 5-20\,\mu\mathrm{m}$, the nucleus is the largest cellular organelle. The nucleus is separated from the cytoplasm by the nuclear envelope, which consists of two lipid bilayers, the inner and outer nuclear membrane, nuclear pores, and the underlying nuclear lamina, a dense protein network consisting mostly of lamin proteins that control the nuclear shape, size, and stability (Fig. 31.4). The outer nuclear membrane is continuous with the endoplasmic reticulum and connects to the inner nuclear membrane at the nuclear pores, thus enclosing the perinuclear space. Nuclei typically contain a few thousand nuclear pores, composed of hundreds of proteins that form the nuclear pore complex that controls transport between the nucleus and the cytoplasm [31.20, 21].

The nuclear interior contains the packaged DNA in the form of chromatin as well as diverse intranuclear compartments referred to as subnuclear bodies that include nucleoli, Cajal bodies, and promyelocytic leukemia (PML) bodies. These subnuclear bodies are not surrounded by membranes but self-organize through processes only incompletely understood. Chromatin can be organized into two distinct forms, the dense and transcriptionally silent heterochromatin often located at the nuclear periphery, and the typically gene-rich and transcriptionally active euchromatin. Chromatin is composed of $\approx 30\,\mathrm{nm}$ fibers that arise from regular wrapping of DNA around histone octamers to form nucleosomes resembling beads on a string and subsequent further compaction facilitated through the linker histone H1. This process allows the $\approx 2\,\mathrm{m}$ of human DNA to be packaged into a nucleus a few micrometers in diameter. The chromatin fibers in turn form higher-order structures such as euchromatin and heterochromatin, which can be dynamically regulated by biochemical modification (e.g., by methylation, acetylation or phosphorylation) of histone proteins and DNA. Furthermore, chromatin fibers from single chromosomes often form distinct and nonoverlapping chromosome territories. In addition to chromatin and nuclear bodies, the nuclear interior also contains several structural proteins, including nuclear actins, myosin, spectrin, and nucleoplasmic lamins A and C [31.22–26]. However, it remains unclear what intranuclear structures (often referred to as nuclear matrix) these proteins form within the nuclear interior and how such structures could contribute to nuclear processes such as transcription, and DNA damage repair.

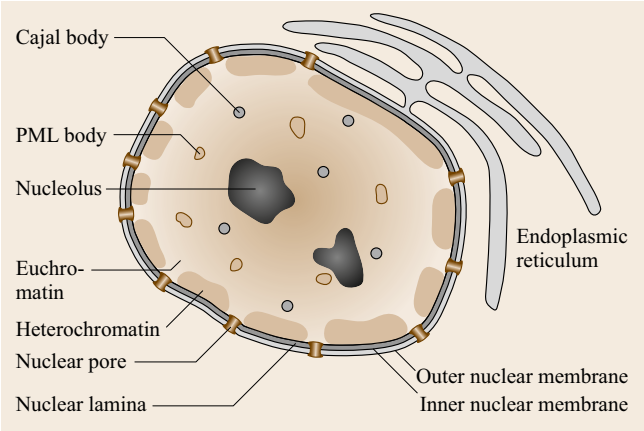


Fig. 31.4 Schematic drawing of a mammalian cell nucleus. The nucleus is surrounded by the inner and outer nuclear membranes and the underlying nuclear lamina. Nuclear pores allow transport between the nuclear interior and the cytoplasm. The nuclear interior consists of densely packed heterochromatin, mostly located at the nuclear periphery, the transcriptionally active euchromatin, and several intranuclear structures such as Cajal and promyelocytic leukemia (PML) bodies and the nucleoli. Not shown are intranuclear structures formed by lamins and other proteins that constitute the nuclear matrix

Importantly, the nucleus cannot be viewed in isolation from the surrounding cytoskeleton. The molecular mechanism by which the nucleus is connected to the cytoskeleton has puzzled researchers for many years, as it was unclear how forces required for nuclear positioning and anchoring could be transmitted across the 50 nm-wide perinuclear space. Recent work in *Caenorhabditis elegans*, *Drosophila melanogaster*, yeast, and mammalian cells led to the discovery of two new families of nuclear envelope proteins that form the linker of nucleoskeleton and cytoskeleton (LINC) complex that transmits forces from the cytoskeleton across the nuclear envelope to the nuclear interior [31.27, 28]. The LINC complex (Fig. 31.5) consists of large nesprin isoforms located on the outer nuclear membrane that can bind to cytoskeletal F-actin, microtubules (via kinesin and dynein), and intermediate filaments (via plectin). At

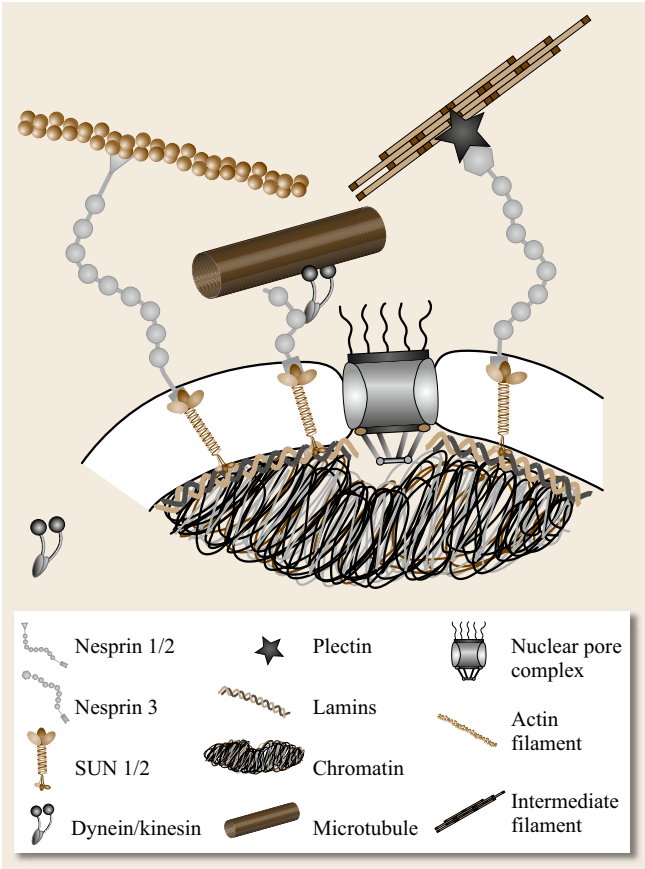


Fig. 31.5 Nucleocytoskeletal coupling across the nuclear envelope. Schematic drawing of the current model of force transmission between the cytoskeleton and nucleus across the LINC complex in mammalian somatic cells. The largest nesprin-1 and nesprin-2 isoforms contain an N-terminal actin-binding domain that can interact with cytoskeletal actin filaments. Shorter isoforms of nesprin 1 and 2 do not contain an actin-binding domain, but may still interact with other cytoskeletal elements through their large spectrin repeat regions (*brown and grey spheres*). Nesprins-1, -2, and -4 can also interact with microtubules via dynein and/or kinesin molecular motors. Nesprin-3 can directly bind to plectin, which can associate with intermediate filaments. Nesprins interact with trimers of the inner nuclear membrane proteins SUN1 and SUN2 across the perinuclear space through their C-terminal KASH (Klarsicht, ANC-1, Syne Homology) domain. SUN proteins can in turn bind to the nuclear lamina, nuclear pore complexes, and other inner nuclear membrane proteins. These proteins interact with chromatin and intranuclear proteins, thus completing the physical link between the cytoskeleton and the nuclear interior. However, how the interaction between SUN proteins and nesprins, or their binding to other nuclear and cytoplasmic proteins, is regulated remains a matter of active research

the same time, nesprins physically interact across the perinuclear space through their conserved C-terminal KASH (Klarsicht, ANC-1, Syne Homology) domain with Sad1p/UNC-84 (SUN) proteins, which are located at the inner nuclear membrane. There, SUN proteins can bind to lamins, chromatin, nuclear pores, and other nuclear envelope proteins, thus creating a physical link between the cytoskeleton and the nucleus [31.29, 30]. Due to this intricate coupling between the nucleus and the cytoskeleton, defects in nuclear envelope proteins can have direct effects on cytoskeletal structure and mechanics; For example, fibroblasts lacking the nuclear envelope proteins lamin A and C have reduced cytoskeletal stiffness and disturbed actin, vimentin, and microtubule organization [31.31–33]. Mutations in lamins, nesprins, and SUN proteins can result in muscular dystrophies and cardiomyopathies [31.34], in part by disturbing nucleocytoskeletal coupling and impairing nuclear stability [31.35–37].

31.2.4 Cell Contractility and Motor Proteins

All muscle cells use the molecular motor composed of actin and myosin to produce active contraction. These are arranged in a well-defined structure, the sarcomere, and the regularity of the sarcomeres gives rise to the characteristic striated pattern seen in skeletal muscle cells and cardiac myocytes. Importantly, even nonmuscle cells contain contractile machinery, which they use for a variety of functions such as maintaining cell tension, changing cell shape, and cell migration. Prominent bundles of actin filaments, called stress fibers, are contractile and house myosin filaments. The macromolecular organization of stress fibers bears similarities to sarcomeres – actin filaments are in parallel arrangement with the filament polarity alternating alongside; laterally, a space of $\approx 10\,\mathrm{nm}$ is maintained that situates parallel bipolar myosin II filaments in between. The myosin head moves toward the positive pole of the actin filaments during a power stroke, producing a net contractility for an alternating-polarity arrangement. Stress fibers (like sarcomeres) have a troupe of proteins, e.g., α-actinin, filamin, troponin, caldesmon, and tropomyosin, which control the arrangement of actin and myosin, and regulate their interaction. The formation and strengthening of stress fibers is intricately linked to the mechanism by which cells respond to their physical environment. First, stress fibers are known to form between points in the cell where actin–myosin contractility is resisted. Typically these are hot spots of protein activity known as focal adhesions, where the actin cytoskeleton is anchored to transmembrane inte-

grins, which are anchored to the matrix proteins. The focal adhesion also includes a cascade of proteins that relay biochemical signals, regulate its strengthening, and organize the cytoskeleton for the growth of stress fibers (e.g., cross-linking actin and recruiting myosin).

31.2.5 Adhesion Complexes

Adhesion complexes are collections of proteins forming a physical linkage between cytoskeleton and extracellular matrix (ECM) and between cells. A cell can use focal adhesions to gain traction on the ECM during the process of spreading and migration. The assembly of focal adhesions is a dynamical process that is closely regulated by the mechanical and chemical cues that the cell experiences. Both intracellular and extracellular mechanical stresses transmitted through focal adhesions are important in the formation of a focal adhesion complex, whereas the release of stress results in the turnover of a focal adhesion. The myosin-mediated contractile force transmitted to the ECM and the tension applied to the adhesion complex are necessary for promoting focal adhesion development. In contrast, disruption of myosin activity effectively inhibits the formation of focal adhesions. In the absence of myosin contractile force, externally applied mechanical forces can also promote the formation of focal adhesions. This force-regulated focal adhesion assembly allows a cell to probe the mechanical stiffness of its surroundings and respond accordingly, for example, by migrating in the direction of increasing substrate stiffness. Adhesion complexes also play an important role in helping tissues form by holding the cells together. The cells of most tissues are bound directly via cell–cell junctions. Cell–cell adhesion complexes are found in many different types, depending on tissue and cell type, and serve in both mechanical coupling of cells as well as intercellular transport.

31.3 Experimental Methods

31.3.1 Methods of Force Application

Measuring cellular or subcellular biomechanics often requires the application of precisely controlled forces to single or multiple cells and quantification of the induced deformation, although some techniques rely on detecting forces generated by the cells or on observing particles within the cytoplasm subjected to thermal motion. Consequently, experimental methods can be divided into active (Table 31.2; Fig. 31.6) and passive techniques. In the active techniques, applied forces are generally in the same range as physiological forces acting at the cellular and molecular level (Table 31.3), and induced displacements are on the nanometer or micrometer scale. For most methods – whether they are active or passive – the cellular deformations are detected based on computer-based image analysis of bright-field or fluorescence microscopy images, but some techniques also apply quadrant photodiode detectors for faster and higher-resolution tracking of microspheres embedded in the cytoplasm.

Active Measurements

Atomic Force Microscopy. Atomic force microscopy (AFM) – only developed in 1986 – is now routinely used in cell biology to *image* cells, measure cytoskeletal stiffness, and quantify single-molecule interactions. In an atomic force microscope, a small tip attached to a flexible cantilever is controlled with (sub)nanometer precision to carefully probe (i. e., indent) the cell surface. The deflection of the cantilever can then be used to infer the indentation depth and calculate the applied force. The tip of the cantilever typically has the shape of a pyramid (with a tip radius of $\approx 20\,\text{nm}$), but some applications use small polystyrene beads ($0.1-5\,\mu\text{m}$ diameter) attached to the cantilever or tipless probes to provide a larger contact area with the cell. The mechanical properties of the probed cell (e.g., Young's modulus or shear modulus) are inferred from the force–indentation curves, often assuming a Hertz model for a linear elastic, isotropic material [31.38], although several modifications have been proposed to account, for example, for the finite thickness of thin cell extensions such as lamellipodia [31.39] or the pericellular brush present on many cells [31.40]. These theoretical models often provide a surprisingly good fit to

Table 31.2 Methods of force application

Method	Typical force range
Atomic force microscopy	10 pN–100 nN
Microindenter	1–100 nN
Microplate stretcher	1–100 nN
Magnetic bead microrheology (twisting)	10 pN–1 nN
Magnetic bead microrheology (pulling)	100 pN–10 nN
Optical traps	1–500 pN
Micropipette aspiration	1–100 nN
Substrate strain	1–30% strain
Shear flow	1–100 Pa
MEMS devices	0.5–1500 nN

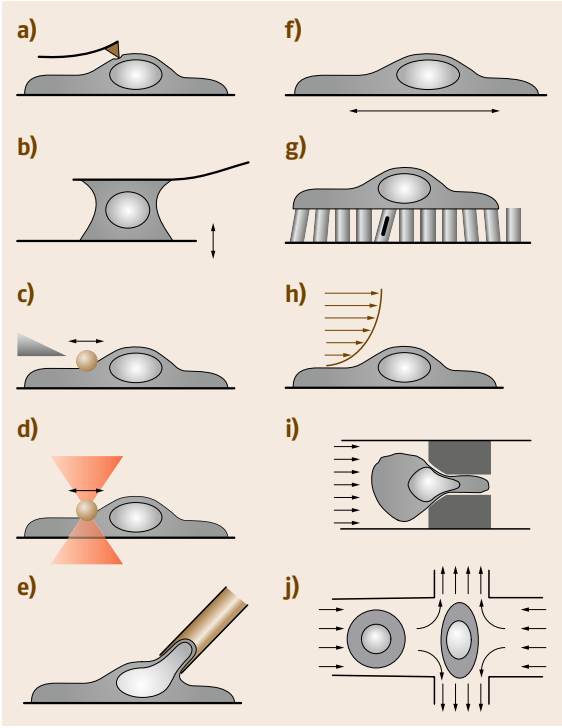


Fig. 31.6 Experimental methods for cellular force application. Overview of techniques to apply precisely controlled forces or deformations to cells for active cellular biomechanics measurements: **(a)** atomic force microscopy, **(b)** parallel microplate stretcher, **(c)** magnetic bead microrheology, **(d)** optical trap, **(e)** micropipette aspiration of adherent cell, **(f)** substrate strain experiments, **(g)** micropillars that deform (i. e., deflect) under cellular traction forces (this passive technique can be modified by embedding small magnetic particles in some of the pillars, which can then be actively manipulated to apply a highly localized force on the basal cell surface), **(h)** flow chamber for fluid shear-stress application, **(i)** perfusion of cells through microfluidic pores or constrictions (cell deformability and friction can be inferred from the transit times), and **(j)** hydrodynamic cell stretching using two opposing fluid streams

the experimental data, despite the fact that cells are nonisotropic, nonhomogeneous, and nonlinear elastic materials. One benefit of the atomic force microscope is that the usable force range spans several orders of magnitude (Table 31.2), depending on the spring constant of the selected cantilever. Therefore, AFM can also be used for single-molecule measurements, for example, to measure bond strength. In this case, the AFM probe is functionalized with low concentrations of the protein of interest and then brought into contact with the appropriate binding partner immobilized on a rigid surface.

Table 31.3 Typical force ranges in cell biology

Biological force	Force range
Force generated by motor proteins (e.g., kinesin, myosin)	$\approx 1-10$ pN
Force transmitted by protein–protein interactions (rate dependent)	$\approx 1-200$ pN
Force required for (partial) protein unfolding	≈ 100 pN
Force generated by migrating or contracting cells	≈ 1 nN–10 μN

After binding occurs, the AFM tip is then carefully retracted until the molecules dissociate, which results in a sudden drop of force. Rupture forces at different velocities can be directly inferred from AFM data.

Magnetic Twisting Cytometry. In this technique, mechanical measurements are based on the displacements of small ($\approx 0.2-5\,\mu$m diameter), ferromagnetic or paramagnetic beads attached to the cell surface and subjected to a magnetic force. Early versions of this technique [31.41, 42] used two orthogonal magnetic fields, one to magnetize the ferromagnetic particles with a brief, intense pulse, and the second one to induce a twisting, magnetic torque to the beads. The induced bead rotation is then measured by a change in the orientation of the induced magnetic field or through microscopic observations. The latter approach offers the advantage that it can detect rotation and displacement of individual beads, so that loosely attached beads that might rotate freely can be excluded from the overall measurements. In a related variant of this technique, often referred to as magnetic trap or tweezers, paramagnetic beads on the cell surface or inside the cytoplasm are manipulated by a single- or multipole electromagnet controlled through a computer. The force acting on a paramagnetic bead inside a magnetic field is given by the equation $F = \mu_0 \chi V \nabla(\boldsymbol{H} \cdot \boldsymbol{H})$, where F is the magnetic force, μ_0 is the permeability constant, χ is the volume susceptibility, V is the bead volume, and $\boldsymbol{H}$ is the external magnetic field strength. Thus, the applied force increases with increasing field strength and increasing field gradient. For a single-pole magnetic trap, the magnetic field decays rapidly with increasing distance from the tip of the pole. This results in a steep gradient and large forces near the tip that exponentially decay away from the tip, requiring careful calibration of the magnetic trap. One of the limitations of single-pole magnetic traps is the unidirectional force direction, i. e., forces can only be exerted in the direction towards the magnetic trap (pulling), but this can be overcome by using multipole magnetic traps. Another limitation is that the bead localization on the cell is random, so care must be taken to only compare results from cells with similar bead positions (e.g., on the nucleus, the nuclear

periphery, or the lamellipodia). Lastly, the induced bead rotation and displacement are strongly dependent on the bead attachment angle, i. e., how deep the magnetic bead is embedded in the cell surface, requiring careful controls or confirmation by confocal three-dimensional (3-D) reconstruction.

Optical Traps/Tweezers. This technique is similar to the magnetic trap experiments, as the induced displacement of microscopic beads (50 nm–1 μm) attached to the cell surface or inside the cytoplasm subjected to a precisely controlled force is used to determine the mechanical properties of the cytoskeleton. The major difference is that, in an optical trap, a focused laser beam is used to position and displace a bead with high refractive index on the cell, allowing precise bead manipulation in all directions. The optical trap acts as an elastic spring with a tunable spring constant, so the force applied to the bead can also be controlled with high precision (Chap. 20). Another advantage of the optical trap system is that multiple beads can be independently controlled by splitting the laser beam, so that cells can, for example, be stretched between two beads that are slowly moved apart. The biggest limitation of using optical traps in cell mechanics experiments is that the maximal force level is limited to < 1 nN, as larger forces would require higher laser power that could excessively heat the cell. While the small forces generated by an optical trap are ideally suited for single-molecule studies and are sufficient for experiments that measure, for example, membrane tethers, they are often too small to induce large-scale cellular deformations, especially when probing stiffer cells such as myocytes.

Micropipette Aspiration. In these experiments, single adherent or suspended cells are partially aspirated into a micropipette with a $\approx 2-10$ μm-diameter opening by applying precisely controlled suction pressure (typically 100−10 000 Pa). The aspirated cell is imaged on a microscope and cellular deformations such as the aspirated tongue lengths are computed from the cell geometry. In a technique called fluorescent/confocal-imaged microdeformation, fluorescent labeling of specific intracellular components such as the nucleus, the nuclear lamina, or nucleoli is used to provide additional information on the subcellular deformations during micropipette aspiration [31.43–46]. One potential limitation of the micropipette aspiration technique is that the interpretation of the experiments is not always straightforward. Analytical or computational models are often necessary to derive material properties from the geometric measurements of the aspirated cells and the applied pressure, and the underlying assumptions may at times be difficult to validate. However, this technique has been proved to be very useful when studying cells with relatively homogeneous structural organization such as red blood cells or neutrophils [31.44, 47–49]. In recent years, microfluidic devices have been developed to enable higher-throughput measurements applying the same principles as micropipette aspiration, but using arrays of microfabricated channels to aspirate or perfuse many cells in parallel [31.50–55].

Substrate Strain. Unlike the other techniques, these experiments do not apply controlled forces to single cells, but instead use carefully controlled strain application to induce deformation in cells plated on a flexible substrate. Generally, cells are plated on transparent, elastic silicone membranes coated with extracellular matrix proteins and subjected to uniaxial or biaxial strain. Depending on the particular experiment, the strain can be held constant or varied over time (e.g., cyclic strain application). Strain levels normally do not exceed 30% in order to avoid damaging the cells, but the exact levels are cell-type dependent. Cells are imaged under a microscope before, during, and after strain application, and small markers or fluorescently labeled components of the cells are used to calculate intracellular deformations and applied membrane strain. Since the cell is firmly attached to the extracellular matrix on the silicone membrane through cell surface receptors which are connected to the cytoskeleton, the cytoskeleton will experience strain levels comparable to the applied substrate strain, whereas the stiffer nucleus typically deforms significantly less [31.32, 56]. The advantages of this technique are that force can be applied to several cells at once and that the strain application closely resembles physiological mechanical stress, e.g., in muscle cells or endothelial cells subjected to blood vessel expansion. Recently, this technique has even been adapted to study nuclear deformation in muscle tissue of living worms [31.36]. The major limitations are that this technique is only suitable for adherent cells and that the applied forces cannot be determined directly.

Shear Flow. The most commonly used devices for shear stress application are the cone and plate rheometer and flow (or perfusion) chambers. Cone and plate rheometers allow precise control over the applied shear stress, but are generally not equipped to image cells during shear stress application, making it difficult to visualize induced cellular deformations. On the other hand, flow chambers are routinely used to study cellular deformations under shear stress [31.57] or to investigate cellular responses to shear stimulation such as calcium influx or cytoskeletal remodeling. Parallel-plate flow chambers with channel heights ≈ 100 μm

are now widely available using standard soft lithography and microfluidics, allowing for precise control of the fluid shear stress acting on a cell monolayer and simultaneous imaging of the cells using a microscope. The shear stress at the cell surface can be calculated for a Newtonian fluid in the parallel-plate geometry as $\tau = 6Q\mu/wh^2$, valid when $h \ll w$, where τ is the fluid shear stress at the wall (and cell surface), Q is the flow rate, μ is the viscosity, w is the channel width, and h is the channel height. The shear stress can thus be adjusted to physiological shear stress levels ($\approx 0.1-10$ Pa) by altering the height of the flow chamber or modulating the flow rate, and experiments can be carried out with either constant or pulsatile flow. However, the actual shear stress at the cell surface might deviate from the predicted wall shear stress, as small variations in the cell height and topology can cause local variations, which might in fact contribute to the alignment of endothelial cells to the flow direction. Recent work has extended these methods to characterize the mechanical phenotype of populations of cells in suspension [31.58, 59]. This approach also has potentially important clinical implications, such as in the use of mechanical markers of diseases such as cancer [31.55].

Passive Measurements

In contrast to the active measurements, these experiments measure forces generated by the cells themselves or quantify the random motion of particles embedded in the cytoplasm subjected to thermal fluctuations. Typical examples of passive measurement techniques include particle tracking of beads in the cytoplasm and traction force microscopy. In passive bead microrheology, small ($0.1-2\,\mu$m diameter) beads are injected into the cytoplasm or taken up by endocytosis and are then tracked with high spatial and temporal resolution with a laser beam and quadrant photodetector. The complex cytoplasmic shear modulus $G(s)$ can then be computed from the unilateral Laplace transform of the measured mean-squared displacement $\langle \Delta r^2(s) \rangle$ using a generalized Stokes–Einstein equation as $G(s) = k_B T/(\pi a s \langle \Delta r^2(s) \rangle)$, where s is the Laplace frequency, k_B is the Boltzmann constant, T is the temperature, and a is the bead radius [31.60], although care must be taken with this approach to avoid artifacts associated with active, motor-driven motions [31.61]. Traction force microscopy (TFM) was originally based on wrinkles generated in thin silicone sheets by the contraction of cells plated on top of these sheets, but has subsequently been refined for more quantitative force determination by plating cells on polyacrylamide gels with fluorescent beads embedded in the gel. The mechanical stiffness of the polyacrylamide gel can be tuned based on the cross-linker concentration, and microscopic measurements of the induced displacement of beads near the surface can be converted into cellular forces exerted on the gel by elastic theory [31.62, 63]. In another alternative to TFM, cells are coated on a fine grid of microfabricated micropillars made of polydimethylsiloxane (PDMS) [31.64]. In this case, the stiffness of the micropillars can be controlled by adjusting their cross-sectional area or length, and the cellular force exerted on each micropillar can be measured based on the deflection of the pillar using beam bending theory. This technique offers the advantage that force application is localized to individual micropillars, and that the geometry and stiffness can be independently adjusted.

An important recent advance has enabled TFM to be performed on cells embedded in a 3-D tissue. If the mechanical properties of the surrounding matrix are known, then the tissue displacements can be used to calculate stresses within the material or at the cell surface [31.65]. The methods of analysis are similar to those used in 2-D, but are much more complex in three dimensions. Moreover, interpretation of the data is complicated by the use of realistic matrix materials that are subject to both remodeling and degradation by the cells, leading to uncertainties regarding the matrix properties that become both heterogeneous and nonisotropic.

Another method used to infer single-molecule strains is to tether a pair of fluorophores to the molecule with the emission wavelength of one (the *donor*) corresponding to the absorption wavelength of the other (the *acceptor*), termed fluorescence (or Förster) resonant energy transfer (FRET) microscopy. In the most common versions, the intramolecular strain or tension sensors, a pair of FRET fluorophores is separated by an elastic linker domain. As the linker domain becomes stretched under applied force, the ratio of the emitted light from the acceptor to that from the donor changes, and this ratio can be used as a measure of the distance between them [31.66]. Such molecular strain sensors have been successfully applied to focal adhesions (vinculin), nesprins, and various other molecules. In addition to intramolecular FRET, the FRET fluorophores can also be incorporated into two interacting molecules, such as cell surface receptors [31.67]. In both cases, the FRET system can be calibrated using AFM or optical traps so that the change in the ratio of energy from the acceptor to that from the donor (the FRET ratio) can be converted into a force, although this conversion is fraught with numerous potential errors. As a result, the probes are most often used to indicate semiquantitative changes in the relative stretch/tension between two conditions of the cell.

Passive methods can also be applied for measurement of cell or matrix stiffness in the absence of any

added label or particles. A recent development in this connection utilizes the frequency change of scattered light by the internal filaments of the cytoskeleton to measure the uniaxial longitudinal compression modulus in confined compression [31.68]. Measured values of frequency shift have shown strong correlations with the moduli measured at more typical physiologically important frequencies.

31.3.2 Rheological Properties

The rheological properties of the cell, such as its viscoelastic properties and its diffusion parameters, are key to the cell's ability to accomplish its diverse functions in health and disease. A wide range of computational and phenomenological models as well as experimental techniques have been proposed over the past two decades to describe the cell, giving rise to several, often contradictory, theories for describing the rheology of the cytoskeleton. The highly heterogeneous structure of the cytoskeleton, coupled with a small linear response regime [31.69] and active dynamics and continuous remodeling, present a major challenge for quantitative measurements and descriptions of its rheology [31.70].

Nonetheless, a wide range of computational models exist for cytoskeletal rheology and mechanics, ranging from continuum to discrete descriptions of the cytoskeleton (reviewed in [31.71]). A major challenge in cytoskeletal rheology and mechanics is how to relate experimental observations to theoretical and phenomenological models. Much effort has recently focused on the interesting rheological behavior of cells, measured by one of the methods described above. Most often, cellular viscoelastic properties are expressed in terms of the complex shear modulus (see *Viscoelastic Solid/Poroelastic Solid*), the real part G' indicating the elastic component, and the imaginary part G'' the viscous component. While each measurement method has its drawbacks, perhaps the most comprehensive data have been obtained through magnetic twisting cytometry (Fig. 31.6). These measurements have shown that a cell exhibits a relatively simple power-law behavior over much of the frequency domain, with [31.72]

$$G' + \mathrm{i}G'' = G_0 \left(\frac{\omega}{\omega_0}\right)^{x-1} (1+\mathrm{i}\eta)\Gamma(2-x) \times \cos\left[\frac{\pi}{2}(x-1)\right] + \mathrm{i}\omega\mu \,, \qquad (31.1)$$

where Γ is the Gamma function, μ is the viscosity, and $\eta = \tan((x-1)\pi/2)G_0$, ω_0, and x are parameters of the model. All but x, however, have been found to be nearly constant in all experiments. This turns out to be similar to the behavior of soft glassy materials, although the fundamental basis for this behavior remains a topic of some debate.

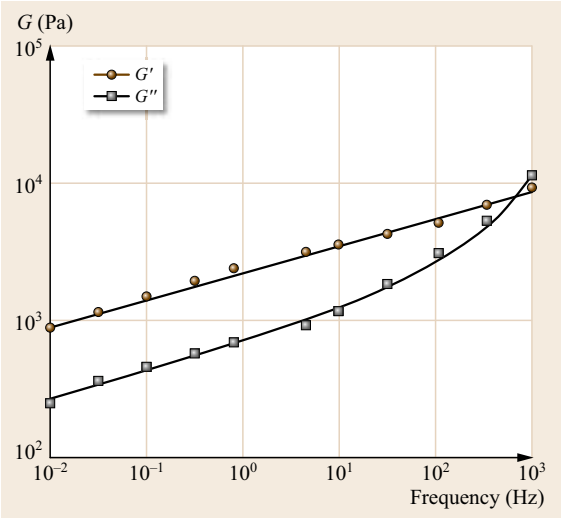


Fig. 31.7 Typical experimental results for the frequency-dependent shear moduli of a wide range of cell types and conditions obtained by magnetic twisting cytometry (after [31.72])

The data in Fig. 31.7 [31.72] apply to small deformations in the linear regime. However, linearity persists up to surprisingly large deformations. Much of what we know about nonlinear effects has been obtained from experiments on reconstituted gels, typically of actin with one or more actin cross-linking protein. From these experiments, the following observations have been made:

1. Linearity persists up to $\approx 30\%$ strain.
2. Values of G' and G'' in this linear regime are often orders of magnitude lower than are measured in cells.
3. Strain stiffening is observed at strains $> 30\%$, followed by a precipitous drop in G' at strains of $\approx 70\%$.

All of these issues continue to be actively studied in the hope of generating a comprehensive understanding of cytoskeletal rheology.

31.3.3 Active Force Generation

The cytoskeleton is an active structure that maintains cell shape and facilitates its motion. The cytoskeletal network is in a nonequilibrium state that drives motor proteins as force-generating features in cells. The cytoskeleton is activated by these molecular motors, which are nanometer-sized force-generating proteins, e.g., myosin. To better understand the nature of

such active force-generation systems, researchers have developed simplified models of the active cytoskeleton by mixing actin with actin binding proteins (e.g., filamin and α-actinin) and molecular motor proteins (myosins) [31.73]. Such a reconstituted synthetic cytoskeleton exhibits local contractions reminiscent of living cells. Tension generated by contraction can lead to drastic increase of the cytoskeletal stiffness. Such models demonstrated that a remarkably simple system, with just three components (myosin, actin, and ATP), can reproduce key phenomena also observed in far more complex living cells [31.67].

31.3.4 Biological Responses

The cellular responses to mechanical stimuli take place over several time scales and range from intracellular signaling to changes in cellular morphology and function. The fastest responses occur within seconds to minutes of stimulation and include changes in intracellular ion concentrations (especially Ca^{2+}) through opening of mechanosensitive ion channels and activation of cellular signaling pathways such as nuclear factor-κB (NF-κB), phosphatidylinositol-3-kinase (PI3K)/protein kinase B (Akt), protein kinase C (PKC), Rho family GTPases, and mitogen-activated protein kinases (MAPKs) [31.74, 75]. These first steps do not require synthesis of new proteins, but involve modification (e.g., phosphorylation) and intracellular translocation of existing proteins. These initial events often trigger activation of mechanosensitive immediate early genes such as *egr-1*, *c-fos*, *c-jun*, or *c-myc* encoding transcription factors that turn on additional downstream genes. These later response genes often include proteins involved in cellular structure (e.g., actin, myosin) or extracellular matrix remodeling. However, cytoskeletal remodeling in response to mechanical stress or strain can start even before synthesis of new proteins as early signaling pathways can also mediate polymerization and depolymerization of cytoskeletal structures; For example, endothelial cells exposed to fluid shear stress begin to align with the flow direction within 20 min, and the process is completed within 24 h. The multifaceted response to mechanical stimulation can effectively mediate several cellular functions, including DNA and RNA synthesis, hypertrophy (increase in cell size), proliferation, apoptosis, migration, and extracellular matrix remodeling.

31.3.5 Nonlinear Effects

As described above, cells are composed of intricate networks of filamentous structures, called the cytoskeleton. These networks exhibit unique properties including relatively large shear moduli with strong signatures of nonlinear response in which, for example, the shear modulus can increase drastically under modest strains [31.76, 77]. Models of semiflexible polymer networks have emerged to describe these unique properties and the dynamics of the cytoskeletal networks. These models involve a semiflexible description of the constituent actin filaments [31.78]. Quantitative models of the nonlinear behavior of the cytoskeletal network are essential for understanding the complex, dynamic, and nonlinear behavior of cells, and ultimately tissues and organs.

31.3.6 Homogeneity and Anisotropy

Cells are inhomogeneous structures composed of various intracellular components, including the lipid bilayer membrane that encases the cell, the actin cortex, which is a dense actin network providing stability for the cell membrane, cytoskeletal networks, and the nucleus, which itself comprises various important substructures. The cytoskeleton is primarily responsible for the structural integrity and stiffness exhibited by a cell. It is composed of a system of highly entangled protein filaments that permeate the microfluidic space of the cytosol. The major components of the cytoskeletal network, i. e., the actin filaments, intermediate filaments, microtubules, and their cross-linking proteins, offer an exquisite microenvironment which is highly inhomogeneous and anisotropic in its structure and geometry.

31.4 Theoretical and Computational Descriptions

As is evident from the discussion above, the cytoskeleton is a complex biopolymer network with varying degrees of connectivity, existing in a state of dynamic equilibrium. The dynamic state arises from the ongoing polymerization and depolymerization of the constituent filaments and the changing density of cross-links between filaments of the same or different families. This picture is complicated further by the milieu of other intracellular constituents that may or may not affect structural properties exhibited by the cell.

Our objective in this section is to come to a better appreciation of how the elastic properties and geometric arrangement of the constituent filaments give rise to the material properties observed by the various ex-

perimental methods just described. These experiments have shown that the elastic modulus of cells can vary considerably from one cell type to another. Cells of the epidermis, for example, require greater structural integrity than red blood cells that are subjected to relatively low shear stresses in the blood. Even within a given cell type, the elastic properties can change. Skeletal muscle, for example, changes its modulus by over an order of magnitude within a small fraction of a second. Other cells change too, but more often over a longer time period, in response, for example, to changes in their mechanical environment.

According to the various measurements that have been made, cell shear moduli fall in the range $\approx$ 10–10 000 Pa, and therefore exhibit a stiffness somewhat lower than collagen gels (or common gelatins) at low concentration or relaxed skeletal muscle. This wide range of moduli can be attributed to differences in the models used as a basis to infer the shear modulus from the data as well as to real cell-to-cell variability. At best, these numbers inferred from experiment should be viewed as measures of an *effective stiffness*, and comparisons between different measurement methods and interpretation should be made with caution. Other biological materials (e.g., bone, wood) exhibit a much higher modulus, but not because of their cellular content. Rather, their high stiffness is due to the calcification in bone and the matrix molecules found in wood and most plants. In tissues, too, the stiffness we measure is more often associated with the extracellular matrix with its elastin and collagen, than the resident cells. Even within the cytoskeleton, individual filaments (i. e., F-actin, intermediate filaments, and microtubules) exhibit moduli much greater than the measured bulk modulus of the cytoskeleton. In the next section we explore different approaches that have been developed to relate the properties of the individual fibers to those of the assembled network.

31.4.1 Continuum Models

Elastic Solid

In several of the experiments described earlier, the assumption of a homogeneous and isotropic elastic solid is used to infer a value for an effective Young's modulus for the cell. It is not our intention here to provide a comprehensive description of elasticity theory, but rather to outline some of the basics. For a full derivation of the governing equations, the reader is referred to any of a number of excellent textbooks [31.79]. Assuming equilibrium conditions (i. e., that all forces balance each other) and a material that can be modeled as a Hookean (i. e., linear) elastic solid, the constitutive equations relating the mechanical stress and strain can be written in tensor notation as $\tau_{ij} = C_{ijkl}\varepsilon_{kl}$, where the summation convention is used. Here τ_{ij} are the elements of the stress tensor, ε_{kl} are the elements of the strain tensor, and C_{ijkl} is the 81-element coefficient matrix. In the case of an isotropic material, the coefficient matrix reduces to just two independent elastic constants, with the constitutive equation now reduced to the following simplified form

$$\tau_{ij} = 2Ge_{ij} + \lambda\varepsilon_{kk}\delta_{ij}\,, \tag{31.2}$$

where λ and G are the Lamé constants with G being termed the shear modulus. This expression can also be written in the inverted form

$$\varepsilon_{ij} = \left[\frac{(1+\nu)}{E}\right]\tau_{ij} - \left(\frac{\nu}{E}\right)\tau_{kk}\delta_{ij}\,. \tag{31.3}$$

Here the two material constants are now given as ν, the Poisson's ratio, and E, the Young's modulus. When using these to solve a particular problem, a set of appropriate boundary conditions also needs to be posed; For example, in the case of measurements by an indentation probe, the displacement of the surface in contact with the probe might be given. In addition, it would be necessary to specify that the opposite side of the cell is held fixed, and that the unsupported sides and the region on the upper surface not in contact with the probe have zero applied stress. Given these, the equations above provide a complete solution.

The use of the above equations presumes that the response of the cytoskeleton can be represented by an *elastic continuum* lacking any discernible microstructure. The validity of this clearly depends upon the length scale of interest in the problem, and is a serious concern in the context of cytoskeletal mechanics where the typical length scale (spacing distance between filaments) may be on the order of 10–100 nm. This becomes comparable to the linear dimension of the region of interest in some of the experimental procedures (e.g., cell poking, AFM measurements) and needs to be kept in mind. Furthermore, while this is a convenient approach for analyzing how cells deform under loading, it provides no insight into the relationship between the macroscopic properties that are measured and the elastic properties of the constituent matrix elements.

Viscoelastic Solid/Poroelastic Solid

A viscoelastic material is to an elastic material as a spring and dashpot network is to a system containing only springs. In both the discrete and continuous systems, the instantaneous displacement or strain is a function of the stress history. Similarly, the instantaneous stress depends on the strain history. This simple

realization, in combination with the assumption that all functions describing the material properties are continuous, leads to a set of generalized constitutive equations for a viscoelastic solid.

Many of the experiments performed on cells have been interpreted in the context of either an elastic or viscoelastic model. However, it might be argued that, like bone, cartilage, and many other biological tissues, it may be more appropriate to view the cytoskeleton as a poroelastic material. While attempts to apply a poroelastic model to studies of cell mechanics are just beginning, it remains to be seen whether a viscoelastic or poroelastic description is more appropriate. For a description of poroelastic theory, the reader is referred to [31.80].

Oscillatory Simple Shear. A useful example of viscoelastic behavior derives from the case in which the upper surface of a sample is being oscillated sinusoidally in the plane of the surface so that the shear strain satisfies $\varepsilon_{21} = \varepsilon_{21}^* \sin \omega t$ and the material experiences an oscillatory shear stress. In this case, the stress in the viscoelastic material will also vary sinusoidally, but out of phase with the strain. The relationship between stress and strain is often described by introducing the complex shear modulus (as introduced above), $G^* = G' + \mathrm{i}G''$, satisfying the following expressions

$$\tau_{21}(t) = \varepsilon_{21}^*(G' \sin \omega t + G'' \cos \omega t) \,, \tag{31.4}$$

$$\begin{aligned} \tau_{21}(t) &= \tau_{21}^* \sin(\omega t + \phi) \\ &= \tau_{21}^* \cos\phi \sin \omega t + \tau_{21}^* \sin\phi \cos \omega t \,, \end{aligned} \tag{31.5}$$

so that G', G'', and ϕ, the phase lag, are related to the amplitudes of strain and stress through the expressions

$$G' = \omega \int_0^\infty G(t') \sin \omega t' \, \mathrm{d}t' = \left(\frac{\tau_{21}^*}{\varepsilon_{21}^*}\right) \cos\phi \,, \tag{31.6}$$

$$G'' = \omega \int_0^\infty G(t') \cos \omega t' \mathrm{d}t' = \left(\frac{\tau_{21}^*}{\varepsilon_{21}^*}\right) \sin\phi \,. \tag{31.7}$$

Note that G' represents the component of stress in phase with the imposed strain and G'' the out-of-phase component. For a purely elastic material (in which strain and stress are in phase), G' is simply the shear modulus G and G'' is zero. On the other hand, it can be shown that, for a viscous Newtonian fluid, G' is zero and $G'' = \mu\omega$, where μ is the Newtonian viscosity.

It should be noted that these results could also have been obtained by expressing the time-varying strain as a complex quantity. In so doing, the shear modulus is complex, and G' and G'' are, respectively, the real and imaginary parts of the stress–strain ratio.

31.4.2 Biopolymer Models

Cytoskeletal networks can be viewed as a polymer gel in which the matrix is considered to consist of relatively straight segments connecting junctions where the filaments are either chemically cross-linked, or effectively so due to entanglements (Fig. 31.8a). Using concepts from polymer physics [31.81], the force required to change the length of one segment of a polymer filament (F-actin, for example) of length L_e by an amount δ can be expressed as

$$F_\mathrm{T} \approx \frac{k_\mathrm{B} T l_\mathrm{p}^2}{L_\mathrm{e}^4}\delta \approx \frac{K_\mathrm{b}^2}{k_\mathrm{B} T L_\mathrm{e}^4}\delta \,, \tag{31.8}$$

where k_B is Boltzmann's constant, T is temperature, K_b is the bending stiffness of the filament, l_p is the persistence length, and L_e is the distance between the points where the tension is applied (or the distance between points of entanglement or cross-linking between network filaments). This expression arises from a consideration of the curvature of the polymer resulting from Brownian or thermal fluctuations, and is based on the assumption that thermal energy is equally partitioned among the different modes of oscillation. The polymer filaments are therefore assumed to be bent prior to the application of stress. Externally imposed forces either increase or decrease the end-to-end length of these filaments, and the deformation produced depends both on the intrinsic bending stiffness of the filaments and on their initial degree of curvature due to thermal fluctuations.

For a network composed of such filaments in which the distance separating points at which physical bonds exist between the filaments or regions of entanglement is L_e, the change in filament length between bonds due to a shear strain θ is $\delta \propto \theta L_\mathrm{e}$. Making use of the fact that the number of filaments per unit area parallel to the surface on which the stress is applied scales inversely with the square of the characteristic mesh spacing ξ, the shear stress required to produce a strain θ is given by

$$\tau \propto F_\mathrm{T} \frac{(\text{number of filaments})}{(\text{unit area})} \propto \frac{k_\mathrm{B} T l_\mathrm{p}^2 \theta}{\left(L_\mathrm{e}^3 \xi^2\right)} \,, \tag{31.9}$$

from which a shear modulus of the (elastic) network G_n can be determined as the ratio of stress to strain

$$G_\mathrm{n} \propto \frac{\tau}{\theta} \propto \frac{k_\mathrm{B} T l_\mathrm{p}^2}{\left(L_\mathrm{e}^3 \xi^2\right)} \,. \tag{31.10}$$

The scaling of G_n clearly depends on both the distance between cross-links or entanglements and the mesh

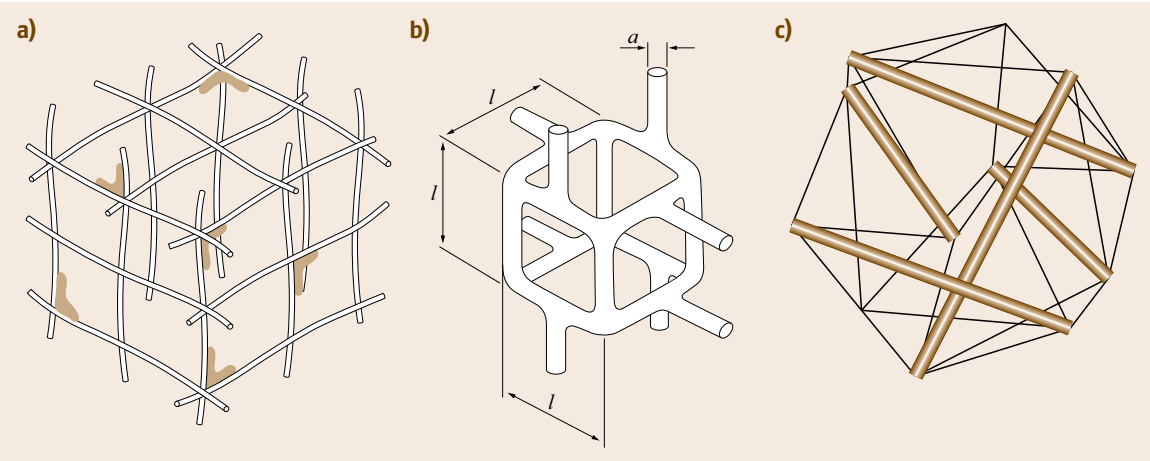


Fig. 31.8a–c Cell mechanics models. **(a)** Schematic of a network of biopolymers consisting of a fiber matrix with cross-links. **(b)** Unit cell used in the cellular solids model (after [31.82]). **(c)** Tensegrity structure showing the balance between elements in compression (*cylinders*) and others in tension (*lines*) (after [31.83])

size. As the concentration of polymer increases, ξ decreases and so likely does L_e, at least in terms of the degree of entanglement. As the concentration of actin binding protein (ABP) increases, L_e will decrease. The mesh spacing ξ also depends on the monomer concentration or solid fraction Φ (monomer volume/total volume). Assuming that the filaments are stiff and homogeneously dispersed through the medium, this relationship can be expressed as $\xi \propto a/\Phi^{1/2}$, where a is the monomer or filament radius. Substituting yields

$$G_n \propto \frac{k_B T l_p^2 \Phi}{\left(L_e^3 a^2\right)} . \tag{31.11}$$

In the limit of a highly cross-linked network, in which case $L_e \approx \xi$, this leads to

$$G_n \propto k_B T l_p^2 \left(\frac{\Phi}{a^2}\right)^{5/2} . \tag{31.12}$$

Alternatively, we can express this in terms of the density of cross-links ρ_c. Note that, if ρ_c is the number of cross-links per unit volume, it should vary inversely with the volume associated with each bond or entanglement, i. e., as L_e^{-3}. Combining this with the expressions (31.10–31.12) gives

$$G_n \propto k_B T l_p^2 \frac{\Phi \rho_c}{a^2} . \tag{31.13}$$

The Young's modulus of the network E_n can be obtained by a similar procedure, and can be shown to scale in the same manner as G_n.

31.4.3 Cellular Solids

The theory of cellular solids was developed for the purpose of relating the macromechanical properties of low-density cellular materials to their microstructural characteristics. The approach used is based on the concept that the material can be modeled as being composed of many *unit cells*, one representation of which is shown in Fig. 31.8b. When a cellular solid is stressed under tension or compression, the fibers act like struts and beams that deform under stress. The unit cell model in Fig. 31.8b has struts or fiber elements of length L and cross-section of radius a. The relative density of the material is defined by the volume fraction of solid material Φ. This is calculated as the solid volume contained within a unit cell ($\propto a^2 L$) divided by the total unit cell volume ($\propto L^3$) or $\Phi \propto (a/L)^2$. Beam theory gives the deflection δ of a beam of length L subject to a force F acting at its midpoint as

$$\delta \propto \frac{F L^3}{E_f I} , \tag{31.14}$$

where E_f is the stiffness of the beam constitutive material and I is the beam's second moment of area. The moment of area for a beam of thickness a is given by $I \propto a^4$. The stress τ is the force per unit area, or $\tau \propto F/L^2$. The strain is related to the beam deflection δ by $\varepsilon \propto \delta/L$. Using these, combined with the expressions above, the network Young's modulus or elastic modulus can be expressed as

$$E_n = \frac{\tau}{\varepsilon} = \frac{c_1 E_f I}{L^4} , \tag{31.15}$$

where c_1 is a constant of proportionality, or

$$\frac{E_n}{E_f} = c_1 \Phi^2 \,. \tag{31.16}$$

Data from a wide range of materials and cell geometries give a value for c_1 of ≈ 1 [31.82]. A similar analysis for cellular materials subjected to shear stresses results in an expression for the network shear modulus G_n [31.82]

$$\frac{G_n}{E_f} \propto c_2 \Phi^2 \,, \tag{31.17}$$

where $c_2 \approx 3/8$. If the material is linearly elastic and isotropic, these values lead to a value of $1/3$ for the Poisson ratio ν.

31.4.4 Tensegrity

Networks can also derive their structural integrity from an interaction between members that are in compression and members in tension. Some familiar large-scale examples include circus tents and geodesic domes. In the case of the circus tent, the rigidity of the structure is due to the balance between the tent poles in compression and the ropes anchored to the ground in tension. The rigidity of the structure, in this case, is related to the elastic characteristics of the tension elements.

Ingber [31.84] first proposed that the cytoskeleton behaves like a tensegrity structure with the microtubules acting in compression and the F-actin microfilaments acting in tension. In support of this concept, microtubules have been shown to be capable of supporting compressive loads and the F-actin network exhibits behavior at junctions consistent with their being in tension. Intermediate filaments may also be involved, although their contributions at this stage are unclear.

Analysis of a fully three-dimensional tensegrity network begins, as in the case of the cellular solids model, with a unit cell consisting of an interconnected system of elements in tension in balance with other elements in compression. When the three-dimensional tensegrity network of Fig. 31.8c is used, analysis shows that prestress plays an important role. Intuitively, it is not surprising that networks with greater pretension in the elastic members should exhibit greater resistance to deformation. The effect of prestress increases the network Young's modulus for small strains, and only in the limit of infinite prestress does E_n become independent of prestress. *Wang* and coworkers [31.85] showed that the cytoskeleton exhibits this same tendency, becoming increasingly stiff when, for example, the cell passes from a spherical (low prestress) to flattened (high prestress) state.

Nonetheless, evidence suggests that the simplicity of this model may undermine the accuracy of its predictions. Continuum mechanics predicts that a free, simply supported beam tends to buckle in the first mode of buckling and that is the case for in vitro loading of a single microtubule. However, in vivo imaging of microtubules indicates that the buckling mostly occurs in higher modes. This buckling mode switch takes place mostly because of lateral support of microtubules via their connections to actin and intermediate filaments, which themselves are tensile members of the tensegrity cytoskeleton model. Since these loads are exerted throughout the microtubule length, they introduce a considerable amount of microtubule bending behavior [31.86]. The sensitivity of the tensegrity model of the cytoskeleton to flexural deflections of microtubules has been explored recently. Microtubules, as crucial components of the cytoskeleton, cooperate with actin filaments to confer the cell with dynamic stability. Using an anisotropic continuum model for microtubules, it has been suggested that the bending energy could exceed the axial energy of microtubules by 40-fold, necessitating a modification to the tensegrity model in order to account for the flexural response of microtubules.

31.5 Mechanics of Subcellular Structures

31.5.1 Cell–Cell and Cell–Matrix Adhesions

Cells can adhere to their surroundings by either nonspecific or receptor–ligand (specific) bonding. Though both mechanisms are likely active in most situations, receptor–ligand bonding is the stronger of the two, by a considerable margin, and is therefore the most relevant mechanically. In situations for which either the receptors or their ligands are not present, however, such as might be the case in certain in vitro experiments, nonspecific binding can be important as well, and can produce a net attractive force per unit area of ≈ 100 Pa at a separation distance of ≈ 25 nm.

Models for Receptor-Mediated Adhesion

Cell–cell adhesion has obvious similarities to adhesion of cells to substrates, and the extracellular matrix, so similar approaches can be used for all of these situations. This requires, in addition, consideration of such factors as the distribution, type, and density of receptor–

ligand bonds or potential bonds, and the elastic properties of the structures to which they are anchored. On the intracellular side, this involves the series of couplings that link the receptor to the cell. In the simplest case, this might simply be a link to the lipid bilayer if the receptor has no intracellular connections. More typically, especially for couplings with a structural role, it involves a series of proteins ultimately linking the receptor to the cytoskeleton.

The typical setting in vivo is one in which cells adhere to other cells or to the extracellular matrix. Adhesions are more easily probed, however, through in vitro experiments, where cell adhesion more often occurs to an artificial substrate mediated through one of several extracellular proteins that are used to coat the surface. Typically, either collagen or fibronectin is used. Cells are often adhered to these substrates, but to produce a more controlled environment, rigid beads are sometimes coated with the appropriate receptors so that one specific receptor–ligand interaction can be probed. While these systems are useful as models of certain adhesion phenomena, it is important to recognize in the interpretation of these experiments that, when bound to a rigid substrate or bead, the binding proteins cannot freely diffuse, as they would in a more natural environment. In particular, the formation of focal adhesions would not occur in bead–substrate experiments because the receptors would be constrained from aggregating.

It is useful to consider a single adhesion bond, for example, one linking the actin matrix of the cytoskeleton to a β_1 integrin, and the integrin receptor binding to the extracellular matrix beyond the cell membrane. If the bond is stressed, as for example if the cell experiences a force relative to the ECM, it will at first stretch an amount dictated by the level of force in the bond and the stiffness of the complex. Each bond, as well as each protein in the bond complex, can be thought of as having a certain stiffness, giving rise to a picture in which several springs are considered connected in series. Forces acting on the adhesion complex are transmitted via this series of bonded proteins between the cytoskeleton (CSK) and the ECM, producing local deformations and stresses in the corresponding matrices that decay with distance from the adhesion site. On the intracellular side, these forces are transmitted via a complex involving vinculin, α-actinin, paxillin, and talin. Attachment to the extracellular matrix is mediated by an arginine–glycine–aspartic acid (RGD) sequence (in fibronectin, for example), which in turn, has binding sites for collagen and fibrin. If the force is sufficiently strong, above a certain threshold value F_0 say, and applied for a sufficiently long time, the bond might be severed (decoupled). Typical values of this threshold force lie in the range of 10−100 pN [31.87] but depend, as well, on the rate at which the force is increased. Under more rapid force application, the threshold is high compared with when the force is increased slowly. This simply reflects the fact that detachment is a stochastic process and that, at any given level of force, there exists a finite probability of detachment, and this probability increases with time. It has also been shown that some bonds, particularly those associated with cell–cell adhesion, strengthen when subjected to force; That is, the unbinding rate decreases with increasing force. These important phenomena are reviewed in [31.88].

We need also to consider that, in equilibrium, a fraction $\Phi = K(1+K)^{-1}$ (where K is the equilibrium constant) of the receptors will be bound at any given time, and that the receptor–ligand complexes are continually cycling between the bound and unbound states, characterized by their respective rate constants. On a larger scale, numerous adhesion sites typically act in parallel, each contributing some amount to the total adhesion force. This gives rise to an average local stress τ that can be thought of as the product of the average force per bond times the bond density. This stress is typically nonuniform. In the adhesion of a cell to a flat substrate, the bonds near the periphery of the contact zone determine the strength of the cell against detachment, and when the cell is subjected to a detaching force, the stresses are concentrated in this peripheral region. In the vicinity of the edge of adhesion, several factors will influence the stress distribution, including the mechanical properties of the membrane (in particular the bending stiffness K_b), the receptor density, the spring constants of the bond, and the bond strength.

The interplay between receptor–ligand bonding and membrane stiffness also determines whether a cell will spread over a surface or peel away from it under a given tension. Membrane bending stiffness acts to maintain the membrane and substrate within close proximity near the edge of contact. Spreading can occur only if their separation distance is within the extended length of the receptor ligand bond over a distance comparable to the linear spacing between neighboring bonds. Under high tension, the surfaces rapidly diverge and spreading is prevented; if the tension is sufficiently low, receptor–ligand bonding can occur and spreading can occur.

Transient Cell Adhesion

Concepts of transient adhesion and release are of particular importance in the case of leukocyte rolling, adhesion, and transmigration across the endothelial layer of a blood vessel. In addition, cell migration through extracellular matrix or along a substrate requires the ability of the cell to both form and release adhesions. One problem that has received considerable attention is the interaction between an adherent or rolling leuko-

cyte and the adhesion receptors on the endothelium. These studies are complicated by the three-dimensional nature of the flow, the compliance of the interacting surfaces, and receptor–ligand dynamics. In a series of studies, *Hammer* and coworkers [31.89] addressed several of these issues, computing the viscous force and torque acting upon a sphere near to a planar wall from the mobility matrix [31.90] and using the Bell model of receptor binding [31.90], but neglecting the effects of cell deformation.

A Monte Carlo method is employed at each time step in the calculation to determine bond formation or breakage. In these studies, the nature of the leukocyte–endothelial interaction is characterized as firm adhesion, rolling adhesion, bimodal adhesion or no adhesion, and mapped as a function of the two parameters of the Bell model. Most literature values for the Bell model parameters for a variety of selectin receptors (known to be instrumental in leukocyte rolling) correspond to the rolling adhesion regime for typical shear rate of $100\,s^{-1}$. An interesting, but counterintuitive, result from this study was the observation that, as the bond stiffness was increased from the value used in most calculations (100 dyn/cm), adhesiveness *decreased*, which the authors attributed to reduced deflection for a given level of force, leading subsequently to more rapid dissociation.

31.5.2 Cell Membranes

For the purpose of analysis, we treat the cell membrane as a homogeneous two-dimensional (2-D) plate or sheet completely enclosing the cytoplasm. The membrane referred to here can be thought of either as the lipid bilayer by itself, or more typically, as the bilayer plus the associated cortex of cytoskeletal filaments and the glycocalyx on the extracellular surface. In addition, though not explicitly recognized in the analysis, transmembrane proteins and their attachments to the intracellular and extracellular milieu are included in terms of their influence on the continuum properties of the model. Were it not for these, the membrane would exhibit little resistance to shear deformation.

In qualitative terms, the lipid bilayer can be thought of as a two-dimensional fluid, within which the individual lipid molecules, or other molecules embedded in the membrane, are relatively free to move about by diffusion or directed motion. Phospholipid molecules in either of the two layers resist being pulled apart, however, so each layer is highly inextensible. This also contributes to the bending stiffness, which is low in absolute terms but high for such a thin layer, since bending requires one layer to expand while the other is compressed. By contrast, the two layers readily slide relative to each other. These qualitative notions are put in more quantitative terms in the next section.

Types of Deformation

Any deformation of the membrane can be thought of, in general terms, as a superposition of several simpler deformations. For small strains for which linearization is appropriate, the principle of superposition is rigorously valid. For larger strains, however, linear theory breaks down and superposition can only be used as a rough, qualitative guide in visualizing combined influences. Here we present the three primary types of deformation: pure extension, pure bending, and pure shear.

Pure Extension. In discussing the extensional stiffness of the membrane, we need to distinguish the behavior at low tension from that at high tension. As an extensional stress is first applied at the edges of a lipid bilayer, the projected or *apparent* membrane area first increases while the actual or true membrane area remains constant. This results from the suppression of out-of-plane undulations. Forces acting to resist membrane flattening originate from entropic effects analogous to those seen in the case of a flexible polymer as its end-to-end distance is increased – many more membrane configurations exist with undulations compared with the single, perfectly flat state. Only when these undulations have been eliminated does the *true* membrane area, proportional to the surface area per molecule group, begin to increase, and this is associated with a relatively abrupt increase in extensional stiffness. For now, we consider only the stiffness of a flat membrane and leave the discussion of undulations of entropic origin to a later point in the chapter. Hence, we initially neglect all entropic effects or, equivalently, consider the membrane to be at zero temperature. Based on this assumption, consider an infinitesimal plate initially of area $A_0 = L_0^2$ that is deformed by a uniform normal stress $\tau_{11} = \sigma_1 = \tau_{22} = \sigma_2$ applied to its edges (Fig. 31.9a) to a new area A. The expressions relating stress and strain in two dimensions, and in the absence of stresses normal to the x_1–x_2 plane, can be written as

$$\sigma_\alpha = \frac{E}{1-\nu^2}(\varepsilon_\alpha + \nu\varepsilon_\beta)\,, \tag{31.18}$$

where the length of one edge is $L_\alpha = L_0(1+\varepsilon_\alpha)$; note that the subscripts α, β are used here rather than i, j to distinguish stresses and strains in two dimensions from those, more generally, in three. Thus, whereas i and j can be either 1, 2 or 3, α and β are restricted to being either 1 or 2. When this stress is uniform in the plane of the membrane (the x_1–x_2 plane) it can be replaced, without loss of generality, by a surface tension N_α (force per

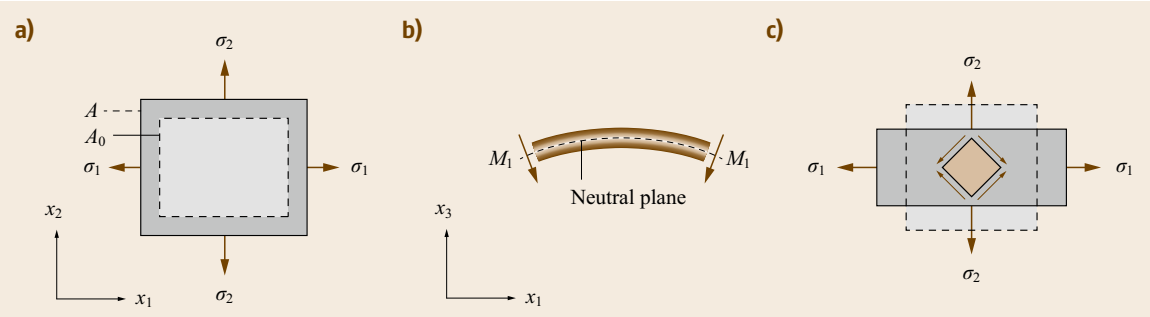


Fig. 31.9a–c Membrane mechanics. **(a)** A membrane, initially of area A_0, subjected to a uniform extensional stress along its edges ($\sigma_1 = \sigma_2$), causing an increase in area to A. **(b)** Application of equal moments M_1 at the two edges of a membrane sheet. **(c)** Shear deformations produced by the application of nonequal surface tensions on the boundaries of a membrane sheet. The inner square experiences shear strains as shown by the *arrows*

unit length) defined as $\sigma_\alpha h$, where h is the thickness of the membrane. These can be combined in the case when $N_1 = N_2 =$ constant and, consequently, $\varepsilon_1 = \varepsilon_2 = \varepsilon$ to give

$$N = \frac{Eh}{1-\nu}\varepsilon \,. \tag{31.19}$$

In terms of a plate stretched uniformly in both directions, we can define the areal strain as

$$\frac{\Delta A}{A_0} = \frac{A - A_0}{A_0} = \frac{L_0^2(1+\varepsilon)^2 - L_0^2}{L_0^2} \cong 2\varepsilon \,, \tag{31.20}$$

where the last approximation is appropriate for small strains. By combining this result with (31.19), we can define the *area expansion modulus* K_e (units of N/m) using

$$N = \frac{Eh}{2(1-\nu)}\frac{\Delta A}{A_0} \equiv K_e \frac{\Delta A}{A_0} \,. \tag{31.21}$$

Note that, although we used the continuum structural equations in our analysis, the final result can also be viewed as simply the definition of the area expansion modulus and applies regardless of whether or not the membrane can be modeled as a continuum.

Experimental measurements of K_e lie in the range of 0.1–1 N/m for various types of lipid bilayers and ≈ 0.45 N/m (450 dyn/cm) for red blood cell membranes [31.91]. These numbers suggest that cell membranes are quite resistant to extension and, for that reason, are often treated as inextensible. This discussion neglects the effects of thermal fluctuations in the membrane that give rise to a much more compliant behavior at the smallest areal strains. When surface stress is sufficient to smooth out most thermal fluctuations, the cell or vesicle will exhibit the large moduli given here.

This high resistance to area change is in large part due to the energy penalty associated with exposing the hydrophobic core of the membrane to water that occurs as the spacing between individual amphiphilic molecules is increased (for a detailed description of bilayer structure and thermodynamics, see [31.92]). Continuing to increase extensional stress, the lipid bilayer eventually ruptures, but at very small extensional strains, in the vicinity of 2–3% [31.93]. Note that a bilayer in a lipid vesicle, for example, stretches primarily by increasing the area per molecule since recruitment of additional material to the membrane occurs very slowly.

Using these expressions, the surface tension at which the membrane would rupture can be computed to be ≈ 0.06 N/m if we use a value near the higher end of the observed range $K_e = 1$ N/m. Cells also often exhibit an intrinsic surface tension. Reported values are small, however, lying in the range of $\approx 10^{-5}$–10^{-4} N/m [31.94].

Pure Bending. By contrast, lipid bilayers exhibit very low bending stiffness, so low that it is often neglected in models of membrane mechanics. It can be important in certain situations, however, and is essential, for example, in analyzing the thermal fluctuations of vesicles. Bending stiffness arises from the same type of molecular interactions that cause extensional stiffness. When an initially flat bilayer is bent, the hydrophilic head groups on the outside of the bend move further apart, while on the inside, intermolecular spacing decreases; both represent departures from the equilibrium, unstressed state and require energy.

Returning to our simple continuum plate model, consider a bending moment applied to the two ends, causing the plate to curve slightly (Fig. 31.9b). If the bending is due to moments applied at the two ends about the x_2-axis, then the bending moment per unit

length is related to the deflection by

$$M_\alpha = -\frac{Et^3}{12(1-\nu^2)}\left(\frac{\partial^2 u_3}{\partial x_\alpha^2}\right) = -K_\mathrm{b}\left(\frac{\partial^2 u_3}{\partial x_\alpha^2}\right), \tag{31.22}$$

where K_b is termed the *bending stiffness* having units of N m. Implicit in this expression are the assumptions that there exists a mid-plane (*neutral plane*) on which the in-plane stress and strain are both zero, and that straight lines perpendicular to this mid-plane remain straight and normal to this surface after deformation.

Typical values for the bending stiffness K_b lie in the range of 10^{-19} N m (10^{-12} dyn cm) for a red blood cell or lipid bilayer [31.93]. This value is larger, on the order of $(1–2)\times 10^{-18}$ N m [31.95], for other cell types (e.g., neutrophils, endothelial cells) that possess a more extensive cortex.

Pure Shear. Shear deformations arise when a membrane is stretched in one direction by a surface tension N_1 (units of force/length) while the lateral surface contracts under a lesser tension N_2, at constant surface area and in the absence of bending (Fig. 31.9c). Surfaces oriented at 45° to the boundaries experience pure shear stresses of magnitude $(N_1 - N_2)/2$. When subjected to shear stresses in the plane of the membrane, a pure lipid bilayer behaves essentially as a liquid. It exhibits a membrane viscosity in that it poses a resisting force proportional to the *rate* of shear deformation, but only a small shear modulus to *static* shear deformations. It is not clear, in fact, whether or not pure lipid bilayers exhibit a nonzero shear modulus. Typical cell membranes *do* exhibit a shear modulus, however, largely due to the cortex of cytoskeletal filaments that lie on the intracellular side of the membrane. In a red blood cell, this matrix, as discussed above, consists of interconnected filamentous spectrin and actin with attachments to the membrane via ankyrin. The equations relating shear force per unit length of membrane to shear deformation (Hooke's law) can be expressed as $N_{12} = \tau_{12}h = 2G\varepsilon_{12} = K_\mathrm{s}\varepsilon_{12}$, where we define the membrane shear modulus K_s in units of N/m. Typical values lie in the range $(6–9)\times 10^{-6}$ N/m for a red blood cell membrane. A pure lipid bilayer exhibits a *viscous* resistance to shear deformations characterized by a shear viscosity of $\approx 10^{-6}$ N s/m [31.48].

31.5.3 Cell Nuclei

Measurements based on micropipette aspiration, atomic force microscopy, and substrate strain experiments suggest that the nucleus is $\approx 2–10$ times stiffer than the surrounding cytoskeleton [31.56, 96–99]. What makes the nucleus so hard to deform? This question may in part depend on the particular mechanical load application. In cellular compression experiments, the highly condensed chromatin provides a large resistance to nuclear compression. In contrast, the nuclear lamina can act as an elastic molecular shock absorber [31.100] that seems ideally suited to sustain large strain when the cell or nucleus is stretched. This idea is supported by micropipette aspiration experiments on isolated nuclei that reveal that, in condensed nuclei, chromatin is the major contributor to nuclear stiffness, whereas the nuclear lamina carries much of the applied mechanical load in swollen nuclei that closely resemble nuclei in intact cells [31.43]. Similarly, condensed nuclei were found to be much stiffer than swollen nuclei [31.43]. Based on the nuclear structure, the mechanical contributions of the nucleus can be separated into three interconnected components: the nuclear membranes, the nuclear lamina, and the nuclear interior, which is mostly composed of chromatin but could also contain a nuclear matrix of lamins, actins, and other proteins. Each of these structural components has its own physical properties. The specific contributions have been addressed in experiments that induced osmotic swelling of isolated nuclei by varying the buffer salt concentration to separate the nuclear envelope from the underlying chromatin [31.12, 43] and in experiments on the large nuclei of *Xenopus* oocytes [31.101]. Additional insights have come from micropipette aspiration experiments, in which specific components of the nucleus (e.g., the lamina, the nuclear membrane or the chromatin) are fluorescently labeled and monitored during aspiration. These experiments have revealed that the nuclear interior behaves as a compressible, viscoelastic gel and can be compacted by more than 60% during micropipette aspiration before becoming resistant to further compression [31.46]. The reduction in nuclear volume during micropipette aspiration results in buckling of the nuclear envelope outside the pipette [31.43, 46], further demonstrating tight physical attachment between the nuclear envelope and the chromatin interior. Nucleoli and other intranuclear structures deform along with the nuclear interior in these experiments [31.45, 46], but might have slightly different mechanical stiffness. The viscoelastic nature of the chromatin also appears responsible for the persistent plastic deformations of nuclei seen after micropipette aspiration of isolated nuclei [31.45, 102]. In contrast to the gel-like nuclear interior, the nuclear lamina itself behaves as a two-dimensional elastic shell. This can be illustrated when monitoring the fluorescence intensity of green fluorescent protein (GFP)–lamin A incorporated into the nuclear lamina during micropipette aspiration experiments. While the fluorescence intensity is

very uniform in the nonaspirated section of the nucleus, the fluorescence intensity decays exponentially inside the aspirated segment, as the nuclear lamina is stretched under the aspiration pressure. Such a fluorescence intensity profile can be well fit by the predicted strain profile of a solid elastic shell [31.46, 103]. Unlike the nuclear lamina, the nuclear membranes have fluid-like characteristics; i. e., they do not show any induced strain during micropipette aspiration as they can flow in response to pressure application [31.46]. Further illustrating the fluid-like nature is the fact that small membrane vesicles continue to form blebs in the aspirated section of the nucleus as long as the suction pressure persists [31.46]. Consequently, the nuclear membrane cannot carry much of the mechanical load, leaving the nuclear interior and the nuclear lamina as the major contributors to nuclear stiffness and stability. This interpretation is further supported by recent studies that demonstrate that compression of cells using microindenters or during migration through tight spaces can result in nuclear membrane blebbing and subsequent rupture at sites of local defects in the nuclear lamina [31.104, 105].

In addition to the various structural components, the differentiation state of a cell can also modulate the mechanical properties of its nucleus. A recent study that probed nuclei from human embryonic stem cells and differentiated cells found that nuclei from embryonic stem cells were significantly more deformable than nuclei from differentiated cells and became up to sixfold stiffer relative to the cytoplasm during differentiation [31.45]. While some of this increase in nuclear stiffness could be caused by changes in chromatin structure and organization, much of it can be attributed to a stiffer lamina through increased expression of lamins A and C, as independent experiments have shown that lamins A and C are the main contributors to nuclear stiffness [31.109, 110].

Overall, micropipette aspiration has been the most widely used technique to probe nuclear mechanics, although cellular strain experiments, AFM, microbead rheology, and parallel-plate compression experiments have also been successfully applied to measure nuclear stiffness (Table 31.4). While some groups have described continuous creep of aspirated nuclei into the micropipette [31.43, 45, 100, 111], others [31.46, 97, 103] found that the nucleus stabilized after an equilibration period of $\approx 10\,\mathrm{s}$. These differences may be due to variations in the experimental setup or the particular cell type studied. Depending on the experimental technique used, various mathematical models have been applied to describe the induced nuclear deformations. For micropipette aspiration experiments, the creep compliance is often inferred from the following equation

$$J(t, \Delta P) = \Phi \frac{4\pi}{3} \frac{L}{D} \frac{1}{\Delta P} , \tag{31.23}$$

where ΔP is the applied pressure difference, L is the aspirated tongue length, and D is the micropipette diameter, and the dimensionless parameter Φ (value = 2.1) accounts for the effects of the micropipette geometry [31.112]. The challenge in finding an analytical expression for the mechanical behavior of the nucleus is that the nucleus is composed of various, structurally very different domains, as discussed above. Finite-element modeling can address some of these complex interactions, but often has to make assumptions on the specific material parameters that are difficult to verify. An alternative approach is to directly compare nuclear deformations under identical experimental conditions, e.g., the aspirated lengths of the nucleus during a constant applied pressure gradient in micropipette aspiration experiments or the induced nuclear strain under a given applied substrate strain in cell strain experiments [31.52].

31.5.4 Mechanosensing Proteins

While the complex biochemical pathways have been explored in considerable detail by biologists during the past several decades, work has only recently begun to understand the mechanical intracellular pathways for the propagation of force through the cell; For example, *Geiger* and *Bershadsky* [31.113] mapped the complex and interconnected pathways that can be

Table 31.4 Reported values for nuclear stiffness. (* Power-law rheology, not single spring constant; the nuclear envelope expansion modulus was measured to be $\approx 325\,\mathrm{mN/m}$ for these cells, after [31.43])

Cell type	Experimental method	Nuclear stiffness	Reference
Chondrocytes	Micropipette aspiration	1–2 kPa	[31.97]
Bovine endothelial cells	Parallel-plate compression	5–8 kPa	[31.96]
Monkey kidney epithelial cells	AFM and micropipette aspiration	$\approx$ 1–10 kPa*	[31.43]
Fibrosarcoma cells (HT1080)	AFM	0.2–2.5 kPa	[31.106]
Bladder cancer cell lines (RT4 and T24)	AFM	6–8 kPa	[31.98]
Mouse fibroblasts	Intranuclear particle tracking	18 Pa	[31.107]
Human cervical cancer cells (HeLa)	Magnetic tweezers	250 Pa	[31.108]

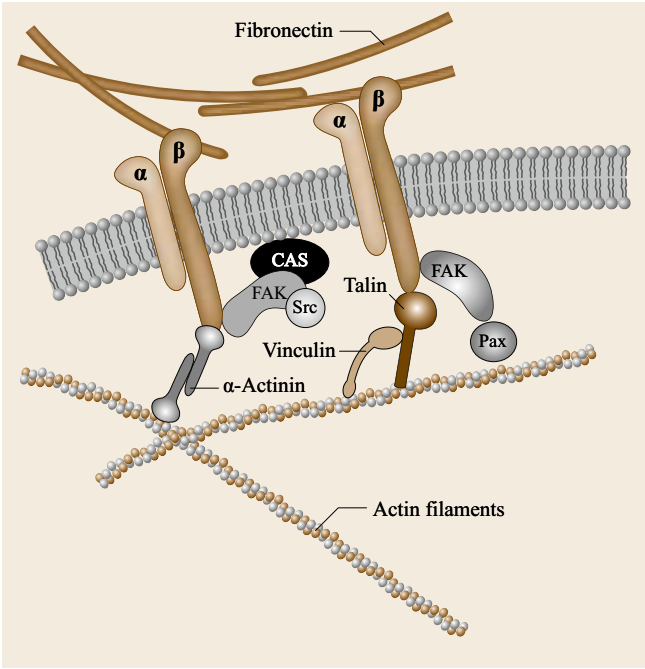


Fig. 31.10 A small sampling of the proteins found in a focal adhesion complex. Forces are typically transmitted from the extracellular matrix (e.g., fibronectin), via the integral membrane adhesion receptors (α- and β-integrins), various membrane-associated proteins (focal adhesion kinase (FAK), paxillin (Pax), talin, Crk-associated substrate (CAS)), to actin-binding proteins (α-actinin) that link the focal adhesion complex to the cytoskeleton. Note that the exact positioning of proteins with the focal adhesion is a matter of ongoing research, and be either closer to or further from the membrane than shown (after [31.113], with permission)

traced from the extracellular domain to the cytoskeletal matrix within the confines of a focal adhesion, a small portion of which is depicted in Fig. 31.10. Forces transmitted to the cell, either by tethering to the extracellular matrix or via ligand-coated beads in various in vitro experiments, are transmitted via the integrins (or other transmembrane receptors) through a collection of membrane-associated proteins, ultimately linking up with the cytoskeleton and propagated throughout the cell. Many of the proteins that form this pathway are known signaling molecules, and several have been implicated in mechanotransduction. The process by which changes in protein conformation give rise to protein clustering at a focal adhesion or initiate intracellular signaling, however, remains largely unknown. This has called for close examination of the individual proteins or protein complexes that transmit the forces throughout the cell and their interconnections. Consider one example pathway from Fig. 31.11. It can be seen that integrins bind to talin, which in turn connects with F-actin either directly or via vinculin. Talin and integrin also bind to focal adhesion kinase (FAK). This force pathway alone gives rise to several possibilities for mechanotransduction.

Even in the absence of biochemical signals, mechanical stimulation can still elicit the formation of viable focal adhesions [31.114]. These focal adhesions are produced as a result of direct molecular conformational changes in focal adhesion-forming molecules. Force-induced conformational changes in talin, vinculin, and focal adhesion kinase (FAK) initiate focal adhesion formation by anchoring integrin molecules to actin filaments and recruiting other focal adhesion-forming molecules to the site of interaction. Other mechanosensing molecules such as filamin and α-actinin stabilize actin filaments through mechanobracing and making scaffolds. The conformational changes in the focal adhesion-initiating molecules and the actin filament scaffold bracing molecules demonstrate the mechanisms of molecular mechanotransduction. In the following, we briefly review some recent work in molecular dynamics simulation of some mechanosensing proteins, namely talin interaction with vinculin, focal adhesion kinase interaction with paxillin, and molecular mechanics of actin binding proteins α-actinin and filamin.

Integrins

Integrins are signal receptors within the cell membrane that is first encountered by the cell mechanical microenvironment. Integrins act as a bridge between the cytoskeleton and ECM by transferring signals bidirectionally into and out of the cell. Several diseases, such as atherosclerosis and different types of cancer, have been linked to altered signaling patterns of integrins. Structurally, integrins are α–β heterodimeric receptors that consist of large extracellular domains (700−1000 residues), two transmembrane α-helices, and short cytosolic tails (50–70 residues). Integrin receptors reside in the cell membrane and link the cytoskeleton to ECM, making them essential for both *inside-out* and *outside-in* signaling. Signal transmission from the cell exterior to the cytoskeleton is called "outside-in signaling," whereas "inside-out signaling" occurs when a biochemical signal is relayed from the cytoskeleton outwards by inducing a cascade of conformational changes within mechanosensing proteins on the mechanical pathways that link the cytoskeleton to the ECM. Inside-out signaling is putatively triggered by separation of the two integrin subunits at their cytoplasmic and transmembrane domains, which occurs in response to association of talin to the integrin cytoplasmic β-subunit domain. A major restraint that holds

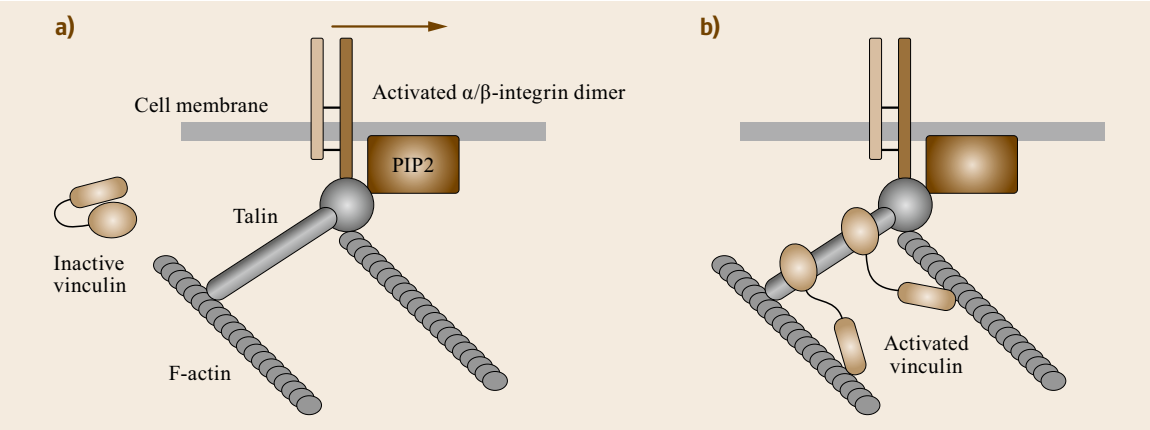


Fig. 31.11 **(a)** Model of the initial adhesion consisting of integrin–talin–F-actin linkage. Vinculin is present in the cytosol in an inactive, autoinhibitory conformation, and tensile force is applied to the integrin dimer from the outside of the cell membrane. **(b)** Transmitted force through the linkage alters the talin configuration, and recruits vinculin to reinforce the initial adhesion linkage (PIP2 – phosphatidylinositol 4,5-bisphosphate)

integrins in an inactive mode is the interaction between transmembrane and cytoplasmic domains of the α- and β-subunits. Separation of these domains is sufficient to allow unbending of the ligand-binding headpiece and conformational changes that increase ligand-binding affinity. Conversely, binding of the ligand to the integrin ectodomain results in an extended conformation of integrin whereby the legs of α- and β-subunits are separated. Hence, integrin-mediated signaling involves conformational changes of integrins leading to alterations in their binding affinity for ECM ligands and may be initiated from interacting with proteins at their cytoplasmic tail.

Talin and Vinculin

Talin interacts with the membrane-bound integrin molecule at its head domain, and with vinculin in its tail domain. It has been implicated as a site of molecular activation by external force. The forces acting on the cells from the ECM are transmitted from the integrin molecules to the talin head and then to all regions of the talin tail (Fig. 31.11). The talin tail has 11 vinculin binding sites (VBSs), some of which are on amphipathic α-helices and are buried in the hydrophobic core of the talin tail. VBS1 is one of the cryptic binding sites buried in the hydrophobic core of the TAL5 region of talin's tail. TAL5 has several surface polar residues that are sites of hydrogen bonding with other secondary structures in the rod domain. Transduced external forces from the ECM are applied to the TAL5 domain via these residues [31.115]. In order to simulate the force transduction through the cryptic VBS1 of TAL5, *Lee* et al. [31.116] applied forces averaging 21.2 pN at the residues near the N-terminal near the talin head while

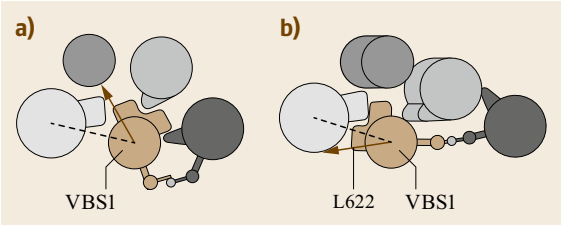


Fig. 31.12a,b With external force applied to the polar residues in the TAL5 region of the talin rod, domain activation of the VBS1 is achieved through rotation of the buried VBS residues to the surface of the molecule. **(a)** TAL5 region before application of external force. The VBS1 is buried in the hydrophobic patch. **(b)** After the external force is applied, TAL5 undergoes a conformational change and the VBS residues are exposed to the surface. In this representation, movement of the VBS to the surface is shown by the movement of the *arrow* (edge of VBS) past the *dotted lines*

constraining fixed the polar residues at the C-terminal, which interact with the adjacent secondary structure domain. The external locally applied force caused TAL5 to elongate, H1 to move in the direction of the pulling force, H1 to apply torque to H4 via its hydrogen bonds, and H4 to rotate, exposing the VBS1 (Fig. 31.12). Once exposed to solvent, the VBS1 is activated and can interact with vinculin. The force-induced activation of VBS sites in the talin tail together with force-induced activation of vinculin can anchor the stimulated integrin molecules to the actin cytoskeleton and thereby initiate focal adhesions.

An alternative mechanism was recently proposed by *Hytönen* and *Vogel* [31.117], which involves the unfold-

ing of the rod domain in order to expose the vinculin binding site VBS1. Using steered molecular dynamics, Hytönen and Vogel demonstrated that the talin rod can be fragmented into three helix subbundles, which is followed by the sequential exposure of vinculin-binding helices to water. The unfolding of a vinculin-binding helix into a completely stretched polypeptide might then inhibit further binding of vinculin.

The conformational changes that enable the recruitment of vinculin to ECM-bound talin can go beyond the exposure of cryptic vinculin binding sites. The orientation of the talin dimer itself can be affected by its mechanical environment. Recent molecular dynamics simulations [31.118] have probed these possible dimer orientation changes. Molecular dynamics has shown that the talin dimer will adopt either an elongated conformation, when bound to two integrins under extension, or a compact conformation, when bound to two integrins free of external strain. The elongated conformation then allows for further recruitment of vinculin by exposure of additional regions on the talin rod domain with cryptic binding sites. The combined reorientation of the dimer and the exposure of the cryptic binding sites vastly enhances the recruitment of vinculin to the growing focal adhesion.

The mechanisms via which vinculin reinforces contacts between the ECM and F-actin through interaction with talin have also been explored [31.119–121]. Vinculin is natively in an autoinhibited conformation in which the VBS binding region of vinculin head domain (Vh) prevents interaction with the vinculin tail domain (Vt) by steric hindrance. Competing hypotheses have been suggested for how vinculin is activated via interaction with talin and actin: (A) vinculin activation after simultaneous interaction with talin and actin, or (B) vinculin activation solely by interaction with talin VBS. Recent molecular dynamics models seem to suggest vinculin is more likely to become activated through simultaneous interaction rather than through talin VBS insertion alone [31.122, 123]. It has been postulated that the activation of vinculin and its binding with talin's VBS could involve the following mechanism:

1. VBS links to D1 via surface hydrophobic interactions.
2. Vinculin undergoes activation, and D1 is moved away from the vinculin tail.
3. Helices in D1 undergo conformational change to allow VBS binding.
4. VBS inserts itself into the hydrophobic core of D1.

One major actin filament cross-linker is α-actinin, which maintains dynamic rearrangement of actin filaments and serves as a scaffold for parallel actin filaments, such as those in stress fibers or in muscle z-disk formations. α-Actinin is an antiparallel homodimer with three major domains: an actin binding domain, a calmodulin homology domain, and a central rod domain [31.124]. In order to act as a molecular scaffold, cross-linking actin filaments in a stressed cytoskeletal network α-actinin must have a semiflexible structure. By exploring the natural frequencies of α-actinin and its response to external forces, *Golji* et al. [31.119] characterized the flexible and rigid regions of α-actinin and demonstrated the molecular conformational changes that underlie the semiflexible cross-linking and mechanical bracing of α-actinin. Besides its cross-linking role for filamentous actin, α-actinin binds with integrin as well and is speculated to play a role in the activation and clustering of integrins and thus focal adhesion formation. However, it is debated whether α-actinin plays an inhibitory or facilitating role in integrin activation. Detailed molecular interactions of α-actinin and integrin cytoplasmic tail have been probed in order to reconcile this controversy [31.120]. While talin is essential for focal adhesion formation, it does not play a significant role in focal adhesion maturation, and other molecules such as α-actinin are responsible for that. Since α-actinin is an antiparallel dimer and each monomer contains an integrin-binding site at the N-terminal end of its rod domain, it can bind to two integrin tails simultaneously, which may give rise to integrin clustering. Integrin clustering may occur via maintaining the distance between two α-actinin-bound integrins, allowing them to interact more effectively. Integrins diffuse freely in the lipid membrane, thus the probability of assembling individual integrins is only related to their concentration in the membrane. Directional motion of integrins may highly influence the time scale of their clustering, leading to adhesion maturation. α-Actinin may play a role in this by binding to actins at both ends and transmitting actomyosin forces to the plasma membrane through their interaction with integrins, which leads to directional migration of integrins governed by cytoskeletal forces [31.121, 125].

Microtubules

As yet another mechanosensing protein system, one can name the microtubule filaments [31.126]. The microtubule cytoskeleton, which is composed of individual microtubules typically nucleated from a microtubule organizing center near the nucleus, usually forms a radial network that radiates outward to the cell periphery. These microtubules are inherently polarized due to their head-to-tail assembly, and this allows for processive movement of cargo via motor proteins that move specifically towards either the plus or the minus end.

Microtubular filaments resist mechanical deformation and facilitate transport of intracellular cargo via coupling to molecular motors. These filaments are typically in compression, effectively balancing the tensile forces exhibited by the actin–myosin system [31.84].

Stretch-Sensitive Ion Channels

Another class of mechanosensing proteins are the mechanosensitive ion channels, e.g., the mechanosensitive channel of large conductance (MscL), which has been studied extensively [31.127, 128]; molecular dynamics simulation has been used to show how stresses in the cell membrane act directly on the channel and cause it to change its conductance [31.129]. A recently discovered cation channel, Piezo2, is present in mechanosensory neurons and important for proprioception [31.130]. These channels have emerged as another potential gateway for the cell to sense and respond to environmental stimuli.

The Nucleus as a Cellular Mechanosensor

The nucleus has long been speculated to serve as a cellular mechanosensor, and functional loss of nuclear envelope proteins such as lamins and nesprins results in impaired activation of mechanosensitive genes [31.32, 131]. Force-induced deformation of the nucleus could modulate gene expression through a variety of mechanisms, including:

1. Changing the accessibility of chromatin to transcriptional regulators
2. Causing conformational changes in nuclear envelope proteins that alter their interaction with chromatin and other binding partners
3. Increasing/decreasing the permeability of nuclear pores
4. Perturbing mobility and interaction of nuclear proteins and intranuclear structures (reviewed in [31.132, 133]).

Nuclear swelling and associated nuclear membrane strain have also been shown to activate proinflammatory signals in response to tissue damage, possibly by stretch-induced changes in the nuclear lipid membrane that facilitate recruitment of cytosolic phospholipase A2 ($cPLA_2$) to the nuclear envelope [31.134]. The most striking evidence for nuclear mechanosensing comes from studies on isolated nuclei subjected to nanonewton forces by magnetic beads attached via nesprin-1. Repeated force application resulted in rapid nuclear stiffening and recruitment of lamins A/C to the LINC complex, mediated by phosphorylation of the nuclear envelope protein emerin through Src kinase [31.135]. Understanding the mechanisms of force transmission and the stability of LINC complexes is critically important in unravelling the cellular mechanotransduction pathways.

31.6 Current Understanding and Future Needs

Progress is rapidly being made on both the experimental and computational fronts to further understand the nanomechanics of the cell. Using either AFM or optical tweezers, controlled force applications in the pN range and displacement measures in the nm range are already within current capabilities. Both methods have been used extensively to measure the rupture force of two bound proteins [31.136] or the force needed to unfold a protein and change its binding or signaling properties. In addition, single-molecule fluorescence measurements provide the opportunity to monitor single binding events. On a larger scale, new entirely noninvasive methods are being developed to probe the mechanical properties of subcellular structures. Since we know cells are exquisitely sensitive to mechanical force, these methods will help us to distinguish the passive and active mechanical responses. On the computational side, technical and technological barriers are being slowly surmounted. Molecular dynamics can be applied to proteins to predict their conformational change under force, and docking simulations provide a means for determining binding affinities in different conformational states. The barriers to progress lie primarily in our lack of atomistic models capable of simulating interactions between proteins such as those located in the force transmission pathway, at focal adhesions or nuclear pore complexes, for example. These proteins also tend to be difficult to crystallize, so few structures are available. Cryoelectron microscopy (cryo-EM) is emerging as a promising approach to yield structural information without the requirement to form crystals, which makes it particularly attractive to study filaments and large, complex structures such as the nuclear pore complex. For computational modeling, due to their size, simulations are computationally intensive, especially if the presence of water molecules is included explicitly. Despite these constraints, some progress can be made by using subdomains of the proteins of interest or coarse-graining and multiscale approaches, although these need to be adequately evaluated [31.137].

References

31.1 H. Huang, R.D. Kamm, R.T. Lee: Cell mechanics and mechanotransduction: pathways, probes, and physiology, Am. J. Physiol. Cell Physiol. **287**(1), C1–C11 (2004)

31.2 P.A. Janmey, D.A. Weitz: Dealing with mechanics: mechanisms of force transduction in cells, Trends Biochem. Sci. **29**(7), 364–370 (2004)

31.3 M.J. Paszek, C.C. DuFort, O. Rossier, R. Bainer, J.K. Mouw, K. Godula, J.E. Hudak, J.N. Lakins, A.C. Wijekoon, L. Cassereau, M.G. Rubashkin, M.J. Magbanua, K.S. Thorn, M.W. Davidson, H.S. Rugo, J.W. Park, D.A. Hammer, G. Giannone, C.R. Bertozzi, V.M. Weaver: The cancer glycocalyx mechanically primes integrin-mediated growth and survival, Nature **511**(7509), 319–325 (2014)

31.4 V. Vogel, M. Sheetz: Local force and geometry sensing regulate cell functions, Nat. Rev. Mol. Cell Biol. **7**(4), 265–275 (2006)

31.5 D. Manoussaki, W.D. Shin, C.M. Waterman, R.S. Chadwick: Cytosolic pressure provides a propulsive force comparable to actin polymerization during lamellipod protrusion, Sci. Rep. **5**, 12314 (2015)

31.6 P. Isermann, J. Lammerding: Nuclear mechanics and mechanotransduction in health and disease, Curr. Biol. **23**(24), R1113–R1121 (2013)

31.7 H. Lodish, C.A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, A. Amon, K.C. Martin: *Molecular Cell Biology*, 8th edn. (W.H. Freeman, London 2016)

31.8 M.J. Paszek, D. Boettiger, V.M. Weaver, D.A. Hammer: Integrin clustering is driven by mechanical resistance from the glycocalyx and the substrate, PLoS Comput. Biol. **5**(12), e1000604 (2009)

31.9 J. Howard: *Mechanics of Motor Proteins and the Cytoskeleton* (S. Sinauer, Sunderland 2001) pp. 288–289

31.10 T.M. Svitkina, G.G. Borisy: Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia, J. Cell. Biol. **145**(5), 1009–1026 (1999)

31.11 H. Higuchi, Y.E. Goldman: Sliding distance per ATP molecule hydrolyzed by myosin heads during isotonic shortening of skinned muscle fibers, Biophys. J. **69**(4), 1491–1507 (1995)

31.12 Y. Tsuda, H. Yasutake, A. Ishijima, T. Yanagida: Torsional rigidity of single actin filaments and actin-actin bond breaking force under torsion measured directly by in vitro micromanipulation, Proc. Natl. Acad. Sci. USA **93**(23), 12937–12942 (1996)

31.13 R. Yasuda, H. Miyata, K. Kinosita Jr.: Direct measurement of the torsional rigidity of single actin filaments, J. Mol. Biol. **263**(2), 227–236 (1996)

31.14 C.P. Brangwynne, G.H. Koenderink, E. Barry, Z. Dogic, F.C. MacKintosh, D.A. Weitz: Bending dynamics of fluctuating biopolymers probed by automated high-resolution filament tracking, Biophys. J. **93**(1), 346–359 (2007)

31.15 F. Gittes, B. Mickey, J. Nettleton, J. Howard: Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape, J. Cell Biol. **120**(4), 923–934 (1993)

31.16 E. Fuchs, D.W. Cleveland: A structural scaffolding of intermediate filaments in health and disease, Science **279**(5350), 514–519 (1998)

31.17 D.A. Parry, S.V. Strelkov, P. Burkhard, U. Aebi, H. Herrmann: Towards a molecular description of intermediate filament structure and assembly, Exp. Cell Res. **313**(10), 2204–2216 (2007)

31.18 M. Kirschner, E. Schulze: Morphogenesis and the control of microtubule dynamics in cells, J. Cell Sci. Suppl. **5**, 293–310 (1986)

31.19 M. Cohen, K.K. Lee, K.L. Wilson, Y. Gruenbaum: Transcriptional repression, apoptosis, human disease and the functional evolution of the nuclear lamina, Trends Biochem. Sci. **26**(1), 41–47 (2001)

31.20 F. Alber, S. Dokudovskaya, L.M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, A. Sali, M.P. Rout: The molecular architecture of the nuclear pore complex, Nature **450**(7170), 695–701 (2007)

31.21 G. Kabachinski, T.U. Schwartz: The nuclear pore complex–structure and function at a glance, J. Cell Sci. **128**(3), 423–429 (2015)

31.22 J.L. Broers, H.J. Kuijpers, C. Ostlund, H.J. Worman, J. Endert, F.C. Ramaekers: Both lamin A and lamin C mutations cause lamina instability as well as loss of internal nuclear lamin organization, Exp. Cell Res. **304**(2), 582–592 (2005)

31.23 T. Pederson, U. Aebi: Nuclear actin extends, with no contraction in sight, Mol. Biol. Cell **16**(11), 5055–5060 (2005)

31.24 W.A. Hofmann, T. Johnson, M. Klapczynski, J.L. Fan, P. de Lanerolle: From transcription to transport: emerging roles for nuclear myosin I, Biochem. Cell Biol. **84**(4), 418–426 (2006)

31.25 W.A. Hofmann, L. Stojiljkovic, B. Fuchsova, G.M. Vargas, E. Mavrommatis, V. Philimonenko, K. Kysela, J.A. Goodrich, J.L. Lessard, T.J. Hope, P. Hozak, P. de Lanerolle: Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II, Nat. Cell Biol. **6**(11), 1094–1101 (2004)

31.26 K.G. Young, R. Kothary: Spectrin repeat proteins in the nucleus, Bioessays **27**(2), 144–152 (2005)

31.27 G.G. Gundersen, H.J. Worman: Nuclear positioning, Cell **152**(6), 1376–1389 (2013)

31.28 D.I. Kim, K.C. Birendra, K.J. Roux: Making the LINC: SUN and KASH protein interactions, Biol. Chem. **396**(4), 295–310 (2015)

31.29 W. Chang, H.J. Worman, G.G. Gundersen: Accessorizing and anchoring the LINC complex for multifunctionality, J. Cell Biol. **208**(1), 11–22 (2015)

31.30 B.A. Sosa, U. Kutay, T.U. Schwartz: Structural insights into LINC complexes, Curr. Opin. Struct. Biol. **23**(2), 285–291 (2013)

31.31 J.L. Broers, E.A. Peeters, H.J. Kuijpers, J. Endert, C.V. Bouten, C.W. Oomens, F.P. Baaijens, F.C. Ramaekers: Decreased mechanical stiffness in LMNA-/- cells is caused by defective nucleocytoskeletal integrity: implications for the development of laminopathies, Hum. Mol. Genet. **13**(21), 2567–2580 (2004)

31.32 J. Lammerding, P.C. Schulze, T. Takahashi, S. Kozlov, T. Sullivan, R.D. Kamm, C.L. Stewart, R.T. Lee: Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction, J. Clin. Invest. **113**(3), 370–378 (2004)

31.33 J.S. Lee, C.M. Hale, P. Panorchan, S.B. Khatau, J.P. George, Y. Tseng, C.L. Stewart, D. Hodzic, D. Wirtz: Nuclear lamin A/C deficiency induces defects in cell mechanics, polarization, and migration, Biophys. J. **93**(7), 2542–2552 (2007)

31.34 G. Bonne, S. Quijano-Roy: Emery-Dreifuss muscular dystrophy, laminopathies, and other nuclear envelopathies, Handb. Clin. Neurol. **113**, 1367–1376 (2013)

31.35 M. Zwerger, D.E. Jaalouk, M.L. Lombardi, P. Isermann, M. Mauermann, G. Dialynas, H. Herrmann, L.L. Wallrath, J. Lammerding: Myopathic lamin mutations impair nuclear stability in cells and tissue and disrupt nucleo-cytoskeletal coupling, Hum. Mol. Genet. **22**(12), 2335–2349 (2013)

31.36 N. Zuela, M. Zwerger, T. Levin, O. Medalia, Y. Gruenbaum: Impaired mechanical response of an EDMD mutation leads to motility phenotypes that are repaired by loss of prenylation, J. Cell. Sci. **129**(9), 1781–1791 (2016)

31.37 E.S. Folker, C. Ostlund, G.W. Luxton, H.J. Worman, G.G. Gundersen: Lamin A variants that cause striated muscle disease are defective in anchoring transmembrane actin-associated nuclear lines for nuclear movement, Proc. Natl. Acad. Sci. USA **108**(1), 131–136 (2011)

31.38 M. Radmacher, J.P. Cleveland, M. Fritz, H.G. Hansma, P.K. Hansma: Mapping interaction forces with the atomic force microscope, Biophys. J. **66**(6), 2159–2165 (1994)

31.39 R.E. Mahaffy, S. Park, E. Gerde, J. Kas, C.K. Shih: Quantitative analysis of the viscoelastic properties of thin regions of fibroblasts using atomic force microscopy, Biophys. J. **86**(3), 1777–1793 (2004)

31.40 M. Dokukin, Y. Ablaeva, V. Kalaparthi, A. Seluanov, V. Gorbunova, I. Sokolov: Pericellular brush and mechanics of Guinea pig fibroblast cells studied with AFM, Biophys. J. **111**(1), 236–246 (2016)

31.41 N. Wang, J.P. Butler, D.E. Ingber: Mechanotransduction across the cell surface and through the cytoskeleton, Science **260**, 1124–1127 (1993)

31.42 N. Wang, D.E. Ingber: Control of cytoskeletal mechanics by extracellular matrix, cell shape, and mechanical tension, Biophys. J. **66**(6), 2181–2189 (1994)

31.43 K.N. Dahl, A.J. Engler, J.D. Pajerowski, D.E. Discher: Power-law rheology of isolated nuclei with deformation mapping of nuclear substructures, Biophys. J. **89**(4), 2855–2864 (2005)

31.44 D.E. Discher, N. Mohandas, E.A. Evans: Molecular maps of red cell deformation: hidden elasticity and in situ connectivity, Science **266**(5187), 1032–1035 (1994)

31.45 J.D. Pajerowski, K.N. Dahl, F.L. Zhong, P.J. Sammak, D.E. Discher: Physical plasticity of the nucleus in stem cell differentiation, Proc. Natl. Acad. Sci. USA **104**(40), 15619–15624 (2007)

31.46 A.C. Rowat, J. Lammerding, J.H. Ipsen: Mechanical properties of the cell nucleus and the effect of emerin deficiency, Biophys. J. **91**(12), 4649–4664 (2006)

31.47 E. Evans, B. Kukan: Passive material behavior of granulocytes based on large deformation and recovery after deformation tests, Blood **64**(5), 1028–1035 (1984)

31.48 E.A. Evans: Bending elastic modulus of red blood cell membrane derived from buckling instability in micropipet aspiration tests, Biophys. J. **43**(1), 27–30 (1983)

31.49 E.A. Evans, R.M. Hochmuth: Membrane viscoelasticity, Biophys. J. **16**(1), 1–11 (1976)

31.50 L.M. Lee, A.P. Liu: A microfluidic pipette array for mechanophenotyping of cancer cells and mechanical gating of mechanosensitive channels, Lab Chip **15**(1), 264–273 (2015)

31.51 Q. Guo, S. Park, H. Ma: Microfluidic micropipette aspiration for measuring the deformability of single cells, Lab Chip **12**(15), 2687–2695 (2012)

31.52 P. Isermann, P.M. Davidson, J.D. Sliz, J. Lammerding: Assays to measure nuclear mechanics in interphase cells, Curr. Protoc. Cell Biol. **22**(16), 1534–1546 (2012) doi:10.1002/0471143030.cb2216s56, Chapter 22

31.53 M.J. Rosenbluth, W.A. Lam, D.A. Fletcher: Analyzing cell mechanics in hematologic diseases with microfluidic biophysical flow cytometry, Lab Chip **8**(7), 1062–1070 (2008)

31.54 S. Byun, S. Son, D. Amodei, N. Cermak, J. Shaw, J.H. Kang, V.C. Hecht, M.M. Winslow, T. Jacks, P. Mallick, S.R. Manalis: Characterizing deformability and surface friction of cancer cells, Proc. Natl. Acad. Sci. USA **110**(19), 7580–7585 (2013)

31.55 H.T. Tse, D.R. Gossett, Y.S. Moon, M. Masaeli, M. Sohsman, Y. Ying, K. Mislick, R.P. Adams, J. Rao, D. Di Carlo: Quantitative diagnosis of malignant pleural effusions by single-cell mechanophenotyping, Sci. Transl. Med. **5**(212), 212ra163 (2013)

31.56 N. Caille, Y. Tardy, J.J. Meister: Assessment of strain field in endothelial cells subjected to uniaxial deformation of their substrate, Ann. Biomed. Eng. **26**(3), 409–416 (1998)

31.57 N. Mohandas, R.M. Hochmuth, E.E. Spaeth: Adhesion of red cells to foreign surfaces in the presence of flow, J. Biomed. Mater. Res. **8**(2), 119–136 (1974)

31.58 O. Otto, P. Rosendahl, A. Mietke, S. Golfier, C. Herold, D. Klaue, S. Girardo, S. Pagliara, A. Ekpenyong, A. Jacobi, M. Wobus, N. Topfner, U.F. Keyser, J. Mansfeld, E. Fischer-Friedrich, J. Guck: Real-time deformability cytometry: on-the-fly cell mechanical phenotyping, Nat. Methods **12**(3), 199–202 (2015), 4 p following 202

31.59 D.R. Gossett, H.T. Tse, S.A. Lee, Y. Ying, A.G. Lindgren, O.O. Yang, J. Rao, A.T. Clark, D. Di Carlo: Hydrodynamic stretching of single cells for large population mechanical phenotyping, Proc. Natl. Acad. Sci. USA **109**(20), 7630–7635 (2012)

31.60 T.G. Mason, K. Ganesan, J.H. van Zanten, D. Wirtz, S.C. Kuo: Particle tracking microrheology of complex fluids, Phys. Rev. Lett. **79**(17), 3282–3285 (1997)

31.61 D. Wirtz: Particle-tracking microrheology of living cells: principles and applications, Annu. Rev. Biophys. **38**, 301–326 (2009)

31.62 J.P. Butler, I.M. Tolic-Norrelykke, B. Fabry, J.J. Fredberg: Traction fields, moments, and strain energy that cells exert on their surroundings, Am. J. Physiol. Cell Physiol. **282**(3), C595–C605 (2002)

31.63 T. Oliver, M. Dembo, K. Jacobson: Traction forces in locomoting cells, Cell Motil. Cytoskeleton **31**(3), 225–240 (1995)

31.64 J.L. Tan, J. Tien, D.M. Pirone, D.S. Gray, K. Bhadriraju, C.S. Chen: Cells lying on a bed of microneedles: an approach to isolate mechanical force, Proc. Natl. Acad. Sci. USA **100**(4), 1484–1489 (2003)

31.65 J. Toyjanova, E. Bar-Kochba, C. Lopez-Fagundo, J. Reichner, D. Hoffman-Kim, C. Franck: High resolution, large deformation 3D traction force microscopy, PLoS One **9**(4), e90976 (2014)

31.66 A.L. Cost, P. Ringer, A. Chrostek-Grashoff, C. Grashoff: How to measure molecular forces in cells: A guide to evaluating genetically-encoded FRET-based tension sensors, Cell Mol. Bioeng. **8**(1), 96–105 (2015)

31.67 V.P. Ma, Y. Liu, L. Blanchfield, H. Su, B.D. Evavold, K. Salaita: Ratiometric tension probes for mapping receptor forces and clustering at intermembrane junctions, Nano Lett **16**(7), 4552–4559 (2016)

31.68 G. Scarcelli, W.J. Polacheck, H.T. Nia, K. Patel, A.J. Grodzinsky, R.D. Kamm, S.H. Yun: Noncontact three-dimensional mapping of intracellular hydromechanical properties by Brillouin microscopy, Nat. Methods **12**(12), 1132–1134 (2015)

31.69 M. Glogauer, P. Arora, G. Yao, I. Sokholov, J. Ferrier, C.A. McCulloch: Calcium ions and tyrosine phosphorylation interact coordinately with actin to regulate cytoprotective responses to stretching, J. Cell Sci. **110**(1), 11–21 (1997)

31.70 P.A. Janmey: Mechanical properties of cytoskeletal polymers, Curr. Opin. Cell Biol. **3**(1), 4–11 (1991)

31.71 H. Yamaoka, S. Matsushita, Y. Shimada, T. Adachi: Multiscale modeling and mechanics of filamentous actin cytoskeleton, Biomech. Model. Mechanobiol. **11**(3–4), 291–302 (2012)

31.72 B. Fabry, G.N. Maksym, J.P. Butler, M. Glogauer, D. Navajas, N.A. Taback, E.J. Millet, J.J. Fredberg: Time scale and other invariants of integrative mechanical behavior in living cells, Phys. Rev. E **68**(4), 041914 (2003)

31.73 D. Mizuno, C. Tardin, C.F. Schmidt, F.C. Mackintosh: Nonequilibrium mechanics of active cytoskeletal networks, Science **315**(5810), 370–373 (2007)

31.74 J. Lammerding, R.D. Kamm, R.T. Lee: Mechanotransduction in cardiac myocytes, Ann. NY Acad. Sci. **1015**, 53–70 (2004)

31.75 N. Wang, J.D. Tytell, D.E. Ingber: Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus, Nat. Rev. Mol. Cell Biol. **10**(1), 75–82 (2009)

31.76 M.L. Gardel, J.H. Shin, F.C. MacKintosh, L. Mahadevan, P.A. Matsudaira, D.A. Weitz: Scaling of F-actin network rheology to probe single filament elasticity and dynamics, Phys. Rev. Lett. **93**(18), 188102 (2004)

31.77 K.E. Kasza, A.C. Rowat, J. Liu, T.E. Angelini, C.P. Brangwynne, G.H. Koenderink, D.A. Weitz: The cell as a material, Curr. Opin. Cell Biol. **19**(1), 101–107 (2007)

31.78 C. Storm, J.J. Pastore, F.C. MacKintosh, T.C. Lubensky, P.A. Janmey: Nonlinear elasticity in biological gels, Nature **435**(7039), 191–194 (2005)

31.79 J.D. Humphrey, S. DeLange: *An Introduction to Biomechanics* (Springer, New York 2003)

31.80 A.J. Grodzinsky, E.H. Frank: *Fields, Forces, and Flows in Biological Systems*, Vol. xii (Garland Science, London, New York 2011) p. 308

31.81 H. Isambert, A.C. Maggs: Dynamics and rheology of actin solutions, Macromolecules **29**, 1036–1040 (1996)

31.82 L.J. Gibson, M.F. Ashby: *Cellular Solids: Structure and Properties* (Cambridge Univ. Press, Cambridge 1999)

31.83 D. Stamenovic, J.J. Fredberg, N. Wang, J.P. Butler, D.E. Ingber: A microstructural approach to cytoskeletal mechanics based on tensegrity, J. Theor. Biol. **181**(2), 125–136 (1996)

31.84 D.E. Ingber: Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton, J. Cell Sci. **104**(3), 613–627 (1993)

31.85 N. Wang, I.M. Tolic-Norrelykke, J. Chen, S.M. Mijailovich, J.P. Butler, J.J. Fredberg, D. Stamenovic: Cell prestress. I. Stiffness and prestress are closely associated in adherent contractile cells, Am. J. Physiol. Cell Physiol. **282**(3), C606–C616 (2002)

31.86 M. Mehrbod, M.R. Mofrad: On the significance of microtubule flexural behavior in cytoskeletal mechanics, PLoS One **6**(10), e25627 (2011)

31.87 E. Evans, K. Ritchie: Dynamic strength of molecular adhesion bonds, Biophys. J. **72**(4), 1541–1555 (1997)

31.88 E.V. Sokurenko, V. Vogel, W.E. Thomas: Catch-bond mechanism of force-enhanced adhesion: counterintuitive, elusive, but ... widespread?, Cell Host Microbe **4**(4), 314–323 (2008)

31.89 K.C. Chang, D.F. Tees, D.A. Hammer: The state diagram for cell adhesion under flow: leukocyte rolling and firm adhesion, Proc. Natl. Acad. Sci. USA **97**(21), 11262–11267 (2000)

31.90 D.A. Hammer, S.M. Apte: Simulation of cell rolling and adhesion on surfaces in shear flow: general results and analysis of selectin-mediated neutrophil adhesion, Biophys. J. **63**(1), 35–57 (1992)

31.91 R. Waugh, E.A. Evans: Thermoelasticity of red blood cell membrane, Biophys. J. **26**(1), 115–131 (1979)

31.92 J.N. Israelachvili: *Intermolecular and Surface Forces* (Academic Press, San Diego 1992)

31.93 N. Mohandas, E. Evans: Mechanical properties of the red cell membrane in relation to molecular structure and genetic defects, Annu. Rev. Biophys. Biomol. Struct. **23**, 787–818 (1994)

31.94 E.D. Sheets, R. Simson, K. Jacobson: New insights into membrane dynamics from the analysis of cell surface interactions by physical methods, Curr. Opin. Cell Biol. **7**(5), 707–714 (1995)

31.95 D.V. Zhelev, D. Needham, R.M. Hochmuth: Role of the membrane cortex in neutrophil deformation in small pipets, Biophys. J. **67**(2), 696–705 (1994)

31.96 N. Caille, O. Thoumine, Y. Tardy, J.J. Meister: Contribution of the nucleus to the mechanical properties of endothelial cells, J. Biomech. **35**(2), 177–187 (2002)

31.97 F. Guilak, J.R. Tedrow, R. Burgkart: Viscoelastic properties of the cell nucleus, Biochem. Biophys. Res. Commun. **269**(3), 781–786 (2000)

31.98 H. Liu, J. Wen, Y. Xiao, J. Liu, S. Hopyan, M. Radisic, C.A. Simmons, Y. Sun: In situ mechanical characterization of the cell nucleus by atomic force microscopy, ACS Nano **8**(4), 3821–3828 (2014)

31.99 J. Lammerding: Mechanics of the nucleus, Compr. Physiol. **1**(2), 783–807 (2011)

31.100 K.N. Dahl, S.M. Kahn, K.L. Wilson, D.E. Discher: The nuclear envelope lamina network has elasticity and a compressibility limit suggestive of a molecular shock absorber, J. Cell Sci. **117**(20), 4779–4786 (2004)

31.101 J. Schape, S. Prausse, M. Radmacher, R. Stick: Influence of lamin A on the mechanical properties of amphibian oocyte nuclei measured by atomic force microscopy, Biophys. J. **96**(10), 4319–4325 (2009)

31.102 S. Deguchi, K. Maeda, T. Ohashi, M. Sato: Flow-induced hardening of endothelial nucleus as an intracellular stress-bearing organelle, J. Biomech. **38**(9), 1751–1759 (2005)

31.103 A. Rowat, L. Foster, M. Nielsen, M. Weiss, J. Ipsen: Characterization of the elastic properties of the nuclear envelope, J. R. Soc. Interface **2**, 63–69 (2005)

31.104 M. Le Berre, J. Aubertin, M. Piel: Fine control of nuclear confinement identifies a threshold deformation leading to lamina rupture and induction of specific genes, Integr. Biol. (Camb.) **4**(11), 1406–1414 (2012)

31.105 C.M. Denais, R.M. Gilbert, P. Isermann, A.L. McGregor, M. te Lindert, B. Weigelin, P.M. Davidson, P. Friedl, K. Wolf, J. Lammerding: Nuclear envelope rupture and repair during cancer cell migration, Science **352**(6283), 353–358 (2016)

31.106 M. Krause, J. Te Riet, K. Wolf: Probing the compressibility of tumor cell nuclei by combined atomic force-confocal microscopy, Phys. Biol. **10**(6), 065002 (2013)

31.107 Y. Tseng, J.S. Lee, T.P. Kole, I. Jiang, D. Wirtz: Micro-organization and visco-elasticity of the interphase nucleus revealed by particle nanotracking, J. Cell Sci. **117**(10), 2159–2167 (2004)

31.108 A.H. de Vries, B.E. Krenn, R. van Driel, V. Subramaniam, J.S. Kanger: Direct observation of nanomechanical properties of chromatin in living cells, Nano Lett. **7**(5), 1424–1427 (2007)

31.109 J. Lammerding, L.G. Fong, J.Y. Ji, K. Reue, C.L. Stewart, S.G. Young, R.T. Lee: Lamins A and C but not lamin B1 regulate nuclear mechanics, J. Biol. Chem. **281**(35), 25768–25780 (2006)

31.110 J. Swift, I.L. Ivanovska, A. Buxboim, T. Harada, P.C. Dingal, J. Pinter, J.D. Pajerowski, K.R. Spinler, J.W. Shin, M. Tewari, F. Rehfeldt, D.W. Speicher, D.E. Discher: Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation, Science **341**(6149), 1240104 (2013)

31.111 K.N. Dahl, P. Scaffidi, M.F. Islam, A.G. Yodh, K.L. Wilson, T. Misteli: Distinct structural and mechanical properties of the nuclear lamina in Hutchinson-Gilford progeria syndrome, Proc. Natl. Acad. Sci. USA **103**(27), 10271–10276 (2006)

31.112 D.P. Theret, M.J. Levesque, M. Sato, R.M. Nerem, L.T. Wheeler: The application of a homogeneous half-space model in the analysis of endothelial cell micropipette measurements, J. Biomech. Eng. **110**, 190–199 (1998)

31.113 B. Geiger, A. Bershadsky: Exploring the neighborhood: adhesion-coupled cell mechanosensors, Cell **110**(2), 139–142 (2002)

31.114 M.P. Sheetz: Cell control by membrane-cytoskeleton adhesion, Nat. Rev. Mol. Cell Biol. **2**(5), 392–396 (2001)

31.115 B. Patel, A.R. Gingras, A.A. Bobkov, L.M. Fujimoto, M. Zhang, R.C. Liddington, D. Mazzeo, J. Emsley, G.C. Roberts, I.L. Barsukov, D.R. Critchley: The activity of the vinculin binding sites in talin is influenced by the stability of the helical bundles that make up the talin rod, J. Biol. Chem. **281**(11), 7458–7467 (2006)

31.116 S.E. Lee, R.D. Kamm, M.R. Mofrad: Force-induced activation of talin and its possible role in focal adhesion mechanotransduction, J. Biomech. **40**(9), 2096–2106 (2007)

31.117 V.P. Hytonen, V. Vogel: How force might activate talin's vinculin binding sites: SMD reveals a structural mechanism, PLoS Comput. Biol. **4**(2), e24 (2008)

31.118 J. Golji, M.R. Mofrad: The talin dimer structure orientation is mechanically regulated, Biophys. J. **107**(8), 1802–1809 (2014)

31.119 J. Golji, R. Collins, M.R. Mofrad: Molecular mechanics of the alpha-actinin rod domain: bending, torsional, and extensional behavior, PLoS Comput. Biol. **5**(5), e1000389 (2009)

31.120 T. Truong, H. Shams, M.R. Mofrad: Mechanisms of integrin and filamin binding and their interplay with talin during early focal adhesion formation, Integr. Biol. (Camb.) **7**(10), 1285–1296 (2015)

31.121 H. Shams, J. Golji, M.R. Mofrad: A molecular trajectory of alpha-actinin activation, Biophys. J.

103(10), 2050–2059 (2012)
31.122 J. Golji, M.R. Mofrad: A molecular dynamics investigation of vinculin activation, Biophys. J. **99**(4), 1073–1081 (2010)
31.123 X. Hu, C. Jing, X. Xu, N. Nakazawa, V.W. Cornish, F.M. Margadant, M.P. Sheetz: Cooperative vinculin binding to talin mapped by time-resolved super resolution microscopy, Nano Lett **16**(7), 4062–4068 (2016)
31.124 H. Shams, J. Golji, K. Garakani, M.R. Mofrad: Dynamic regulation of alpha-actinin's calponin homology domains on f-actin, Biophys. J. **110**(6), 1444–1455 (2016)
31.125 G. Meacci, H. Wolfenson, S. Liu, M.R. Stachowiak, T. Iskratsch, A. Mathur, S. Ghassemi, N. Gauthier, E. Tabdanov, J. Lohner, A. Gondarenko, A.C. Chander, P. Roca-Cusachs, B. O'Shaughnessy, J. Hone, M.P. Sheetz: α-Actinin links extracellular matrix rigidity-sensing contractile units with periodic cell-edge retractions, Mol. Biol. Cell **27**(22), 3471–3479 (2016)
31.126 D.J. Odde, L. Ma, A.H. Briggs, A. DeMarco, M.W. Kirschner: Microtubule bending and breaking in living fibroblast cells, J. Cell Sci. **112**(19), 3283–3288 (1999)
31.127 G. Chang, R.H. Spencer, A.T. Lee, M.T. Barclay, D.C. Rees: Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel, Science **282**(5397), 2220–2226 (1998)
31.128 O.P. Hamill, B. Martinac: Molecular basis of mechanotransduction in living cells, Physiol. Rev. **81**(2), 685–740 (2001)
31.129 J. Gullingsrud, D. Kosztin, K. Schulten: Structural determinants of MscL gating studied by molecular dynamics simulations, Biophys. J. **80**(5), 2074–2081 (2001)
31.130 S.H. Woo, V. Lukacs, J.C. de Nooij, D. Zaytseva, C.R. Criddle, A. Francisco, T.M. Jessell, K.A. Wilkinson, A. Patapoutian: Piezo2 is the principal mechanotransduction channel for proprioception, Nat. Neurosci. **18**(12), 1756–1762 (2015)
31.131 I. Banerjee, J. Zhang, T. Moore-Morris, E. Pfeiffer, K.S. Buchholz, A. Liu, K. Ouyang, M.J. Stroud, L. Gerace, S.M. Evans, A. McCulloch, J. Chen: Targeted ablation of nesprin 1 and nesprin 2 from murine myocardium results in cardiomyopathy, altered nuclear morphology and inhibition of the biomechanical gene response, PLoS Genet **10**(2), e1004114 (2014)
31.132 G.R. Fedorchak, A. Kaminski, J. Lammerding: Cellular mechanosensing: getting to the nucleus of it all, Prog. Biophys. Mol. Biol. **115**(2–3), 76–92 (2014)
31.133 C. Guilluy, K. Burridge: Nuclear mechanotransduction: forcing the nucleus to respond, Nucleus **6**(1), 19–22 (2015)
31.134 B. Enyedi, M. Jelcic, P. Niethammer: The cell nucleus serves as a mechanotransducer of tissue damage-induced inflammation, Cell **165**(5), 1160–1170 (2016)
31.135 C. Guilluy, L.D. Osborne, L. Van Landeghem, L. Sharek, R. Superfine, R. Garcia-Mata, K. Burridge: Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus, Nat. Cell Biol. **16**(4), 376–381 (2014)
31.136 J.M. Ferrer, H. Lee, J. Chen, B. Pelz, F. Nakamura, R.D. Kamm, M.J. Lang: Measuring molecular rupture forces between single actin filaments and actin-binding proteins, Proc. Natl. Acad. Sci. USA **105**(27), 9221–9226 (2008)
31.137 H. Shams, M. Soheilypour, M. Peyro, M. Mofrad: Looking "under the hood" of cellular mechanotransduction with computational tools: A systems biomechanics approach across multiple scales, ACS Biomater. Sci. Eng. (2017) doi:10.1021/acsbiomaterials.7b00117

32. Nanomechanical Properties of Nanostructures and Scale Effects

Bharat Bhushan

Structural integrity is of paramount importance in all devices. Load applied during the use of devices can result in component failure. Cracks can develop and propagate under tensile stresses, leading to failure. Knowledge of the mechanical properties of nanostructures is necessary for the design of realistic micro/nanoelectromechanical systems (MEMS/NEMS) and biological micro/nanoelectromechanical systems (bioMEMS/bioNEMS) devices. Elastic and inelastic properties are needed to predict deformation from an applied load in the elastic and inelastic regimes respectively. The strength property is needed to predict the allowable operating limit. Some properties of interest are hardness, elastic modulus, bending strength, fracture toughness, and fatigue strength. Many of the mechanical properties are scale dependent. Therefore, these should be measured at relevant scales. Atomic force microscopy and nanoindenters can be used to evaluate the mechanical properties of micro/nanoscale structures.

Commonly used materials in MEMS/NEMS are single-crystal silicon and silicon-based materials, e.g., SiO_2 and polysilicon films deposited by low-pressure chemical vapor deposition. Single-crystal SiC deposited on large-area silicon substrates is used for high-temperature micro/nanosensors and actuators. Amorphous alloys can be formed on both metal and silicon substrates by sputtering and plating techniques, providing more flexibility in surface integration. Electroless deposited Ni-P amorphous thin films have been used to construct microdevices, especially using the so-called Lithographie, Galvanoformung, Abformung (LIGA) techniques. Micro/nanodevices need conductors to provide power, as well as electrical/magnetic signals to make them functional. Electroplated gold films have found wide application in electronic devices because of their ability to make thin films and because they process simply. Polymers, such as poly(methyl methacrylate) (PMMA), poly(dimethylsiloxane) (PDMS), and polystyrene are commonly used in BioMEMS/BioNEMS such as micro/nanofluidic devices because of the ease of manufacturing and reduced cost. Many polymers are biocompatible, so they may be integrated into biomedical devices.

This chapter presents a review of mechanical property measurements on the micro/nanoscale of various materials of interest and stress and deformation analyses of nanostructures.

32.1 **Experimental Techniques for Measurement of Mechanical Properties of Nanostructures** 1103
32.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters 1103
32.1.2 Bending Tests of Nanostructures Using an AFM 1104
32.1.3 Bending Tests of Micro/Nanostructures Using a Nanoindenter 1108
32.2 **Experimental Results and Discussion** . 1109
32.2.1 Indentation and Scratch Tests of Various Ceramic and Metals Using Micro/Nanoindenter 1109
32.2.2 Bending Tests of Ceramic Nanobeams Using an AFM .. 1113
32.2.3 Bending Tests of Metallic Microbeams Using a Nanoindenter 1116
32.2.4 Indentation and Scratch Tests of Polymeric Microbeams Using a Nanoindenter 1117
32.2.5 Bending Tests of Polymeric Microbeams Using a Nanoindenter 1121
32.3 **Finite Element Analysis of Nanostructures with Roughness and Scratches** 1125
32.3.1 Stress Distribution in a Smooth Nanobeam 1126
32.3.2 Effect of Roughness in the Longitudinal Direction 1126

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_32

32.3.3 Effect of Roughness in the Transverse Direction and Scratches 1130
32.3.4 Effect on Stresses and Displacements for Materials Which are Elastic, Elastic–Plastic or Elastic–Perfectly Plastic 1131
32.4 **Summary** 1131
32.A **Appendix: Fabrication Procedure for the Double-Anchored and Cantilever Beams** 1132
References 1134

Microelectromechanical systems (MEMS) refer to microscopic devices that have a characteristic length of less than 1 mm but more than 100 nm (or 1 μm), and nanoelectromechanical systems (NEMS) refer to nanoscopic devices that have a characteristic length of less than 100 nm (or 1 μm). These are referred to as an intelligent miniaturized system comprising of sensing, processing, and/or actuating functions and combine electrical and mechanical components. The acronym MEMS originated in the USA. The term commonly used in Europe is micro/nanosystem technology (MST/NST), and in Japan it is micro/nanomachines. Another term generally used is micro/nanodevices. MEMS/NEMS terms are also now used in a broad sense and include electrical, mechanical, optical, biological, and/or fluidic functions. To put the dimensions in perspective, individual atoms are typically a fraction of a nanometer in diameter, DNA molecules are about 2.5 nm wide, biological cells are in the range of thousands of nm in diameter, and human hair is about 75 μm in diameter. The mass of a micromachined silicon structure can be as low as 1 nN, and NEMS can be built with mass as low as 10^{-20} N with cross-sections of about 10 nm. In comparison, the mass of an eyelash is about 100 nN, and the mass of a drop of water is about 10 μN.

A wide variety of MEMS, including Si-based devices, chemical and biological sensors and actuators, and miniature nonsilicon structures (e.g., devices made from plastics or ceramics) have been fabricated with dimensions in the range of a couple to a few thousand microns (see e.g., [32.1–11]). A variety of NEMS have also been produced (see e.g., [32.12–18]). MEMS/NEMS technology and fabrication processes have found a variety of applications in biology and biomedicine, leading to the establishment of an entirely new field known as BioMEMS/BioNEMS [32.19–26]. The ability to use micro/nanofabrication processes to develop precision devices that can interface with biological environments at the cellular and molecular level has led to advances in the fields of biosensor technology [32.26–30], drug delivery [32.31–34], and tissue engineering [32.35–37]. The miniaturization of fluidic systems using micro/nanofabrication techniques has led to new and more efficient devices for medical diagnostics and biochemical analysis [32.38]. The largest industrial applications of MEMS include accelerometers, pressure sensors, thermal and piezoelectric-type inkjet printer heads, and digital micromirror devices. BioMEMS and BioNEMS are increasingly used in commercial applications. The largest applications of BioMEMS and BioNEMS include a variety of biosensors and drug delivery devices.

Structural integrity is of paramount importance in all devices. Load applied during the use of devices can result in component failure. Cracks can develop and propagate under tensile stresses leading to failure. Friction/stiction and wear limit the lifetimes and compromise the performance and reliability of devices involving relative motion [32.3, 39–41]. Most MEMS/NEMS applications demand extreme reliability. Stress and deformation analyses are carried out for an optimal design. MEMS/NEMS designers require information about mechanical properties at the nanoscale. Elastic and inelastic properties are needed to predict deformation from an applied load in the elastic and inelastic regimes respectively. Understanding of the strength property is needed to predict the allowable operating limit. Some of the properties of interest are hardness, elastic modulus, creep, bending strength (fracture stress), fracture toughness, and fatigue strength. Micro/nanostructures have some surface topography and local scratches dependent upon the manufacturing process. Surface roughness and local scratches may compromise the reliability of the devices, and their effect needs to be studied.

Most mechanical properties are scale dependent [32.40, 41]. Several researchers have measured the mechanical properties of silicon and silicon-based milli- to microscale structures. These include tensile tests and bending tests [32.42–51], resonant structure tests for the measurement of elastic properties [32.52], fracture toughness tests [32.43, 45, 53–57], and fatigue tests [32.55, 58, 59]. Nanomechanical properties of nanoscale structures have been measured using atomic force microscopy (AFM) [32.60, 61] and nanoindentation [32.62–65]. For stress and deformation analyses of simple geometries and boundary conditions, analytical models can be used. For analysis of complex

geometries, numerical models are needed. The conventional finite element method (FEM) can be used down to a few tens of nanometers scale, although its applicability is questionable at the nanoscale. FEM has been used for simulation and prediction of residual stresses and strains induced in MEMS devices during fabrication [32.66], to perform fault analysis in order to study MEMS faulty behavior [32.67], to compute mechanical strain resulting from doping of silicon [32.68], analyze micromechanical experimental data [32.45, 69, 70], and nanomechanical experimental data [32.61]. FEM analysis of nanostructures has been performed to analyze the effect of types of surface roughness and scratches on stresses in nanostructures [32.71, 72].

Commonly used materials for MEMS/NEMS are single-crystal silicon and silicon-based materials (e.g., SiO_2 and polysilicon films deposited by the low-pressure chemical vapor deposition (LPCVD) process) [32.10]. An early study showed silicon to be a mechanically resilient material, in addition to its favorable electronic properties [32.73]. Single-crystal 3C-SiC (cubic or β-SiC) films, deposited by the atmospheric-pressure chemical vapor deposition (APCVD) process on large-area silicon substrates, are produced for high-temperature micro/nanosensor and actuator applications [32.74–76]. Amorphous alloys can be formed on both metal and silicon substrates by sputtering and plating techniques, providing more flexibility in surface integration. Electroless deposited Ni-P amorphous thin films have been used to construct microdevices, especially using the so-called LIGA techniques [32.10, 63]. Micro/nanodevices need conductors to provide power, as well as electrical/magnetic signals to make them functional. Electroplated gold films have found wide applications in electronic devices because of their ability to make thin films and process simplicity [32.63].

As the field of MEMS/NEMS has progressed, alternative materials, especially polymers, have established an important role in their construction. This trend has been driven by the reduced cost associated with polymer materials. Polymer microfabrication processes, including micromolding and hot embossing techniques [32.77], can be orders of magnitude less expensive than traditional silicon photolithography processes. Many polymers are biocompatible, so these can be integrated into biomedical devices with minimal detrimental effects to the host or on the biofluids. An improvement in device functionality (relative to silicon) is also possible due to the mechanical properties of the polymer used, which have properties much closer to biological tissues. Polymer BioMEMS structures involving microbeams have been designed to measure cellular forces [32.64, 65]. Polymer materials most commonly used for biomedical applications include poly(methyl methacrylate) (PMMA), poly(dimethylsiloxane) (PDMS), and polystyrene [32.65, 77–79]. Another material of interest due to ease of fabrication is poly(propyl methacrylate) (PPMA), which has a lower glass transition temperature (T_g) (35−43 °C) [32.80] than PMMA (104−106 °C) [32.81, 82], permitting low-temperature thermal processing [32.64, 65, 79].

This chapter presents a review of mechanical property measurements on the nanoscale of various materials of interest and stress and deformation analyses of nanostructures.

32.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures

32.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters

A nanoindenter is commonly used to measure hardness, elastic modulus, and fracture toughness, and to perform micro/nanoscratch studies to get a measure of scratch/wear resistance of materials [32.83].

Hardness and Elastic Modulus

The nanoindenter monitors and records the dynamic load and displacement of the three-sided pyramidal diamond (Berkovich) indenter during indentation with a force resolution of about 75 nN and displacement resolution of about 0.1 nm. Hardness and elastic modulus are calculated from the load-displacement data [32.83]. The peak indentation load depends on the mechanical properties of the specimen; a harder material requires a higher load for reasonable indentation depth.

Fracture Toughness

The indentation technique for fracture toughness measurement of brittle samples on the microscale is based on the measurement of the lengths of median-radial cracks produced by indentation. A Vickers indenter (a four-sided diamond pyramid) is used in a microhardness tester. A load on the order of 0.5 N is typically used in making the Vickers indentations. The

indentation impressions are examined using an optical microscope with Nomarski interference contrast to measure the length of median-radial cracks, c. The fracture toughness (K_{IC}) is calculated by the following relation [32.84],

$$K_{IC} = \alpha \left(\frac{E}{H}\right)^{1/2} \left(\frac{P}{c^{3/2}}\right) , \tag{32.1}$$

where α is an empirical constant depending on the geometry of the indenter, H and E are hardness and elastic moduli, and P is the peak indentation load. For Vickers indenters, α has been empirically found based on experimental data and is equal to 0.016 [32.83]. Both E and H values are obtained from the nanoindentation data. The crack length is measured from the center of the indent to the end of crack using an optical microscope. For one indent, all crack lengths are measured. The crack length c is obtained from the average value of several indents.

Indentation Creep

For an indentation creep test of polymer samples, the test is performed using a continuous stiffness measurements (CSMs) technique [32.83]. In a study by *Wei* et al. [32.64], the indentation load was typically 30 μN, and the loading rate was 3 μN/s. The tip was held typically for 600 s after the indentation load reached 30 μN. To measure the mean stress and contact stiffness, during the hold segment the indenter was oscillated at a peak-to-peak load amplitude of 1.2 μN and a frequency of 45 Hz.

Scratch Resistance

In micro/nanoscratch studies, in a nanoindenter, a conical diamond indenter having a tip radius of about 1 μm and an included angle of 60° is drawn over the sample surface, and the load is ramped up until substantial damage occurs [32.83]. The coefficient of friction is monitored during scratching. In order to obtain scratch depths during scratching, the surface profile of the sample surface is first obtained by translating the sample at a low load of about 0.2 mN, which is insufficient to damage a hard sample surface. The 500 μm long scratches are made by translating the sample while ramping the loads on the conical tip over different loads dependent upon the material hardness. The actual depth during scratching is obtained by subtracting the initial profile from the scratch depth measured during scratching. In order to measure the scratch depth after the scratch, the scratched surface is profiled at a low load of 0.2 mN and is subtracted from the actual surface profile before scratching.

32.1.2 Bending Tests of Nanostructures Using an AFM

Quasistatic bending tests of fixed nanobeam arrays in the normal direction are carried out using an AFM [32.61, 85]. A three-sided pyramidal diamond tip (with a radius of about 200 nm) mounted on a rectangular stainless steel cantilever is used for the bending tests. The beam stiffness is selected based on the desired load range. The stiffness of the cantilever beams for application of a normal load up to 100 μN is about 150–200 N/m.

For the bending test, the tip is brought over the nanobeam array with the help of the sample stage of the AFM and a built-in high magnification optical microscope (Fig. 32.1) [32.61]. For the fine positioning of the tip over a chosen beam, the array is scanned in contact mode at a contact load of about 2–4 μN, which results in negligible damage to the sample. After scanning, the tip is located at one end of a chosen beam. To position the tip at the center of the beam span, the tip is moved to the other end of the beam by giving the x-piezoelectric electrode an offset voltage. The value of this offset is determined after several such attempts have been made in order to minimize the effects of piezoelectric drift. Half of this offset is then applied to the x-piezoelectric electrode after the tip is positioned at one end of the beam, which usually results in the tip being moved to the center of the span. Once the tip is positioned over the center of the beam span, the tip is held stationary without scanning, and the z-piezoelectric electrode is extended by a known distance, typically about 2.5 μm, at a rate of 10 nm/s, as shown in Fig. 32.1. During this time, the vertical deflection signal (dV_{AFM}), which is proportional to the deflection of the cantilever (D_{tip}), is monitored. The displacement of the piezoelectric tube scanner is equal to the sum of the displacements of the cantilever and the nanobeam. Hence the displacement of the nanobeam (D_{beam}) under the point of load can be determined as

$$D_{beam} = D_{piezo} - D_{tip} . \tag{32.2}$$

The load (F_{beam}) on the nanobeam is the same as the load on the tip/cantilever (F_{tip}) and is given by

$$F_{beam} = F_{tip} = D_{tip} k , \tag{32.3}$$

where k is the stiffness of the tip/cantilever. In this manner, a load displacement curve for each nanobeam can be obtained.

The photodetector sensitivity of the cantilever needs to be calibrated to obtain D_{tip} in nm. For this calibration, the tip is pushed against a smooth diamond sample by moving the z-piezoelectric electrode over

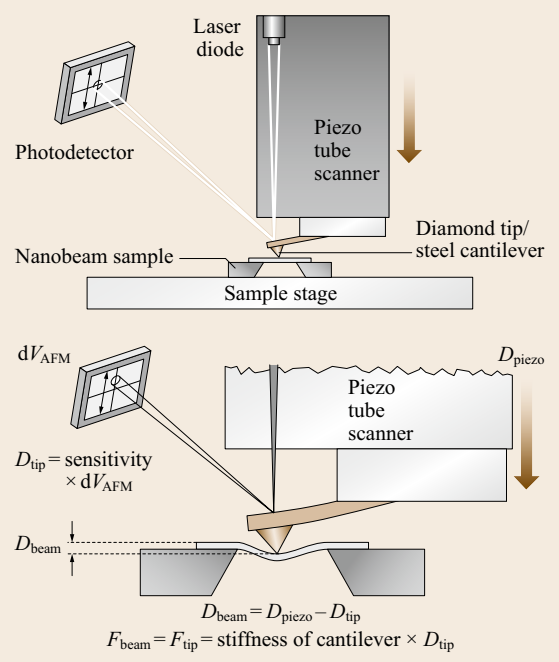


Fig. 32.1 Schematic showing the details of a nanoscale bending test using an AFM. The AFM tip is brought to the center of the nanobeam and the piezoelectric tube scanner is extended over a known distance. By measuring the tip displacement, a load displacement curve of the nanobeam can be obtained (after [32.61])

a known distance. For the hard diamond material, the actual deflection of the tip can be assumed to be the same as the z-piezoelectric electrode travel (D_{piezo}), and the photodetector sensitivity (S) for the cantilever setup is determined as

$$S = \frac{D_{piezo}}{dV_{AFM}} \frac{nm}{V} . \tag{32.4}$$

In the measurements, D_{tip} is given as $dV_{AFM}S$.

Since a sharp tip would result in an undesirable large local indentation, *Sundararajan* and *Bhushan* [32.61] used a diamond tip that was worn (blunt). Indentation experiments using this tip on a silicon substrate yielded a residual depth of less than 8 nm at a maximum load of 120 μN, which is negligible compared to displacements of the beams (several hundred nm). Hence we can assume that negligible local indentation or damage is created during the bending process of the beams and that the displacement calculated from (32.2) is from the beam structure.

Elastic Modulus and Bending Strength

Elastic modulus and bending strength (fracture stress) of the beams can be estimated by equations based on

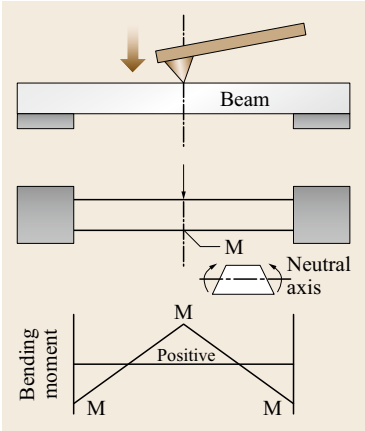


Fig. 32.2 Schematic of the bending moments generated in the beam during a quasistatic bending experiment, with the load at the center of the span. The maximum moments occur under the load and at the fixed ends. Due to the trapezoidal cross-section, the maximum tensile bending stresses occur at the top surfaces at the fixed ends

the assumption that the beams follow the linear elastic theory of an isotropic material. This is probably valid since the beams have high length-to-width ℓ/w and length-to-thickness ℓ/t ratios and also since the length direction is along the principal stress direction during the test. For a fixed elastic beam loaded at the center of the span, the elastic modulus is expressed as

$$E = \frac{\ell^3}{192I} m , \tag{32.5}$$

where ℓ is the beam length, I is the area moment of inertia for the beam cross-section, and m is the slope of the linear region of the load–displacement curve during bending [32.86]. The area moment of inertia for a beam with a trapezoidal cross-section is calculated from the following equation

$$I = \frac{4w_1w_2 + w_1^2 + w_2^2}{36\,(w_1 + w_2)} t^3 , \tag{32.6}$$

where w_1 and w_2 are the upper and lower widths respectively, and t is the thickness of the beam. According to linear elastic theory, for a centrally loaded beam, the moment diagram is shown in Fig. 32.2. The maximum moments are generated at the ends (negative moment) and under the loading point (positive moment) as shown in Fig. 32.2. The bending stresses generated in the beam are proportional to the moments and are compressive or tensile about the neutral axis (line of zero stress). The maximum tensile stress (σ_b, which is the bending strength or fracture stress) is produced on the top sur-

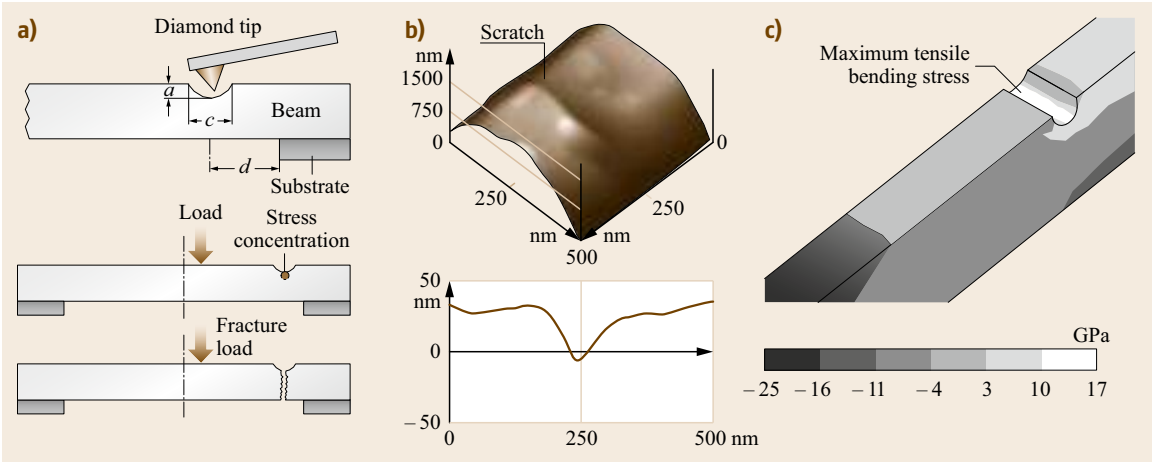


Fig. 32.3 **(a)** Schematic of technique to generate a defect (crack) of known dimensions in order to estimate fracture toughness. A diamond tip is used to generate a scratch across the width of the beam. When the beam is loaded as shown, a stress concentration is formed at the bottom of the scratch. The fracture load is then used to evaluate the stresses using FEM. **(b)** AFM three-dimensional (3-D) image and two-dimensional (2-D) profile of a typical scratch. **(c)** Finite element model results verifying that the maximum bending stress occurs at the bottom of the scratch (after [32.61])

face at both the ends and is given by [32.86]

$$\sigma_b = \frac{F_{max}\ell e_1}{8I} , \tag{32.7}$$

where F_{max} is the applied load at failure, and e_1 is the distance of the top surface from the neutral plane of the beam cross-section and is given by [32.86]

$$e_1 = \frac{t\,(w_1 + 2w_2)}{3\,(w_1 + w_2)} . \tag{32.8}$$

Although the moment value at the center of the beam is the same as at the ends, the tensile stresses at the center (generated on the bottom surface) are less than those generated at the ends (per (32.7)) because the distance from the neutral axis to the bottom surface is less than e_1. This is because of the trapezoidal cross-section of the beam, which results in the neutral axis being closer to the bottom surface than the top Fig. 32.2.

In the preceding analysis, the beams were assumed to have fixed ends. However, in the nanobeams used by *Sundararajan* and *Bhushan* [32.61], the underside of the beams was pinned over some distance on either side of the span. Hence a finite element model of the beams was created to see if the difference in the boundary conditions affected the stresses and displacements of the beams. It was found that the difference in the stresses was less than 1%. This indicates that the boundary conditions near the ends of the actual beams are not that different from that of fixed ends. Therefore the bending strength values can be calculated from (32.7).

Fracture Toughness

Fracture toughness is another important parameter for brittle materials such as silicon. In the case of the nanobeam arrays, these are not best suited for fracture toughness measurements because they do not possess regions of uniform stress during bending. *Sundararajan* and *Bhushan* [32.61] developed a methodology, which is outlined schematically in Fig. 32.3a. First, a crack of known geometry is introduced in the region of maximum tensile bending stress, i. e., on the top surface near the ends of the beam. This is achieved by generating a scratch at high normal load across the width (w_1) of the beam using a sharp diamond tip (radius < 100 nm). A typical scratch thus generated is shown in Fig. 32.3b. By bending the beam as shown, a stress concentration will be formed under the scratch. This will lead to failure of the beam under the scratch once a critical load (fracture load) is attained. The fracture load and relevant dimensions of the scratch are input into the FEM model, which is used to generate the fracture stress plots. Figure 32.3c shows an FEM simulation of one such experiment, which reveals that the maximum stress does occur under the scratch.

If we assume that the scratch tip acts as a crack tip, a bending stress will tend to open the crack in Mode I. In this case, the stress field around the crack tip can be described by the stress intensity parameter K_I (for Mode I) for linear elastic materials [32.87]. In particular

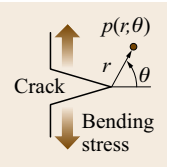


Fig. 32.4 Schematic of crack tip and coordinate systems used in (32.9) to describe a stress field around the crack tip in terms of the stress intensity parameter, K_I (after [32.61])

the stresses corresponding to the bending stresses are described by

$$\sigma = \frac{K_I}{\sqrt{2\pi r}} \cos\left(\frac{\theta}{2}\right)\left[1 + \sin\left(\frac{\theta}{2}\right)\sin\left(\frac{3\theta}{2}\right)\right] \tag{32.9}$$

for every point $p(r, \theta)$ around the crack tip as shown in Fig. 32.4. If we substitute the fracture stress (σ_f) into the left-hand side of (32.9), then the K_I value can be substituted by its critical value, which is the fracture toughness K_{IC}. Now, the fracture stress can be determined for the point ($r = 0, \theta = 0$), i. e., right under the crack tip as explained above. However, we cannot substitute $r = 0$ in (32.9). The alternative is to substitute a value for r that is as close to zero as possible. For silicon, a reasonable number is the distance between neighboring atoms in the (111) plane, the plane along which silicon exhibits the lowest fracture energy. This value was calculated from silicon unit cell dimensions of 0.5431 nm [32.88] to be 0.4 nm (half of the face diagonal). This assumes that Si displays no plastic zone around the crack tip, which is reasonable since in tension silicon is not known to display much plastic deformation at room temperature. *Sundararajan* and *Bhushan* [32.61] used values $r = 0.4–1.6$ nm (i. e., distances up to four times the distance between the nearest neighboring atoms) to estimate the fracture toughness for both Si and SiO_2 according to the following equation

$$K_{IC} = \sigma_f\sqrt{2\pi r}\,, \quad r = 0.4–1.6\,\text{nm}\,. \tag{32.10}$$

Fatigue Strength

In addition to the properties mentioned so far that can be evaluated from quasistatic bending tests, the fatigue properties of nanostructures are also of interest. This is especially true for MEMS/NEMS involving vibrating structures such as oscillators and comb drives [32.89] and hinges in digital micromirror devices [32.90]. To study the fatigue properties of the nanobeams, *Sundararajan* and *Bhushan* [32.61] applied monotonic cyclic stresses using an AFM (Fig. 32.5a). Similar to the bending test, the diamond tip is first positioned at the center of the beam span. In order to ensure that the tip is always in contact with the beam (as opposed to impacting it), the piezoelectric tube scanner is first extended by a distance D_1, which ensures a minimum stress on the beam. After this extension, a cyclic

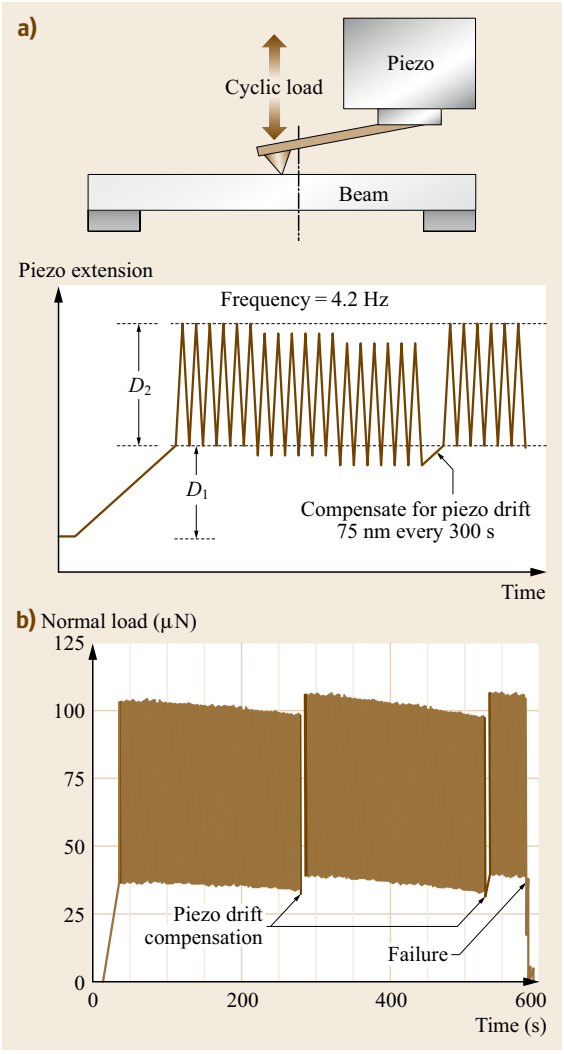


Fig. 32.5 **(a)** Schematic showing the details of the technique to study fatigue behavior of the nanobeams. The diamond tip is located at the middle of the span and a cyclic load at 4.2 Hz is applied to the beam by forcing the piezoelectric tube scanner to move in the pattern shown. An extension is made every 300 s to compensate for the piezoelectric drift, ensuring that the load on the beam is kept fairly constant. **(b)** Data from a fatigue experiment on a nanobeam until failure. The normal load is computed from the raw vertical deflection signal. The compensations for piezoelectric drift keep the load fairly constant (after [32.61])

displacement of amplitude D_2 is applied continuously until failure of the beam occurs. This results in the application of a cyclic load to the beam. The maximum frequency of the cyclic load that could be attained using the AFM by *Sundararajan* and *Bhushan* [32.61] was

4.2 Hz. The vertical deflection signal of the tip is monitored throughout the experiment. The signal follows the pattern of the piezoelectric input up to failure, which is indicated by a sudden drop in the signal. During initial runs, piezoelectric drift was observed that caused the piezoelectric tube scanner to gradually move away from the beam (i. e., to retract), resulting in a continuous decrease in the applied normal load. In order to compensate for this, the piezoelectric tube scanner is given a finite extension of 75 nm every 300 s as shown in Fig. 32.5a. This results in keeping the applied loads fairly constant. The normal load variation (calculated from the vertical deflection signal) from a fatigue test is shown in Fig. 32.5b. The cyclic stress amplitudes (corresponding to D_2) and fatigue lives are recorded for every sample tested. Values for D_1 are set such that minimum stress levels are about 20% of the bending strengths.

32.1.3 Bending Tests of Micro/Nanostructures Using a Nanoindenter

Quasistatic bending tests of micro/nanostructures in both normal and lateral directions are carried out using a nanoindenter (Fig. 32.6) [32.62–65]. The advantage of a nanoindenter is that loads up to about 400 mN, higher than that in AFM (up to about 100 μN), can be used for structures requiring high loads for experiments. Bending experiments in the normal direction have been carried out on the suspended beams using the nanoindenter fitted with a conical tip that has a 1 μm radius of curvature and a 60° included angle. Figure 32.6a shows the schematic of the beam bending experimental setup. To avoid the indenter tip pushing into the specimen, a blunt tip is used in the bending and fatigue tests. For ceramic and metallic beam samples, *Li* et al. [32.63] used a diamond conical indenter with a radius of 1 μm and an included angle of 60°. For polymer beam samples, *Wei* et al. [32.64] and *Palacio* et al. [32.65] reported that the diamond tip penetrated the polymer beams easily and caused considerable plastic deformation during the bending test, which led to significant errors in the measurements. To avoid this issue, the diamond tip was dip-coated with PMMA (about 1–2 μm thick) by dipping the tip in the 2 wt% PMMA solution for about 5 s. Load position used was at the center of the span for the bridge beams and at 10 μm off from the free end of cantilever beams. An optical microscope with a magnification of 1500× or an in situ AFM is used to locate the loading position. Then the specimen is moved under the indenter location with a resolution of about 200 nm in the longitudinal direction and less than 100 nm in the lateral direction.

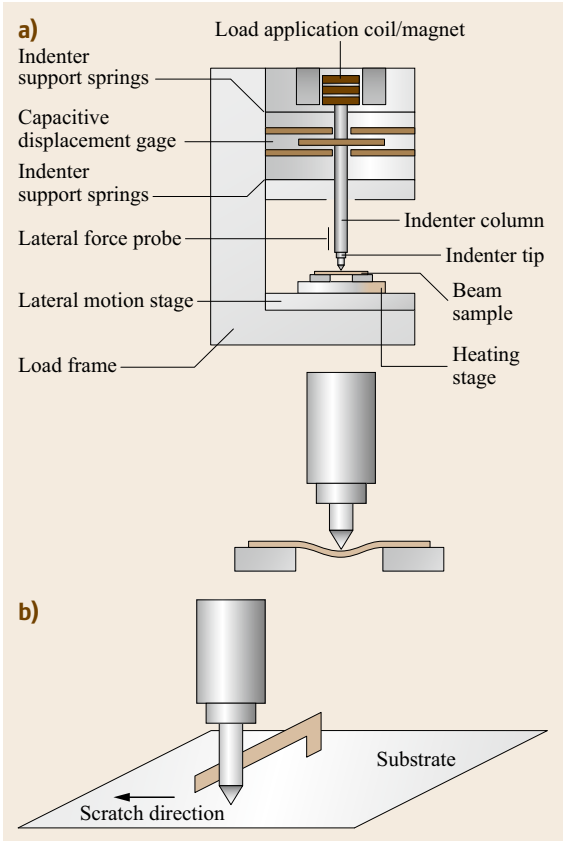


Fig. 32.6a,b Schematic of micro/nanoscale bending experimental setup using a nanoindenter **(a)** in the normal direction, and **(b)** in the lateral direction

Using the analysis presented earlier, elastic modulus and bending strength of the beams can be obtained from the load displacement curves [32.63–65]. For fatigue tests, an oscillating load is applied and contact stiffness is measured during the tests. A significant drop in the contact stiffness during the test is a measure of the number of cycles to failure [32.62].

A schematic of the lateral bending experimental setup is shown in Fig. 32.6b. The nanoindenter was fitted with a conical tip, which has a 1 μm radius of curvature and a 60° included angle. The tip was positioned at 200–300 μm away from the beam anchor and 20 μm from the edge of the beam length. The indenter was programmed to perform a scratch at a constant load of 400 μN, and the stage was moved at a rate of 10 μm/s. Since the lateral force contribution of the substrate surface is constant, any additional force recorded comes from the cantilever beam response. From the linear elastic regime in the measured lateral force as a function of lateral displacement, the elastic modulus can be evaluated. For a cantilever beam with one end clamped, it is

expressed as [32.86],

$$E = \frac{\ell^3}{3I} m \,, \tag{32.11}$$

where, as described earlier, ℓ is the beam length, I the area moment of inertia for the beam cross-section, and m is the slope of the linear region of the force displacement curve.

32.2 Experimental Results and Discussion

32.2.1 Indentation and Scratch Tests of Various Ceramic and Metals Using Micro/Nanoindenter

Studies have been conducted on six different materials: undoped single-crystal Si(100), undoped polysilicon film, SiO_2 film, SiC film, electroless deposited Ni-11.5 wt% P amorphous film, and electroplated Au film [32.63, 75, 76]. A 3 μm-thick polysilicon film was deposited by a low-pressure chemical vapor deposition (LPCVD) process on a Si(100) substrate. The 1 μm-thick SiO_2 film was deposited by a plasma-enhanced chemical vapor deposition (PECVD) process on a Si(111) substrate. A 3 μm-thick 3C-SiC film was epitaxially grown using an atmospheric-pressure chemical vapor deposition (APCVD) process on a Si(100) substrate. A 12 μm-thick Ni-P film was electroless plated on a 0.8 mm-thick Al-4.5 wt% Mg alloy substrate. A 3 μm-thick Au film was electroplated on a Si(100) substrate.

Hardness and Elastic Modulus

Hardness and elastic modulus measurements are made using a nanoindenter [32.63]. The hardness and elastic modulus values of various materials at a peak indentation depth of 50 nm are summarized in Fig. 32.7 and Table 32.1. The SiC film exhibits the highest hardness among the samples examined, of about 25 GPa, with an elastic modulus of about 395 GPa, followed by the undoped Si(100), undoped polysilicon film, SiO_2 film, Ni-P film, and Au film. The hardness and elastic modulus data of the undoped Si(100) and undoped polysilicon film are comparable. For the metal alloy films, the Ni-P film exhibits higher hardness and elastic modulus than the Au film.

Fracture Toughness

The optical images of Vickers indentations made using a microindenter at a normal load of 0.5 N held for 15 s on the undoped Si(100), undoped polysilicon film, and SiC film are shown in Fig. 32.8 [32.76]. The SiC film exhibits the smallest indentation mark, followed by the undoped polysilicon film and undoped Si(100). These Vickers indentation depths are smaller than one-third of the film thickness. Thus, the influence of substrate on the fracture toughness of the films can be ignored. In addition to the indentation marks, radial

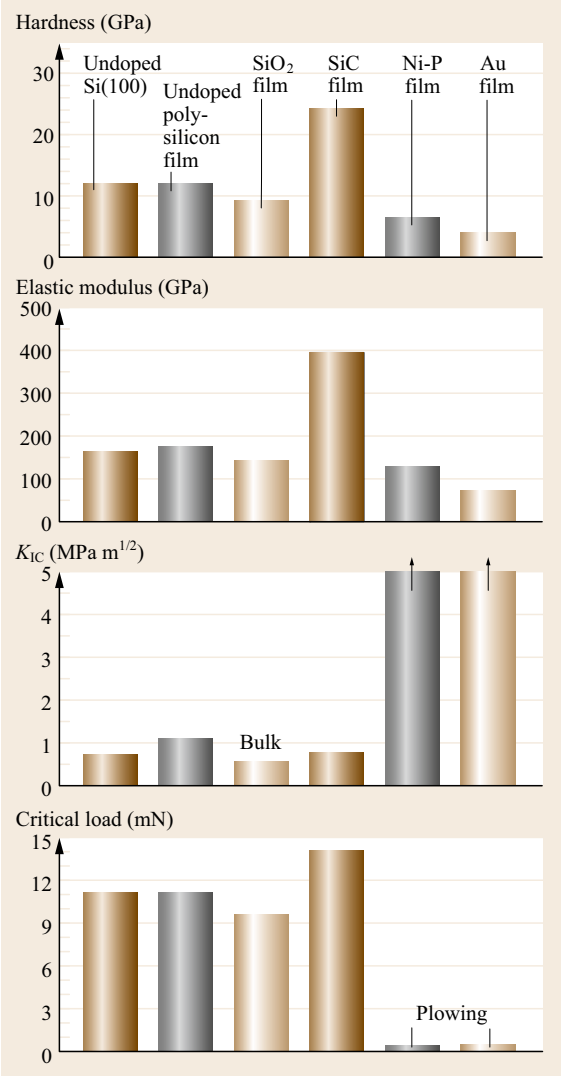


Fig. 32.7 Bar chart summarizing the hardness, elastic modulus, fracture toughness, and critical load (from scratch tests) results of the bulk undoped single-crystal Si(100) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au (after [32.63])

Table 32.1 Hardness, elastic modulus, fracture toughness, and critical load results of bulk single-crystal Si(100) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au

Samples	Hardness (GPa)	Elastic modulus (GPa)	Fracture toughness ($MPa\,m^{1/2}$)	Critical load (mN)
Undoped Si(100)	12	165	0.75	11
Undoped polysilicon film	12	167	1.11	11
SiO_2 film	9.5	144	0.58 (bulk)	9.5
SiC film	24.5	395	0.78	14
Ni-P film	6.5	130		0.4 (plowing)
Au film	4	72		0.4 (plowing)

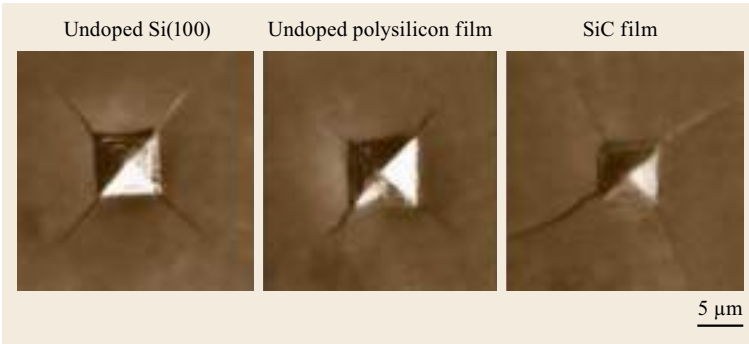


Fig. 32.8 Optical images of Vickers indentations made at a normal load of 0.5 N held for 15 s on the undoped Si(100), undoped polysilicon film, and SiC film (after [32.76])

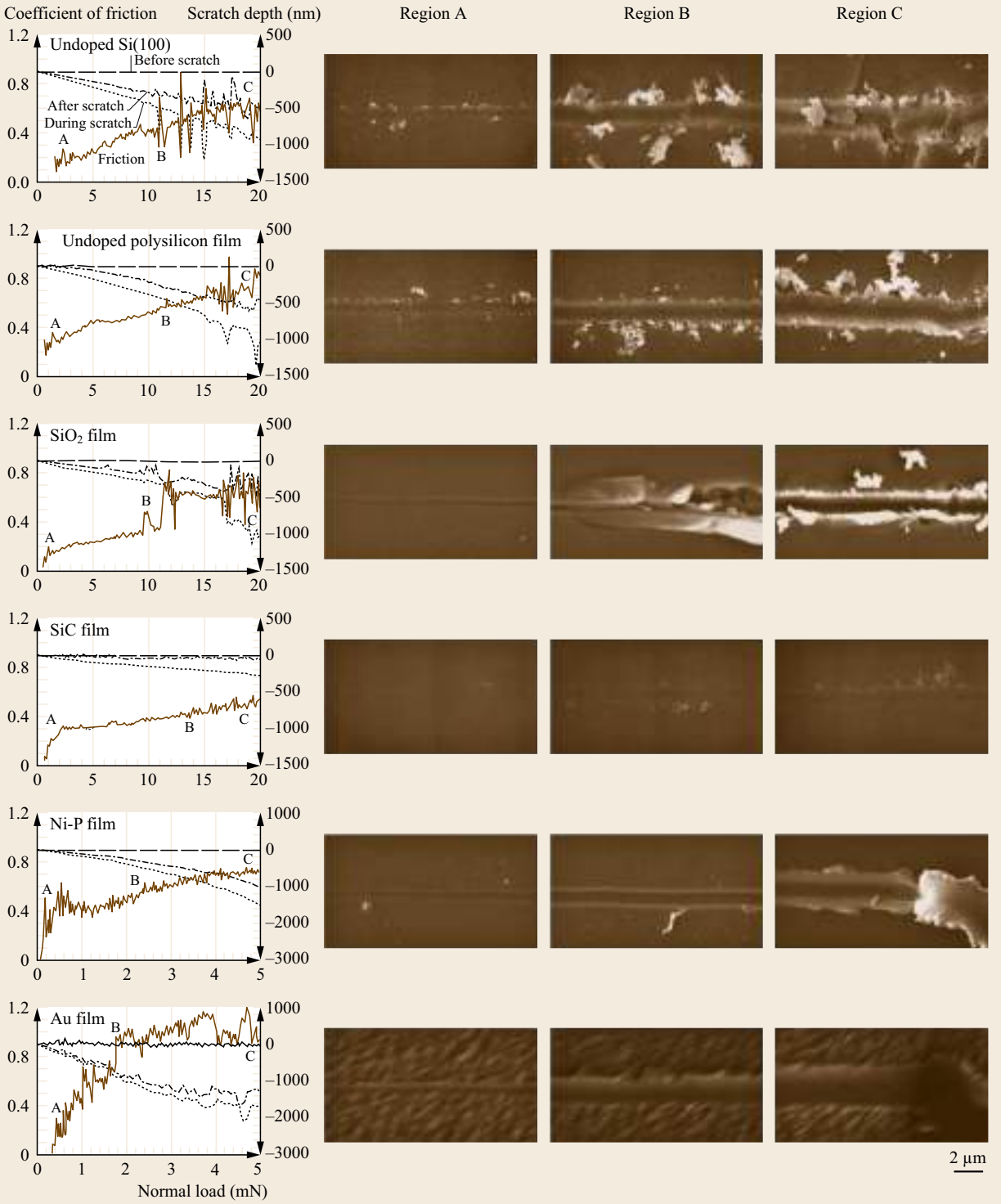
cracks are observed, emanating from the indentation corners. The SiC film shows the longest radial crack length, followed by the undoped Si(100) and undoped polysilicon film. The radial cracks for the undoped Si(100) are straight whereas those for the SiC, undoped polysilicon film are not straight but go in a zigzag manner. The fracture toughness (K_{IC}) is calculated using (32.1).

The fracture toughness values of all samples are summarized in Fig. 32.7 and Table 32.1. The SiO_2 film used in this study is about 1 μm thick, which is not thick enough for fracture toughness measurement. The fracture toughness value of bulk silica is listed instead for a reference. The Ni-P and Au films exhibit very high fracture toughness values that cannot be measured by indentation methods. For other samples, the undoped polysilicon film has the highest value, followed by the undoped Si(100), SiC film, and SiO_2 film. For the undoped polysilicon film, the grain boundaries can stop the radial cracks and change the propagation directions of the radial cracks, making the propagation of these cracks more difficult. Values of fracture toughness for the undoped Si(100) and SiC film are comparable. Since the undoped Si(100) and SiC film are single crystal, no grain boundaries are present to stop the radial cracks and change the propagation directions of the radial cracks. This is why the SiC film shows a lower fracture toughness value than the bulk polycrystal SiC materials of 3.6 $MPa\,m^{1/2}$ [32.91].

Scratch Resistance

The scratch resistance of various materials has been studied using a nanoindenter by *Li* et al. [32.63]. Figure 32.9 compares the coefficient of friction and scratch depth profiles as a function of increasing normal load and optical images of three regions over scratches: at the beginning of the scratch (indicated by A on the friction profile), at the point of initiation of damage at which the coefficient of friction increases to a high value or increases abruptly (indicated by B on the friction profile), and towards the end of the scratch (indicated by C on the friction profile) for all samples. Note that the ramp loads for the Ni-P and Au range are from 0.2 to 5 mN whereas the ramp loads for other samples are from 0.2 to 20 mN. All samples exhibit a continuous increase in the coefficient of friction with increasing normal load from the beginning of the scratch. The continuous increase in the coefficient of friction during scratching is attributed to the increasing

Fig. 32.9 Coefficient of friction and scratch depth profiles as a function of increasing normal load and optical images of three regions over scratches: at the beginning of the scratch (indicated by A on the friction profile), at the point of initiation of damage at which the coefficient of friction increases to a high value or increase abruptly (indicated by B on the friction profile), and towards the end of the scratch (indicated by C on the friction profile) for all samples (after [32.63]) ►


Coefficient of friction
Scratch depth (nm)
Region A
Region B
Region C
Undoped Si(100)
Before scratch
After scratch
During scratch
Friction
A
B
C
Undoped polysilicon film
SiO2 film
SiC film
Ni-P film
Au film
Normal load (mN)
2 µm

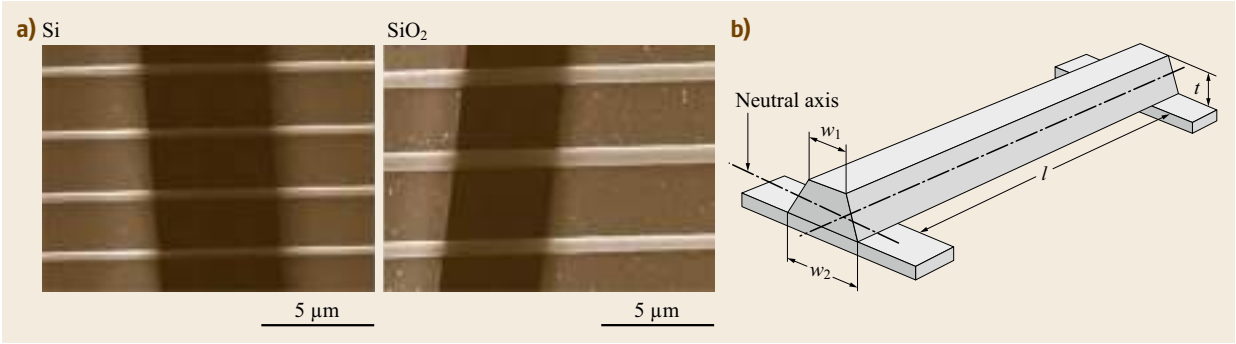


Fig. 32.10 (a) SEM micrographs of nanobeam arrays, and (b) a schematic of the shape of a typical nanobeam. The trapezoidal cross-section is due to the anisotropic wet etching during fabrication (after [32.85])

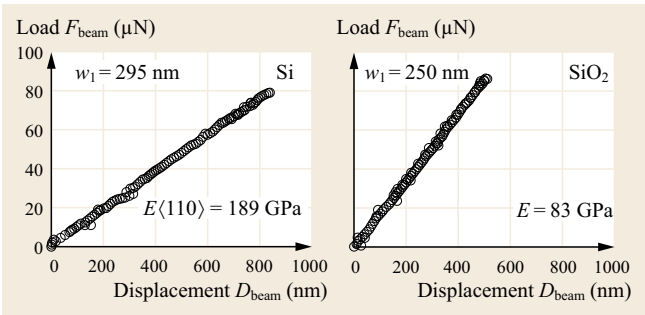


Fig. 32.11 Typical load displacement curves of silicon and SiO_2 nanobeams. The *curves* are linear until sudden failure, indicative of brittle fracture of the beams. The elastic modulus (E) values calculated from the *curves* are shown. The dimensions of the Si beam were $w_1 = 295\,\text{nm}$, $w_2 = 484\,\text{nm}$ and $t = 255\,\text{nm}$, while those of the SiO_2 beam were $w_1 = 250\,\text{nm}$, $w_2 = 560\,\text{nm}$ and $t = 425\,\text{nm}$ (after [32.85])

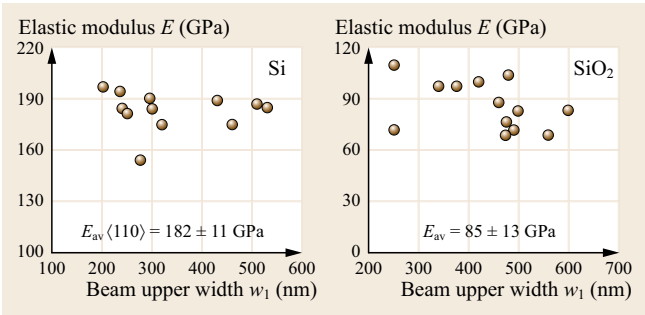


Fig. 32.12 Elastic modulus values measured for Si and SiO_2. The average values are shown. These are comparable to bulk values, which demonstrates that elastic modulus shows no specimen-size dependence (after [32.61])

plowing of the sample by the tip with increasing normal load, as shown in the scanning electron microscopy (SEM) images in Fig. 32.9. The abrupt increase in the coefficient of friction is associated with catastrophic failure as well as significant plowing of the tip into the sample. Before the critical load, the coefficient of friction of the undoped polysilicon, SiC, and SiO_2 films increased at a slower rate, and was smoother than that of the other samples. The undoped Si(100) exhibits some bursts in the friction profiles before the critical load. At the critical load, the SiC and undoped polysilicon films exhibit a small increase in the coefficient of friction whereas the undoped Si(100) and undoped polysilicon film exhibit a sudden increase in the coefficient of friction. The Ni-P and Au films show a continuous increase in the coefficient of friction, indicating the behavior of a ductile metal. The bursts in the friction profile might result from the plastic deformation and material pileup in front of the scratch tip. The Au film exhibits a higher coefficient of friction than the Ni-P film. This is because the Au film has lower hardness and elastic modulus values than the Ni-P film.

The SEM images show that below the critical loads the undoped Si(100) and undoped polysilicon film were damaged by plowing, associated with the plastic flow of the material and formation of debris on the sides of the scratch. For the SiC and SiO_2 films, in region A, a plowing scratch track was found without any debris on the side of the scratch, which is probably responsible for the smoother curve and slower increase in the coefficient of friction before the critical load. After the critical load, for the SiO_2 film, delamination of the film from the substrate occurred, followed by cracking along the scratch track. For the SiC film, only several small debris particles were found without any cracks on the side of the scratch, which is responsible for the small increase in the coefficient of friction at the critical load. For the undoped Si(100), cracks were found on the side of the scratch right from the critical load and up, which is probably responsible for the big bursts in the friction profile. For the undoped polysilicon film, no cracks were found on the side of the scratch at the critical load. This might result from grain boundaries, which can stop

Table 32.2 Summary of measured parameters from quasistatic bending tests

Sample	Elastic modulus E (GPa)		Bending strength σ_b (GPa)		Fracture toughness K_{IC} (MPa m$^{1/2}$)		
	Measured	Bulk value	Measured	Reported (microscale)	Estimated	Reported (microscale)	Bulk value
Si	182±11	169[a]	18 ± 3	< 10[c]	1.67 ± 0.4	0.6–1.65[e]	0.9[f]
SiO_2	85±13	73[b]	7.6 ± 2	< 2[d]	0.60 ± 0.2	0.5–0.9[d]	–

[a] Si[110] [32.92], [b] [32.93], [c] [32.42, 43, 45–48, 70, 94, 95], [d] [32.57], [e] [32.53–56], [f] [32.88]

the propagation of cracks. At the end of the scratch, some of the surface material was torn away and cracks were found on the side of the scratch in the undoped Si(100). A couple of small cracks were found in the undoped polysilicon and SiO_2 films. No crack was found in the SiC film. Even at the end of the scratch, less debris was found in the SiC film. A curly chip was found at the end of the scratch in both Ni-P and Au films. This is a typical characteristic of ductile metal alloys. The Ni-P and Au films were damaged by plowing right from the beginning of the scratch with material pileup at the side of the scratch.

The scratch depth profiles obtained during and after the scratch on all samples with respect to initial profile, after the cylindrical curvature is removed, are plotted in Fig. 32.9. Reduction in scratch depth is observed after scratching as compared to that of during scratching. This reduction in scratch depth is attributed to an elastic recovery after removal of the normal load. The scratch depth after scratching indicates the final depth, which reflects the extent of permanent damage and plowing of the tip into the sample surface, and is probably more relevant for visualizing the damage that can occur in real applications. For the undoped Si(100), undoped polysilicon film, and SiO_2 film, there is a large scatter in the scratch depth data after the critical loads, which is associated with the generation of cracks, material removal, and debris. The scratch depth profile is smooth for the SiC film. It is noted that the SiC film exhibits the lowest scratch depth among the samples examined. The scratch depths of the undoped Si(100), undoped polysilicon film and SiO_2 film are comparable. The Ni-P and Au films exhibit much larger scratch depth than other samples. The scratch depth of the Ni-P film is smaller than that of the Au film.

The critical loads estimated from friction profiles for all samples are compared in Fig. 32.7 and Table 32.1. The SiC film exhibits the highest critical load of about 14 mN, as compared to other samples. The undoped Si(100) and undoped polysilicon film show comparable critical load of about 11 mN whereas the SiO_2 film shows a low critical load of about 9.5 mN. The Ni-P and Au films were damaged by plowing right from the beginning of the scratch.

32.2.2 Bending Tests of Ceramic Nanobeams Using an AFM

Bending tests have been performed on Si and SiO_2 nanobeam arrays [32.61, 85]. The single-crystal silicon bridge nanobeams were fabricated by bulk micromachining incorporating enhanced-field anodization using an AFM [32.60]. The Si nanobeams are oriented along the [110] direction in the (001) plane. Subsequent thermal oxidation of the beams results in the formation of SiO_2 beams. The cross-section of the nanobeams is trapezoidal owing to the anisotropic wet etching process. SEM micrographs of Si and SiO_2 nanobeam arrays and a schematic of the shape of a typical nanobeam are shown in Fig. 32.10. The actual widths and thicknesses of nanobeams were measured using an AFM in tapping mode prior to tests using a standard Si tapping-mode tip (tip radius < 10 nm). Surface roughness measurements of the nanobeam surfaces in tapping mode yielded a σ of 0.7±0.2 nm and peak-to-valley (P–V) distance of 4±1.2 nm for Si and a σ of 0.8±0.3 nm and a P-V distance of 3.1±0.8 nm for SiO_2. Prior to testing, the Si nanobeams were cleaned by immersing them in a *piranha etch* solution (3:1 solution by volume of 98% sulphuric acid and 30% hydrogen peroxide) for ten minutes to remove any organic contaminants.

Bending Strength

Figure 32.11 shows typical load displacement curves for Si and SiO_2 beams that were bent to failure [32.61, 85]. The upper width (w_1) of the beams is indicated in the figure. Also indicated in the figure are the elastic modulus values obtained from the slope of the load displacement curve (32.5). All the beams tested showed linear elastic behavior followed by abrupt failure, which is suggestive of brittle fracture. Figure 32.12 shows the scatter in the values of elastic modulus obtained for both Si and SiO_2 along with the average values (± standard deviation). The scatter in the values may be due to differences in orientation of the beams with respect to the trench and the loading point being a little off center with respect to the beam span. The average values are a little higher than the bulk values (169 GPa for Si[110] and 73 GPa for SiO_2 in Table 32.2). However the values of E

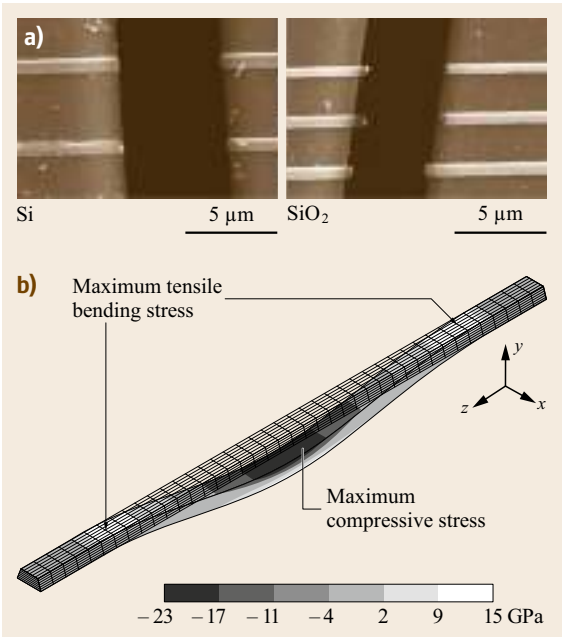


Fig. 32.13 **(a)** SEM micrographs of nanobeams that failed during quasistatic bending experiments. The beams failed at or near the ends, which is the location of maximum tensile bending stress (after [32.85]), and **(b)** bending stress distribution for silicon nanobeam indicating that the maximum tensile stresses occur on the top surfaces near the fixed ends

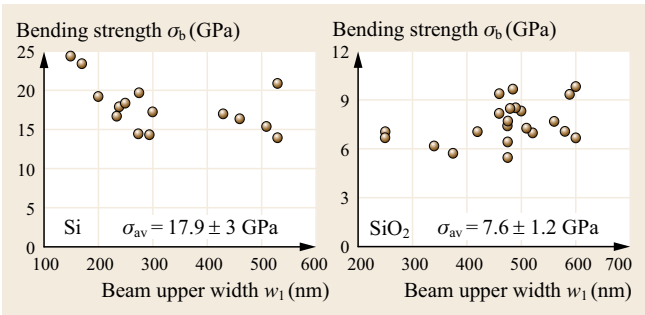


Fig. 32.14 Bending strength values obtained from bending experiments. Average values are indicated. These values are much higher than values reported for microscale specimens, indicating that bending strength shows a specimen-size effect (after [32.61])

obtained from (32.5) have an error of about 20% due to the uncertainties in beam dimensions and spring constant of the tip/cantilever (which affects the measured load). Hence the elastic modulus values on the nanoscale can be considered to be comparable to bulk values.

Most of the beams when loaded quasistatically at the center of the span broke at the ends as shown in Fig. 32.13a, which is consistent with the fact that maximum tensile stresses occur on the top surfaces near the ends. (See FEM stress distribution results in Fig. 32.13b.) Figure 32.14 shows the values of bending strength obtained for different beams. There appears to be no trend in bending strength with the upper width (w_1) of the beams. The large scatter is expected for the strength of brittle materials, since they are dependent on preexisting flaw populations in the material and hence are statistical in nature. The Weibull distribution, a statistical analysis, can be used to describe the scatter in the bending strength values. The means of the Weibull distributions were found to be 17.9 and 7.6 GPa for Si and SiO_2 respectively. Previously reported numbers of strengths range from 1–6 GPa for silicon [32.42, 43, 45–48, 50, 70, 94, 95] and about 1 GPa for SiO_2 [32.57] microscale specimens. This clearly indicates that bending strength shows a specimen-size dependence. Strength of brittle materials is dependent on preexisting flaws in the material. Since for nanoscale specimens the volume is smaller than for micro- and macroscale specimens, the flaw population will be smaller as well, resulting in higher values of strength.

Fracture Toughness

Estimates of fracture toughness calculated using (32.10) for Si and SiO_2 are shown in Fig. 32.15 [32.61]. The results show that the K_{IC} estimate for Si is about 1–2 MPa $m^{1/2}$, whereas for SiO_2 the estimate is about 0.5–0.9 MPa $m^{1/2}$. These values are comparable to values reported by others on larger specimens for Si [32.53–56] and SiO_2 [32.57]. The high values obtained for Si could be due to the fact that the scratches, despite being quite sharp, still have a finite radius of about 100 nm. The bulk value for silicon is about 0.9 MPa $m^{1/2}$ (Table 32.2). Fracture toughness is considered to be a material property and is believed to be independent of specimen size. The values obtained in this study, given its limitations, appear to show that fracture toughness is comparable, if not a little higher on the nanoscale.

Fatigue Strength

Fatigue strength measurements of Si nanobeams have been carried out by *Sundararajan* and *Bhushan* [32.61] using an AFM and *Li* and *Bhushan* [32.62] using a nanoindenter. Various stress levels were applied to nanobeams by *Sundararajan* and *Bhushan* [32.61]. The minimum stress was 3.5 GPa for Si beams and 2.2 GPa for SiO_2 beams. The frequency of applied load was 4.2 Hz. In general, the fatigue life decreased with increasing mean stress as well as increasing stress amplitude. When the stress amplitude was less than 15% of the bending strength, the fatigue life was greater than 30 000 cycles for both Si and SiO_2. However, the mean

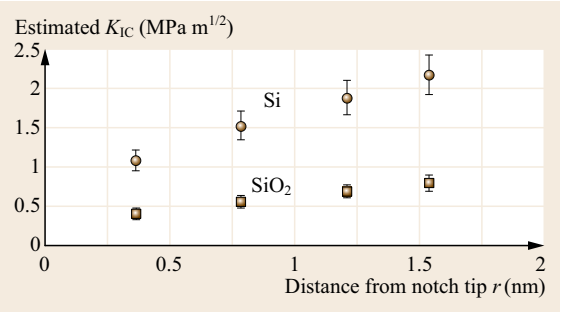


Fig. 32.15 Fracture toughness (K_{IC}) values for increasing values of r corresponding to distance between neighboring atoms in {111} planes of silicon (0.4 nm). Hence r values between 0.4 and 1.6 nm are chosen. The K_{IC} values thus estimated are comparable to values reported by others for both Si and SiO_2 (after [32.61])

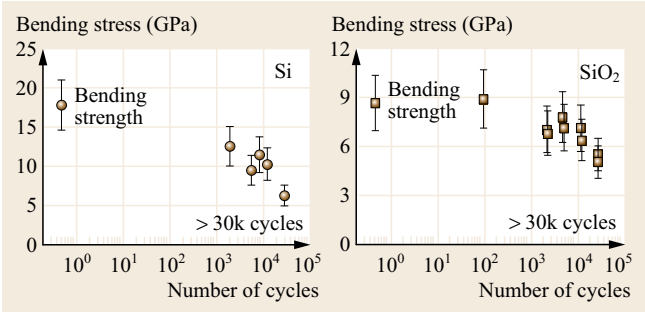


Fig. 32.16 Fatigue test data showing applied bending stress as a function of number of cycles. A single load-unload sequence is considered as one cycle. The bending strength data points are therefore associated with 1/2 cycle, since failure occurs upon loading (after [32.61])

stress had to be less than 30% of the bending strength for a life of greater than 30 000 cycles for Si whereas even at a mean stress of 43% of the bending strength, SiO_2 beams showed a life greater than 30 000 cycles. During fatigue, the beams broke under the loading point or at the ends when loaded at the center of the span. This was different from the quasistatic bending tests, where the beams broke at the ends almost every time. This could be due to the fact that the stress levels under the load and at the ends are not that different and fatigue crack propagation could occur at either location. Figure 32.16 shows a nanoscale *S–N* curve, with bending stress (*S*) as a function of fatigue in number of cycles (*N*) with an apparent endurance life at lower stress. This study clearly demonstrates that fatigue properties of nanoscale specimens can be studied.

SEM Observations of Fracture Surfaces

Figure 32.17 shows SEM images of the fracture surfaces of nanobeams broken during quasistatic bending as well as fatigue [32.61]. In the quasistatic cases, the maximum tensile stresses occur on the top surface, so it is reasonable to assume that fracture initiated at or near the top surface and propagated downward. The fracture surfaces of the beams suggest a cleavage type of fracture. Silicon beam surfaces show various ledges or facets, which is typical for crystalline brittle materials. Silicon usually fractures along the (111) plane due to this plane having the lowest surface energy to overcome by a propagating crack. However, failure has also been known to occur along the (110) planes in microscale specimens, despite the higher energy required as compared to the (111) planes [32.70]. The plane normal to the beam direction in these samples is the (110) plane while (111) planes will be oriented at 35° from the (110) plane. The presence of facets and irregularities on the silicon surface in Fig. 32.17a suggest that it is a combination of these two types of fractures that has occurred. Since the stress levels are very high for these specimens, it is reasonable to assume that crack-propagating forces will be high enough to result in (110)-type failures.

In contrast, the silicon fracture surfaces under fatigue, shown in Fig. 32.17b, appear very smooth without facets or irregularities. This is suggestive of low energy fracture, i. e., of (111)-type fracture. We do not see evidence of fatigue crack propagation in the form of steps or striations on the fracture surface. We believe that for the stress levels applied in these fatigue experiments, failure in silicon occurred via cleavage associated with *static fatigue* type of failures.

SiO_2 shows very smooth fracture surfaces for both quasistatic bending and fatigue. This is in contrast to the irregular surface one might expect for the brittle failure of an amorphous material on the macroscale. However, in larger-scale fracture surfaces for such materials, the region near the crack initiation usually appears smooth or mirror-like. Since the fracture surface here is so small and very near the crack initiation site, it is not unreasonable to see such a smooth surface for SiO_2 on this scale. There appears to be no difference between the fracture surfaces obtained by quasistatic bending and fatigue for SiO_2.

Summary of Mechanical Properties Measured Using Quasistatic Bending Tests

Table 32.2 summarizes the various properties measured via quasistatic bending in this study [32.61]. Also shown are bulk values of the parameters along with values reported on larger-scale specimens by other researchers. Elastic modulus and fracture toughness values appear to be comparable to bulk values and show no dependence on specimen size. However, bending

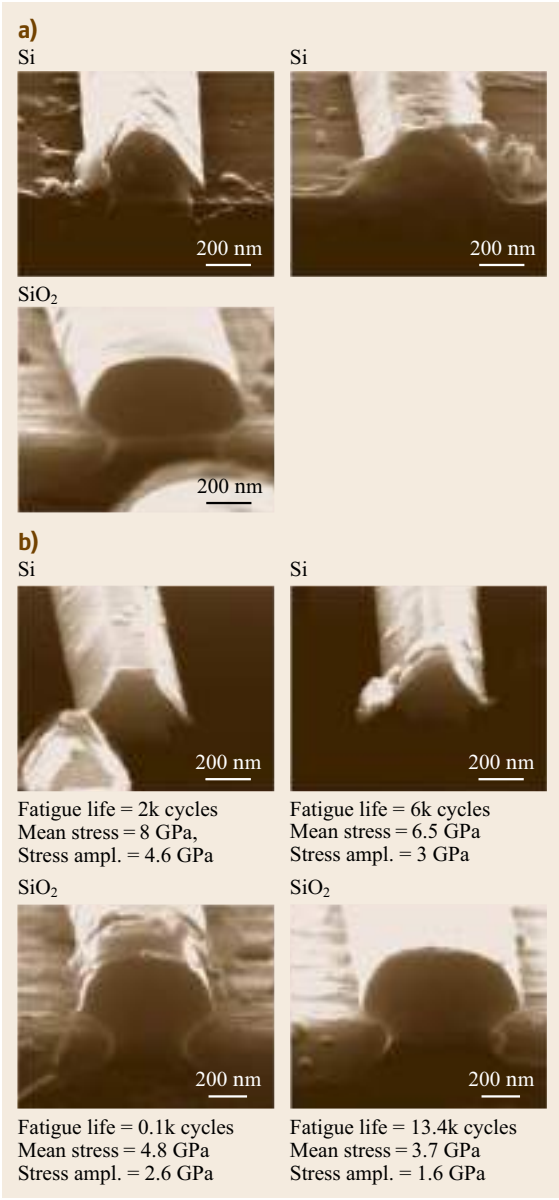


Fig. 32.17a,b SEM micrographs of fracture surfaces of silicon and SiO_2 beams subjected to **(a)** quasistatic bending and **(b)** fatigue (after [32.61])

strength shows a clear specimen-size dependence with nanoscale numbers being twice as large as numbers reported for larger-scale specimens.

32.2.3 Bending Tests of Metallic Microbeams Using a Nanoindenter

Bending tests have been performed on Ni-P and Au microbeams [32.63]. The Ni-P cantilever microbeams were fabricated by a focused ion beam machining technique. The dimensions were $10 \times 12 \times 50\,\mu m^3$. Notches with a depth of $3\,\mu m$ and a tip radius of $0.25\,\mu m$ were introduced in the microbeams to facilitate failure at a lower load in the bending tests. The Au bridge microbeams were fabricated by an electroplating technique.

Figure 32.18 shows the SEM images, load displacement curve, and FEM stress contour for the notched Ni-P cantilever microbeam that was bent to failure [32.63]. The distance between the loading position and the fixed end is $40\,\mu m$. The $3\,\mu m$ deep notch is $10\,\mu m$ from the fixed end. The notched beam showed linear behavior followed by abrupt failure. The FEM stress contour shows that there is a higher stress concentration at the notch tip. The maximum tensile stress σ_m at the notch tip can be analyzed by using the Griffith fracture theory as follows [32.84],

$$\sigma_m \approx 2\sigma_o \left(\frac{c}{\rho}\right)^{1/2} , \qquad (32.12)$$

where σ_o is the average applied tensile stress on the beam, c is the crack length, and ρ is the crack tip radius. Therefore, elastic-plastic deformation will first occur locally at the end of the notch tip, followed by abrupt fracture failure after σ_m reaches the ultimate tensile strength of Ni-P, even though the rest of the beam is still in the elastic regime. The SEM image of the fracture surface shows that the fracture started right from the notch tip with plastic deformation characteristics. This indicates that although local plastic deformation occurred at the notch tip area, the whole beam failed catastrophically. The present study shows that FEM simulation can predict well the stress concentration, and helps in understanding the failure mechanism of the notched beams.

Figure 32.19 shows the SEM images, load displacement curve, and FEM stress contour for the Au bridge microbeam that was deformed by the indenter [32.63]. The recession gap between the beam and substrate is about $7\,\mu m$, which is not large enough to break the beam at the load applied. From the load-displacement curve, we note that the beam experienced elastic-plastic deformation. The FEM stress contour shows that the maximum tensile stress is located at the fixed ends whereas the minimum compressive stress is located around the center of the beam. The SEM image shows that the beam has been permanently deformed. No crack was found on the beam surface. The present study shows a possibility for mechanically forming the Au film into the shape as needed. This may help in designing/fabricating functionally complex smart micro/nanodevices that need conductors for power supply and input/output signals.

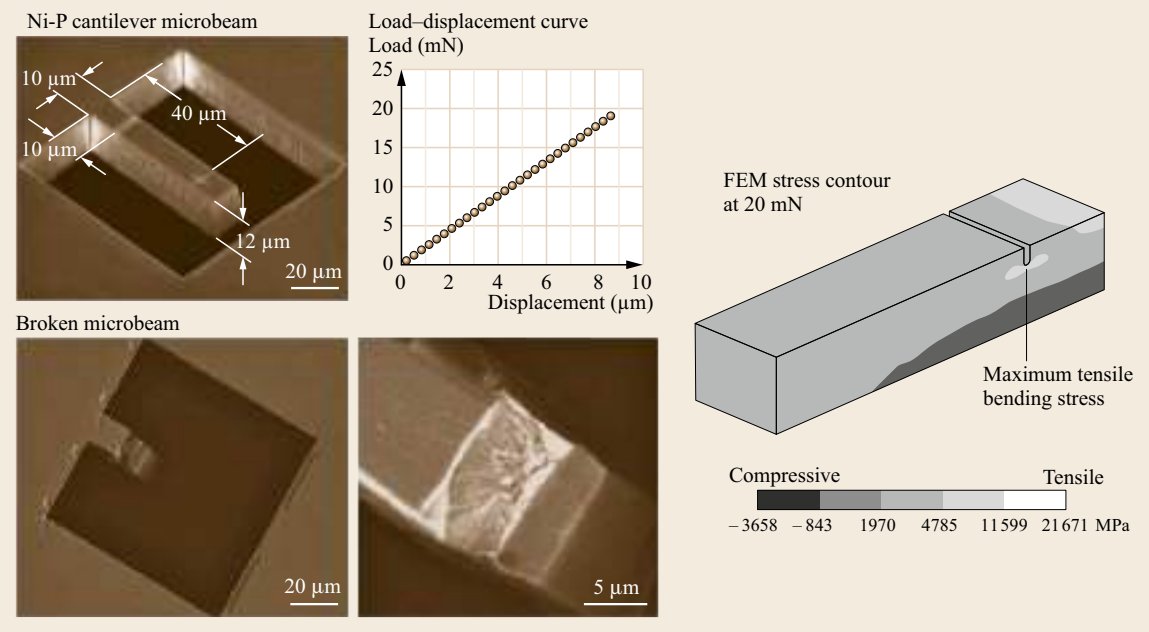


Fig. 32.18 SEM micrographs of the new and broken beams, load displacement curve and FEM stress contour for the notched Ni-P cantilever microbeam (after [32.63])

32.2.4 Indentation and Scratch Tests of Polymeric Microbeams Using a Nanoindenter

Four potential structural materials for BioMEMS devices were selected for this study, namely poly(propyl methacrylate) (PPMA), poly(methyl methacrylate) (PMMA), polystyrene (PS) and a polystyrene-nanoclay composite (PS/clay) [32.65]. The physical properties of PPMA, PMMA and PS are summarized in Table 32.3. PMMA and PS are widely used engineering polymers for a number of commercial applications. Historically, PMMA was the polymer of choice in ophthalmologic devices due to its high refractive index, hardness, and biocompatibility. Its surface can be functionalized with proteins, which promotes the bonding of tissues for in vivo implants [32.96]. PMMA is also employed in chips and valve components for immunosensors and other lab-on-a-chip applications [32.26, 97]. PPMA has a lower glass transition temperature (T_g) (35–43 °C) [32.80] than PMMA (104–106 °C) [32.81, 82], which allows for easier processing at a lower temperature.

Polystyrene is particularly desirable for BioMEMS due to its ubiquitous use in tissue culture applications. Plasma-treated polystyrene is the most commonly used material for in vitro cell biology studies of adherent cells. Thus, there is a wealth of knowledge about cellular behavior on polystyrene. In addition to tissue culture polystyrene, various surface modification techniques can be employed for functionalizing the PS surface in order to promote cell attachment and proliferation [32.96]. Oxygen plasma-modified polystyrene has been shown to improve cell growth, proliferation, and

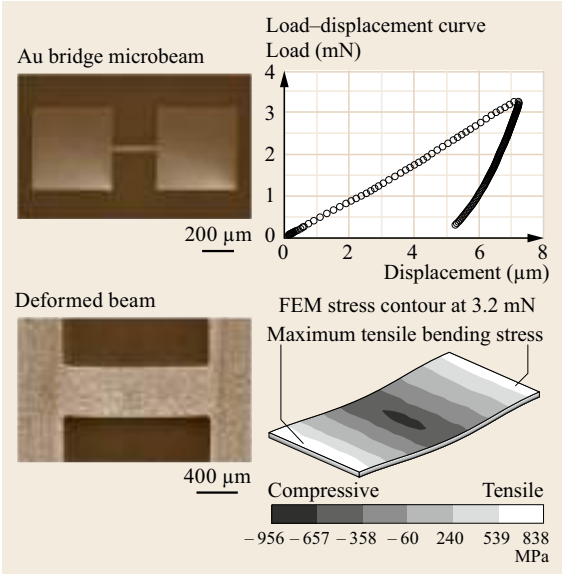
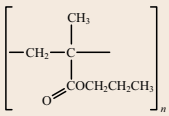
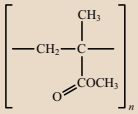
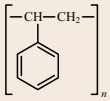


Fig. 32.19 SEM micrographs of the new and deformed beams, load displacement curve and FEM stress contour for the Au bridge microbeam (after [32.63])

Table 32.3 Summary of the physical properties of PPMA, PMMA and PS

	Structure	Density (g/cm^3)	T_g (°C)	Water absorption after 24 h (%)	Thermal expansion coefficient (K^{-1})	Thermal conductivity (0–50 °C) (W/(m K))	Heat capacity (0–100 °C) (kJ/(kg K))
PPMA		–	35–43[a]	–	–	–	–
PMMA		1.188[b]	104–106[b,c]	0.3–0.4[d] 0.1–0.3[e]	$2.0–3.0\times10^{-4}$ [c] ($< T_g$) 6.0×10^{-4} [c] ($> T_g$)	0.193[b]	1.255–1.720[b]
PS		1.040–1.065[c] (amorphous) 1.111–1.127[c] (crystalline)	100[b,c]	0.03–0.1[d]	$1.7–2.1\times10^{-4}$ [c] ($< T_g$) $5.1–6.0\times10^{-4}$ [c] ($> T_g$)	0.105–0.116[b]	1.185–1.838[b]

[a] [32.80], [b] [32.81], [c] [32.82], [d] [32.98], [e] [32.99]

expression of cellular adhesion protein proportional to the surface oxygen concentration [32.100]. The previous knowledge of cellular interactions with polystyrene makes it a logical choice to use in BioMEMS devices for cellular interactions.

Clay nanoparticles in a polymer matrix act to improve the mechanical and thermal properties as compared to the native polymer [32.101, 102]. Nanoclay composites can also be used to improve barrier resistance and improve ionic conductivity. The extent to which polymer properties are affected is determined by the clay content, polymer/clay interfacial strength, and dispersion of the clay particles. The fabrication procedure for the double-anchored and cantilever beams is presented in Appendix 32.A. SEM images of the two types of beams are shown in Fig. 32.20.

The hardness (H), elastic modulus (E) and creep of the beams were measured at the supported region of the beams [32.64, 65]. The indentation location, where the H, E and creep were measured, is indicated by an arrow in Fig. 32.20a.

Since the fabricated polymer beams were too narrow (width about 5 µm) to perform nanoscratch tests, the nanoscratch experiments were conducted on thin films of polymeric materials. PPMA, PMMA, PS and PS/clay were dissolved in anisole at various concentrations and spin coated for 1 min at 3000 RPM to achieve the desired film thickness of 500 nm. All film samples were then baked at 95 °C for 2 min to remove any residual solvent.

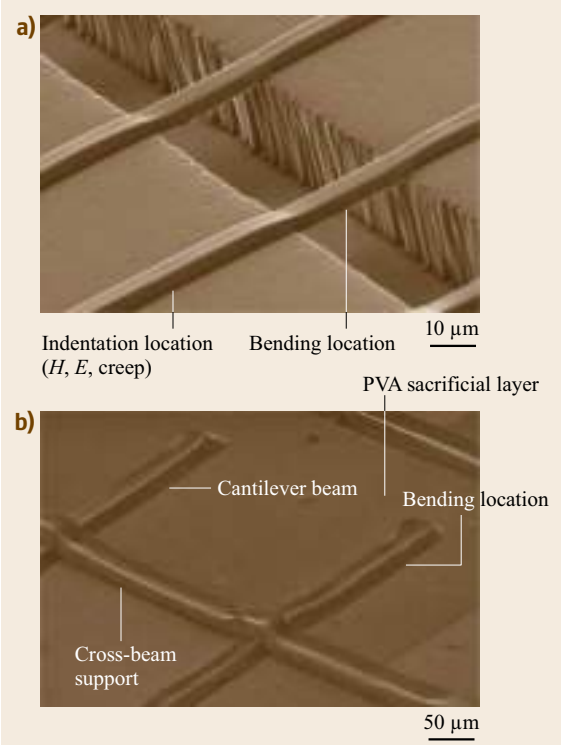


Fig. 32.20 SEM images of polystyrene **(a)** double-anchored beams, and **(b)** cantilever beams (after [32.65])

Hardness and Elastic Modulus

Figure 32.21a shows the H and E of various beams as a function of contact depth [32.65]. Five beams in one wafer sample were indented once, and the average values of H and E at a contact depth of 100 nm are reported in Table 32.4. Among the materials examined, the PS/clay nanocomposite exhibited the highest hardness and elastic modulus (of 390±50 MPa and 5.1±0.4 GPa

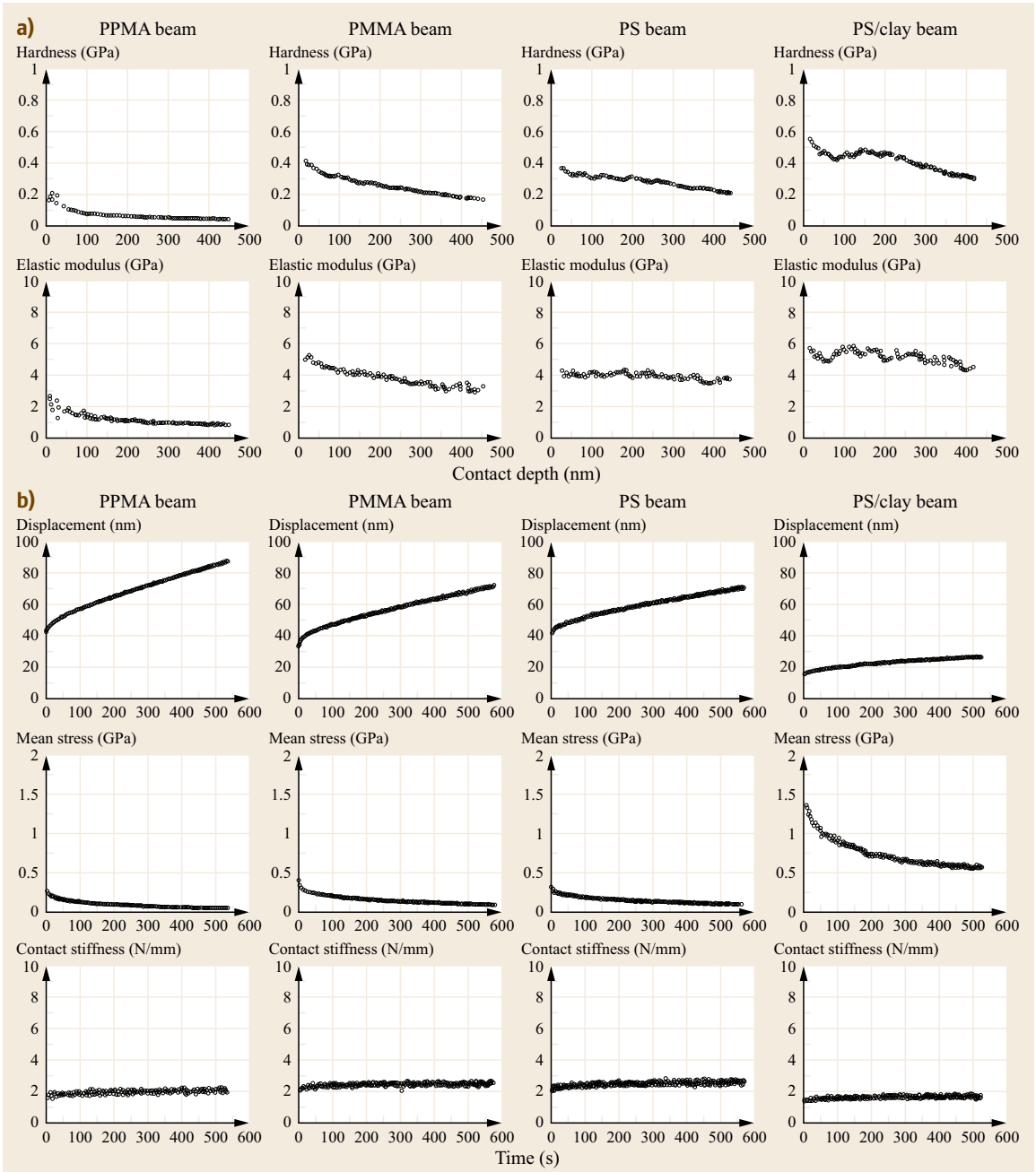


Fig. 32.21 **(a)** Hardness and elastic modulus of PPMA, PMMA, PS and PS/clay beams as a function of contact depth, **(b)** creep displacement, mean stress and contact stiffness as a function of time for PPMA, PMMA, PS and PS/clay, and **(c)** scratch depth profiles and coefficient of friction as a function of increasing normal loads for thin films of PPMA, PMMA, PS and PS/clay. SEM images were taken at three regions: at the beginning of the scratch (A), in the middle of the scratch (B) and at the end of the scratch (C) (after [32.65])

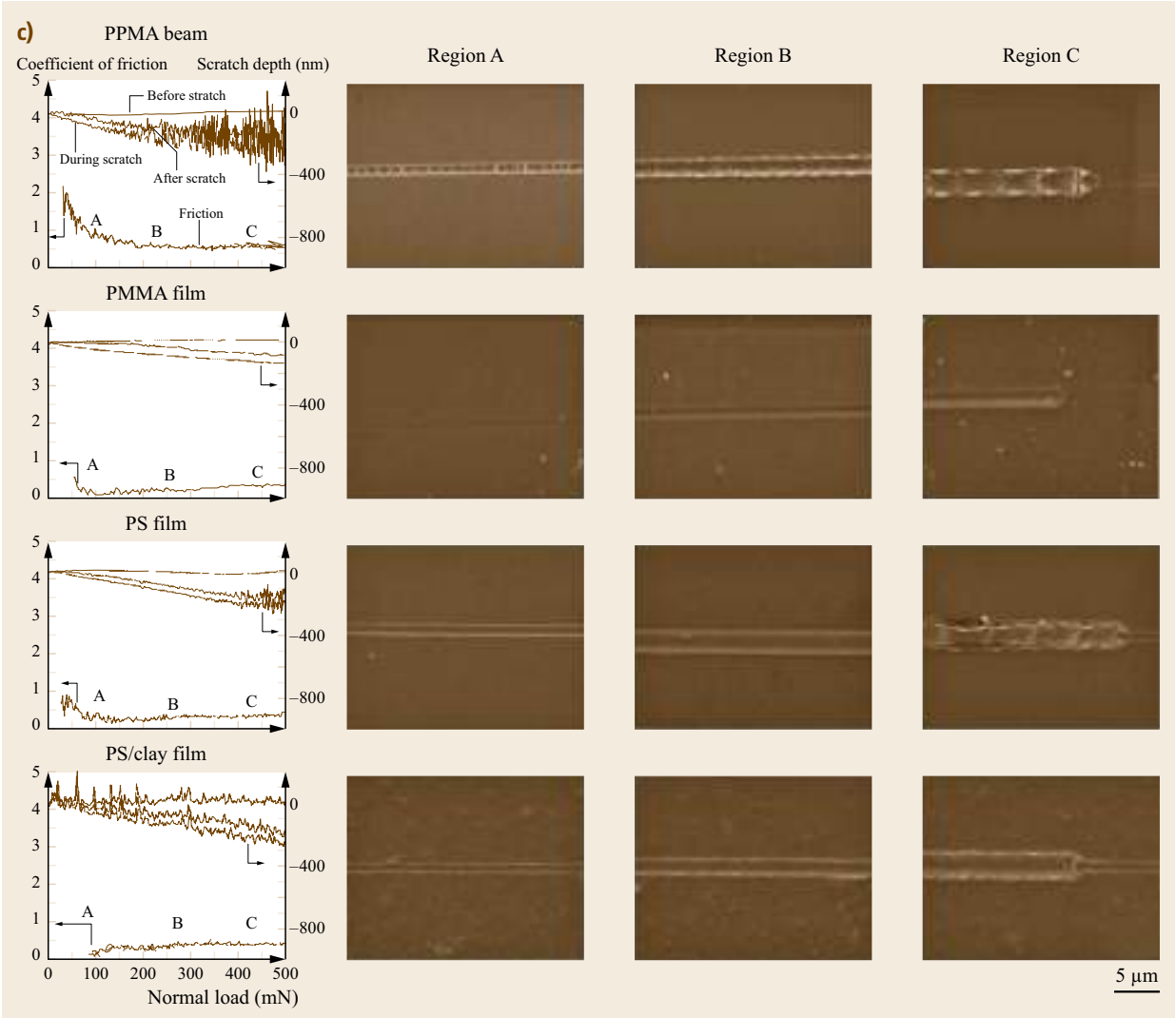


Fig. 32.21 (continued)

respectively), followed by PMMA (340±30 MPa, 4.8±0.5 GPa), PS (290±20 MPa, 3.6±0.4 GPa) and PPMA (110±30 MPa, 1.7±0.5 GPa). The standard deviation was calculated from the five indents performed.

Indentation Creep

Creep is an important aspect of polymer mechanical behavior. Figure 32.21b shows representative creep data for the supported polymer beams using CSM indentation, where the change in displacement, mean stress and contact stiffness was monitored while holding the tip at the maximum imposed load [32.65]. Conclusions on the creep response of the materials are based mainly on the displacement change at constant loading, which is directly measured in the experiment. For each material, three beams on one sample were tested, and it was found that the variation in creep response was no more than 15%. As expected for polymers, the indentation displacement increases with time. The material with the lowest hardness and modulus (PPMA) exhibited the fastest rate of displacement change (i. e., mostly creep), while the PS/clay nanocomposite (which has the highest H and E value) showed minimal creep. This shows a direct correlation between the hardness, modulus, and creep resistance. The mean stress exerted by the tip decreases with time as the polymer deforms viscoelastically underneath the tip during the hold segment. For the PS/clay nanocomposite, it is observed that this quantity decreases more rapidly during the first few seconds of the measurement compared to the other ma-

terials. This could be attributed to the presence of an additional phase (the clay filler). The filler-matrix interface could exhibit yield without necessarily causing further penetration into the material.

Scratch Resistance

The scratch behavior of thin films of the four polymers studied is presented in Fig. 32.21c. Plots on the left side show the depth profile before, during, and after a scratch, along with the coefficient of friction. SEM images were taken at three areas: at the beginning of the scratch (indicated by A on the friction profile), middle of the scratch (B) and towards the end of the scratch (C). Among the three homogeneous films, PMMA had the greatest scratch resistance as it exhibited the least penetration as well as the absence of bursts in the scratch depth profile. PMMA is followed by PS, then PPMA. This order is consistent with the observed trend for the elastic modulus and hardness. The features in the scratch depth profiles can be correlated to the morphology of the scratches. In the PPMA depth profile, the tip jumped up and down during the scratch instead of continuously ramping down. This is accounted for by the SEM images, which reveal deformation bands that are convex with respect to the scratch direction. The bands indicate plastic deformation and material pileup in front of the tip. In PS, the oscillation in the depth profile occurs only towards the end of the scratch, and the deformation bands are observed only in Region C. For PMMA, the smooth depth profile corresponds to a clear scratch track, indicating that minimal pileup occurred. The nanocomposite exhibits enhanced scratch resistance relative to the unfilled PS. The peaks and valleys in the depth profile come from the roughness of the surface and not from scratch-induced material displacement, which is confirmed from the SEM images.

32.2.5 Bending Tests of Polymeric Microbeams Using a Nanoindenter

Suspended Beams

For each material, five suspended beams in one sample were measured, and representative load-displacement data are presented in Fig. 32.22. In this case, the displacement is due to beam bending. The loading profile can be fit into a straight line, indicating that bending at this load range is fully elastic. From the slope of this line and (32.5), the calculated elastic moduli are 0.7±0.2, 1.9±0.3, 1.9±0.1 and 4.6±0.3 GPa for PPMA, PMMA, PS and PS/clay respectively. The standard deviation was based on five measurements.

The elastic moduli from beam bending are compared to values from nanoindentation and the available

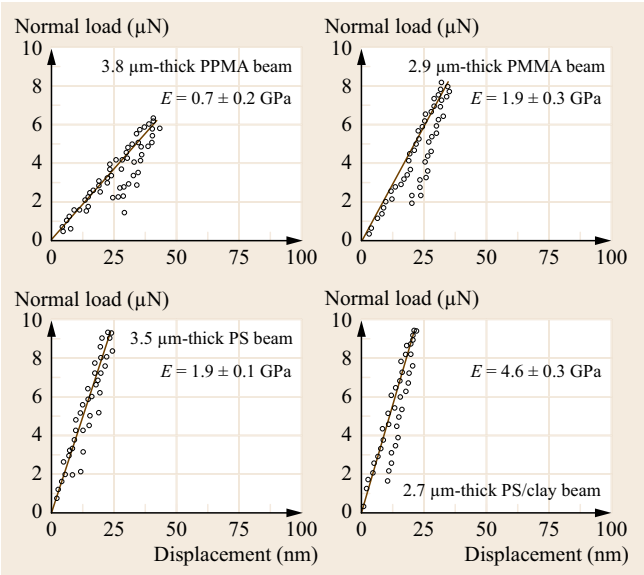


Fig. 32.22 Normal beam bending results for PPMA, PMMA, PS and PS/clay at a maximum load of 10 μN (after [32.65])

bulk literature values. The consistency in the obtained trend in E for the four materials using the three methods provides validation for the beam bending technique. In addition, the beam bending method has been successfully employed on various metal and ceramic materials, as presented earlier [32.63]. The slightly lower elastic moduli calculated from beam bending tests (relative to nanoindentation and bulk literature data) can be attributed to several reasons. Some deformation may have occurred in the boundary between the supported and the suspended region of the beam due to stress concentration at this junction. A small amount of plastic deformation of the beam surface may occur even though the indenter tip was coated with PMMA. This could reduce the slope of the loading curve.

Effect of Soaking and Temperature. The effects of aqueous medium and human body temperature (37.5 °C) were investigated on both the suspended and supported beams. Figure 32.23a is a summary of the effects of beam soaking and elevated temperature on the hardness and elastic modulus as obtained by nanoindentation on the supported part of the beam. Among the four materials studied, soaking had the greatest effect on PPMA. On the other hand, the H and E of the other materials (PMMA, PS and PS/clay) were not significantly affected. The properties of the PPMA beam were adversely affected when the indentation temperature was increased to 37.5 °C. This is because the glass transition temperature (T_g) of PPMA is within 35–43 °C [32.80], such that the glassy to rubbery tran-

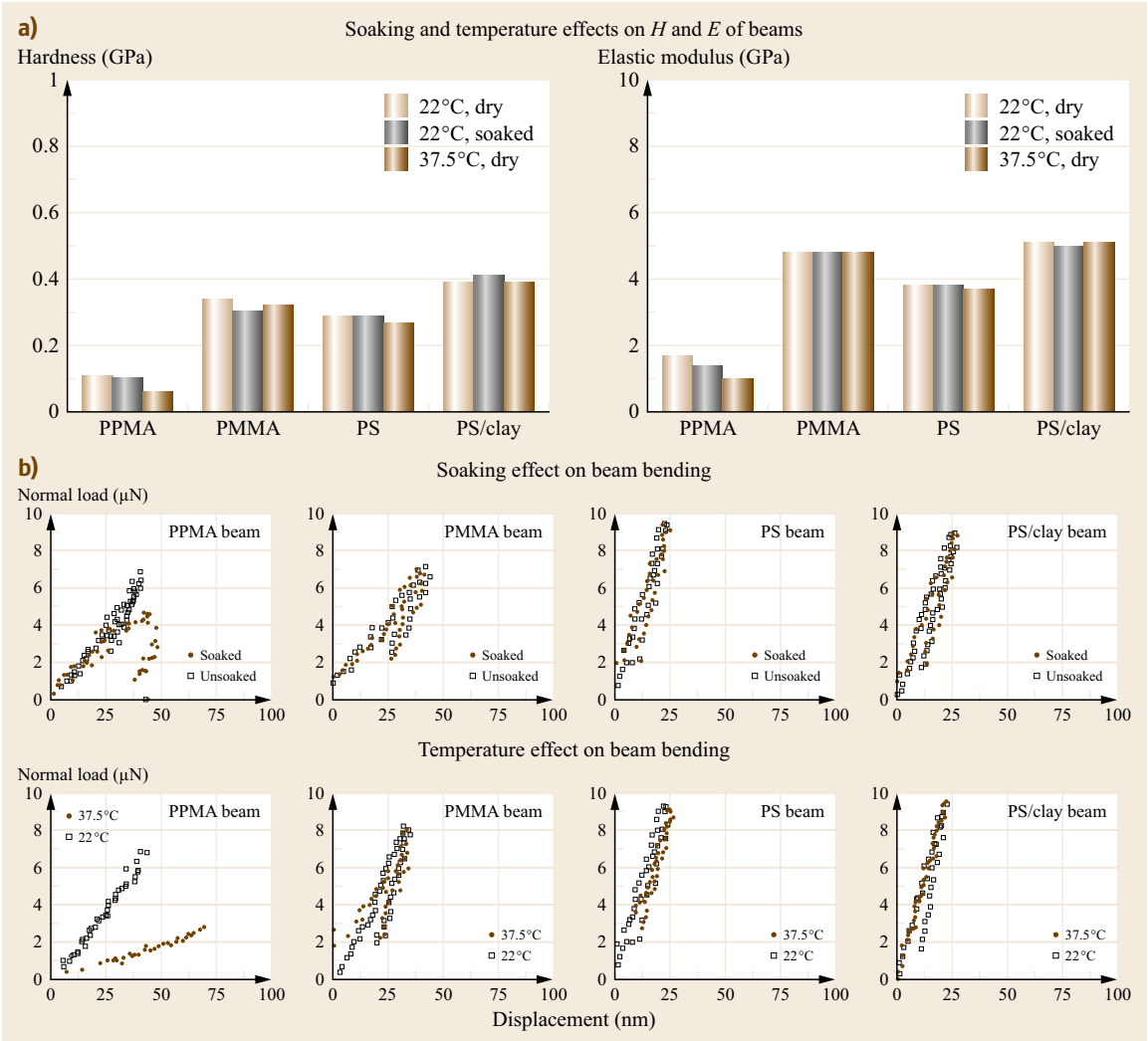


Fig. 32.23 **(a)** Effect of soaking and elevated temperature on suspended beams and **(b)** effect of soaking and elevated temperature on the hardness and elastic modulus of supported beams (after [32.65])

sition is being observed here. On the other hand, no significant decrease in H and E was observed on the other three materials since their values of T_g are much higher, such that the indent was performed well within the glassy regime.

In Fig. 32.23b, beam bending data are shown for each of the four polymers investigated. The PPMA beam is affected by soaking in deionized (DI) water, as shown by the appearance of a large amount of hysteresis in the load-displacement data. Human body temperature has an adverse effect on the stiffness of the PPMA beam. These are consistent with results from indentation. From a materials selection standpoint, these are important findings as they imply that the performance of PPMA structures would be compromised if they were in a device subjected to either soaking in an aqueous medium or above ambient temperatures.

Normal Beam Bending at Elevated Loads. Bending experiments were performed with applied loads in the millinewton range in order to investigate the behavior of polymer beams beyond the elastic deformation regime. The upper half of Fig. 32.24 shows normal load-displacement profiles at maximum imposed loads of 0.5, 1.5 and 3.0 mN applied on three different beams in one wafer sample. For PPMA, PMMA and PS, the loading portion overlaps, indicating uniform beam quality. Slight variability in the loading profile is observed in the PS/clay composite, which is attributed to variable

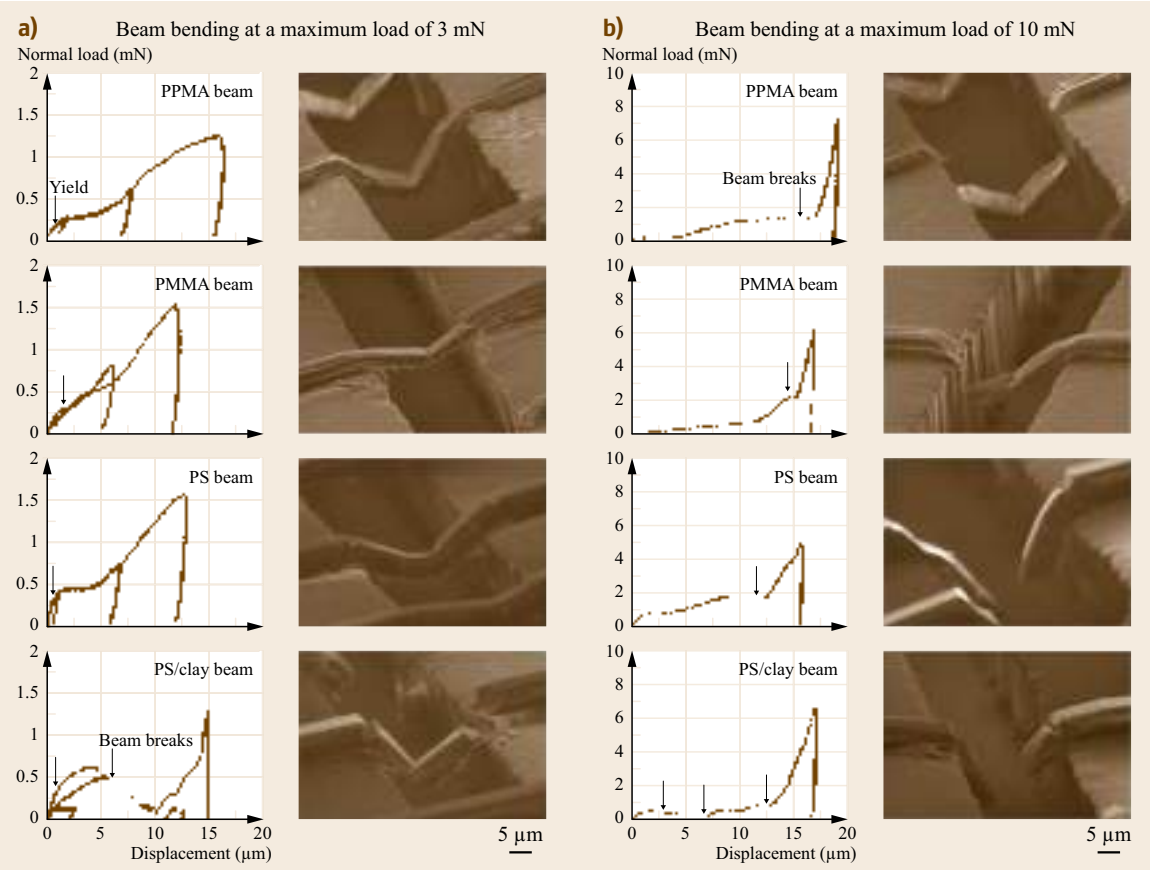


Fig. 32.24a,b Normal beam bending at elevated loads. **(a)** Normal load-displacement profiles at a maximum load of 0.5, 1.5 and 3.0 mN, *arrows* indicate the onset of yield and breaking of the beams. For plots with multiple profiles **(a)**, the SEM image corresponds to the beam subjected to the highest load (3 mN). **(b)** Normal load-displacement profiles and SEM images for beams subjected to 10 mN maximum load (after [32.65])

interfacial adhesion between the filler and the matrix and inhomogeneities in the dispersion of the clay particles.

The first inflection in the load-displacement profiles implies the onset of yield, and its location is indicated by arrows. The three unfilled polymers (PPMA, PMMA and PS) exhibit ductility, as seen in the SEM images for the beams tested with an applied load of 3 mN. The beams stretch at the ends and underneath the center of the beam. These correspond to regions subjected to maximum tensile stress [32.63]. It is worth noting that the beams deformed symmetrically, indicating that the load was applied equidistant from the clamping points. The PS/clay nanocomposite beam exhibited slight yielding, followed by breaking, as indicated by an arrow in the load-displacement profile. After breaking, the load continued to increase because a predefined load is imposed by the nanoindenter. The corresponding SEM image for PS/clay shows a flat fracture surface, indicating that the crack propagated perpendicular to the direction of the applied stress.

From the load-displacement curves, the yield strength was evaluated using the following equation [32.86]

$$\sigma = \frac{3}{4}\frac{Fl}{bh^2}\,, \tag{32.13}$$

where F is the load corresponding to yield of the beam (as indicated by arrows in the load–displacement profile), and l, b and h are the beam's length, width and height respectively. It is assumed that the beam ends are clamped. The yield strength values listed in Table 32.4 are in the same order of magnitude as the bulk yield strength data in the literature.

Bending data at a maximum load of 10 mN are shown in the lower half of Fig. 32.24. As with the bending tests performed at 3 mN, symmetric deformation and failure is observed in the beams. The three unfilled

Table 32.4 Mechanical properties of the polymer beams under investigation. All values have $\pm\sigma$ of about 10% based on measurements on five samples for nanoindentation and three to five samples for beam bending except for the 10–40% variation in breaking strength

Polymer		PPMA	PMMA	PS	PS/clay
Poisson's ratio		–	0.35[a]	0.325[b]	–
Elastic modulus (E) (GPa)	Bulk (literature)	–	3.1–3.3[c,d]	3.2–3.4[c]	–
	Nanoindentation (measured)	1.7	4.8	3.6	5.1
	Normal beam bending (measured)	0.7	1.9	1.9	4.6
	Lateral beam bending (measured)	–	–	0.2	0.6
Hardness (H) (GPa)	Bulk (literature)	–	195[c]	110[d]	–
	Nanoindentation (measured)	110	340	290	390
Yield strength (σ_{ys}) (MPa)	Bulk (literature)	–	53.8–73.1[e]	–	–
	Normal beam bending (measured)	54	71	66	95
	Lateral beam bending (measured)	–	–	21	64
Breaking strength (σ_b) (MPa)	Tensile strength (literature)	–	48–76[d]	30–60[d]	–
	Flexural strength (literature)	–	–	95[d]	–
	Normal beam bending (measured)[f]	600	944	271	128
	Lateral beam bending (measured)	–	–	–	110

[a] [32.99], [b] [32.103], [c] [32.81], [d] [32.82], [e] [32.104], [f] Values may be an overestimate due to limitations of the technique and the geometry of the beams

polymers stretched considerably prior to fracture, and the PS/clay nanocomposite beam exhibited brittle failure as observed in the lower-load experiment.

For brittle materials, the elastic regime is immediately followed by breaking, and the breaking strength can be evaluated using (32.13). Even though plastic deformation was observed in all materials, (32.13) was assumed valid, and the force corresponding to breaking is taken as the load observed prior to the abrupt jump in displacement. The calculated breaking strength (as shown in Table 32.4) is higher and does not appear to correlate with the trends observed for the other mechanical properties reported. This can be attributed to a number of reasons. The load-displacement response may not purely be from beam bending. A higher load can be imposed without beam breaking due to plastic deformation on the beam surface. At elevated loads, a larger contact area between the tip and beam is expected such that permanent deformation on the surface can be significant. In addition, the compliant trench material (SU-8) can help to accommodate a higher stress prior to breaking of the beam.

Lateral Bending of the Cantilever Beams

Figure 32.25 presents the results of the lateral bending experiments on PS and PS/clay cantilever beams. In the left column are the plots for the lateral force recorded as a function of the lateral displacement. From the linear elastic regime (as indicated by the straight line drawn through the data points), the modulus can be evaluated using (32.11). The resulting elastic modulus of PS/clay is three times that of PS, as shown in Table 32.4. This was obtained from measurements on two beams for each material. However, the elastic modulus values obtained from this experiment are lower than those obtained from nanoindentation and beam bending in the normal direction. This is attributed to the applied pressure coming from the addition of the crossbeam reinforcement, inducing deformation on the anchor. In addition, a small curvature is observed on the beam. This diminishes the beam width, which is the dominant term in the equation for evaluating the modulus ((32.5) and (32.11)).

From the load-displacement curves, the yield strength was evaluated using [32.86]

$$\sigma = \frac{6Fl}{bh^2}\,, \tag{32.14}$$

where F, l, b and h are defined similarly as in (32.13). The force used for calculating the yield strength was taken at the point where a change in slope was observed (as indicated by the arrows on the left). The results are listed in Table 32.4. They are lower than the values obtained from normal bending, and this could be attributed to the same causes mentioned above.

For PS, a steady decrease in the lateral load was observed at 160 μm, while for the PS/clay beam, the drop at 80 μm was abrupt (as indicated by the arrows on the right). SEM images reveal that for PS, this corresponds to crack formation at the base of the beam that did not run throughout the entire thickness of the beam. However, for PS/clay, the beam broke. This implies that the addition of the nanoclay filler induces embrittlement of the beam.

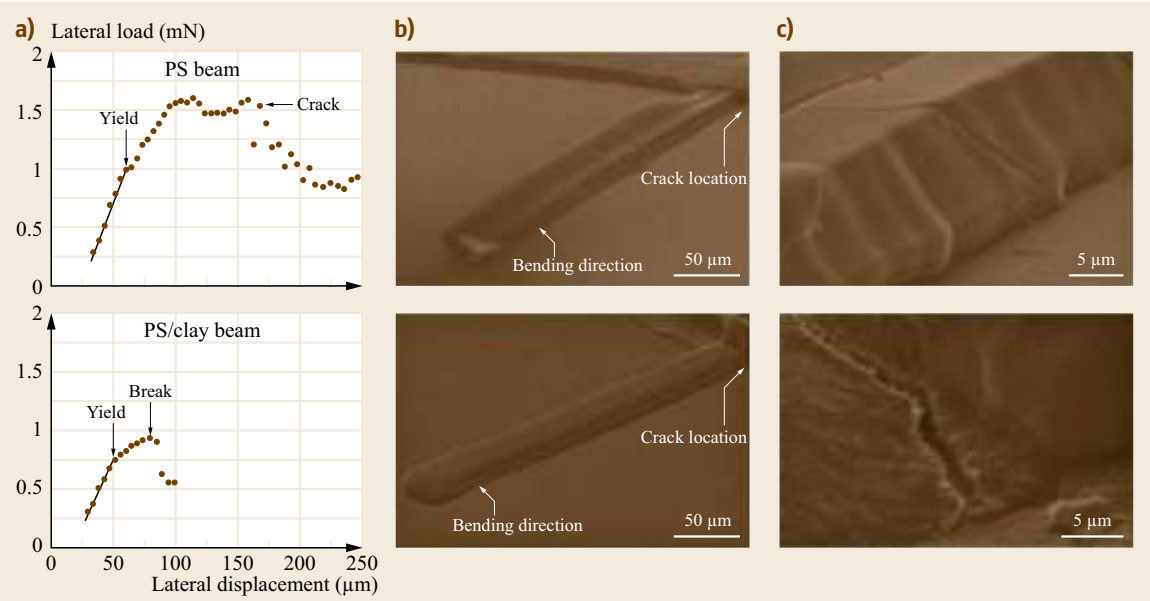


Fig. 32.25a–c Lateral bending of PS and PS/clay cantilever beams. **(a)** Lateral force plotted as a function of the lateral displacement; *arrows* indicate the onset of yield for both PS and PS/clay, cracking for PS and breaking for PS/clay. **(b)** SEM images of the cantilever beams indicate the bending direction and the location of the cracks. The images in **(c)** show the cracks at a higher magnification (after [32.65])

Equation (32.14) is assumed valid for evaluating the breaking strength of this beam. The force used in the calculation was the load prior to the abrupt decrease. The breaking strength for PS/clay obtained from this technique (listed in Table 32.4) is lower than the value from the normal bending experiment. This is possibly due to the beam geometry and processing conditions, as described earlier. However, the large difference in the calculated yield strength between the unfilled and filled PS demonstrates that the technique can adequately differentiate between material compositions.

32.3 Finite Element Analysis of Nanostructures with Roughness and Scratches

Micro/nanostructures have some surface topography and local scratches dependent upon the manufacturing process. Surface roughness and local scratches may compromise the reliability of the devices, and their effect needs to be studied. Finite element modeling is used to perform parametric analysis to study the effect of surface roughness and scratches in different well-defined forms on tensile stresses, which are responsible for crack propagation [32.71, 72]. The analysis has been carried out on trapezoidal beams supported at the bottom whose data (on Si and SiO_2 nanobeams) have been presented earlier in Table 32.2.

The finite element analysis has been carried out by using the static analysis of the ANSYS 5.7 model, which calculates the deflections and stresses produced by applied loading. The type of element selected for the study was SOLID95 type, which allows the use of different shapes without much loss of accuracy. This element is 3-D with 20 nodes, each node having three degrees of freedom, which implies translation in the x, y and z directions. The nanobeam cross-section is divided into six elements each along the width and the thickness and forty elements along the length. SOLID95 has plasticity, creep, stress stiffening, large deflection and large strain capabilities. The large displacement analysis is used for large loads. The mesh is kept finer near the asperities and the scratches in order to take into account variation in the bending stresses. The beam materials studied are made of single-crystal silicon (110) and SiO_2 films whose data have been presented earlier. Based on the bending experiments presented earlier, the beam materials can be assumed to be linearly elastic isotropic materials. Young's modulus of elasticity (E) and Poisson's ratio (ν) for Si and SiO_2 are

169 GPa [32.92] and 0.28 [32.88], and 73 GPa [32.93] and 0.17 [32.93] respectively. A sample nanobeam of silicon was chosen for performing most of the analysis as silicon is the most widely used MEMS/NEMS material. The cross-section of the fabricated beams used in the experiment is trapezoidal and supported at the bottom so nanobeams with trapezoidal cross-section are modeled (Fig. 32.10). The following dimensions are used: $w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, and $P = 6\,\mu\text{m}$. In the boundary conditions, the displacements are constrained in all directions on the bottom surface for 1 μm from each end. A point load applied at the center of the beam is simulated with the load being applied at three closely located central nodes on the beam used. It has been observed from the experimental results that the Si nanobeam breaks at around 80 μN. Therefore, in this analysis, a nominal load of 70 μN is selected. At this load, deformations are large, and a large displacement option is used.

To study the effect of surface roughness and scratches on the maximum bending stresses the following cases were studied. First, the semicircular and grooved asperities in the longitudinal direction with defined geometrical parameters are analyzed (Fig. 32.26a). Next, semicircular asperities and scratches placed along the transverse direction at a distance c from the end and separated by pitch p from each other are analyzed (Fig. 32.26b). Lastly, the beam material is assumed to be either purely elastic, elastic-plastic, or elastic-perfectly plastic. In the following, we begin with the stress distribution in smooth nanobeams followed by the effect of surface roughness in the longitudinal and transverse directions and scratches in the transverse direction.

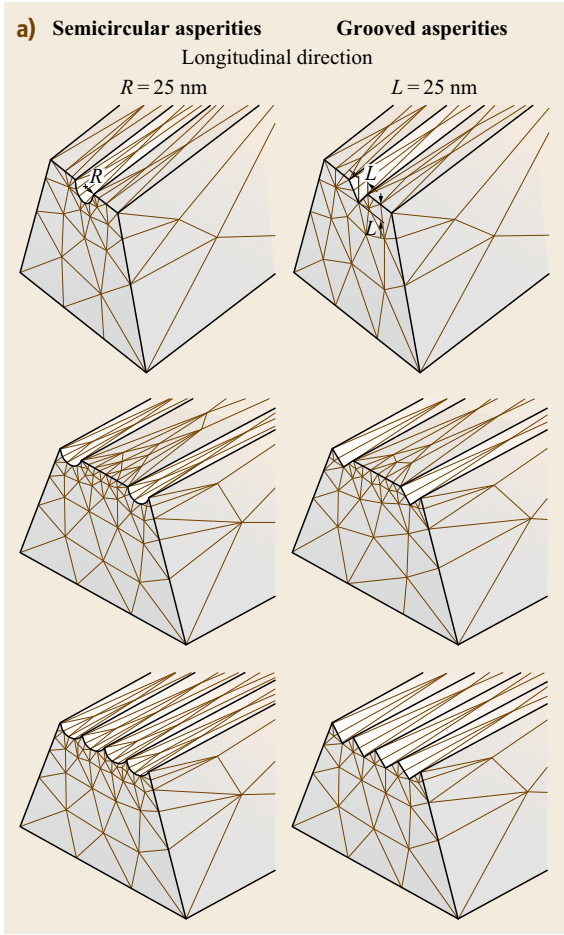


Fig. 32.26 (a) Plots showing the geometries of modeled roughness – semicircular and grooved asperities along the nanobeam length with defined geometrical parameters. (b) Schematic showing semicircular asperities and scratches in the transverse direction followed by the illustration of the mesh created on the beam with fine mesh near the asperities and the scratches. Also shown are the semicircular asperities and scratches at different pitch values

32.3.1 Stress Distribution in a Smooth Nanobeam

Figure 32.27 shows the stress and vertical displacement contours for a nanobeam supported at the bottom and loaded at the center [32.71, 72]. As expected, the maximum tensile stress occurs at the ends while the maximum compressive stress occurs under the load at the center. Stress contours obtained at a section of the beam from the front and side are also shown. In the beam cross-section, the stresses remain constant at a given vertical distance from one side to another and change with a change in vertical location. This can be explained due to the fact that the bending moment is constant at a particular cross-section, so the stress is only dependent on the distance from the neutral axis. However, in cross-section A–A the high tensile and compressive stresses are localized near the end of the beam at top and bottom respectively, whereas the lower values are spread out away from the ends. High values of tensile stresses occur near the ends because of high bending moment.

32.3.2 Effect of Roughness in the Longitudinal Direction

The roughness in the form of semicircular and grooved asperities in the longitudinal direction on the maximum bending stresses are analyzed [32.72]. The radius R and depth L are kept fixed at 25 nm while the number of asperities is varied, and their effect is observed on the maximum bending stresses. Figure 32.28 shows

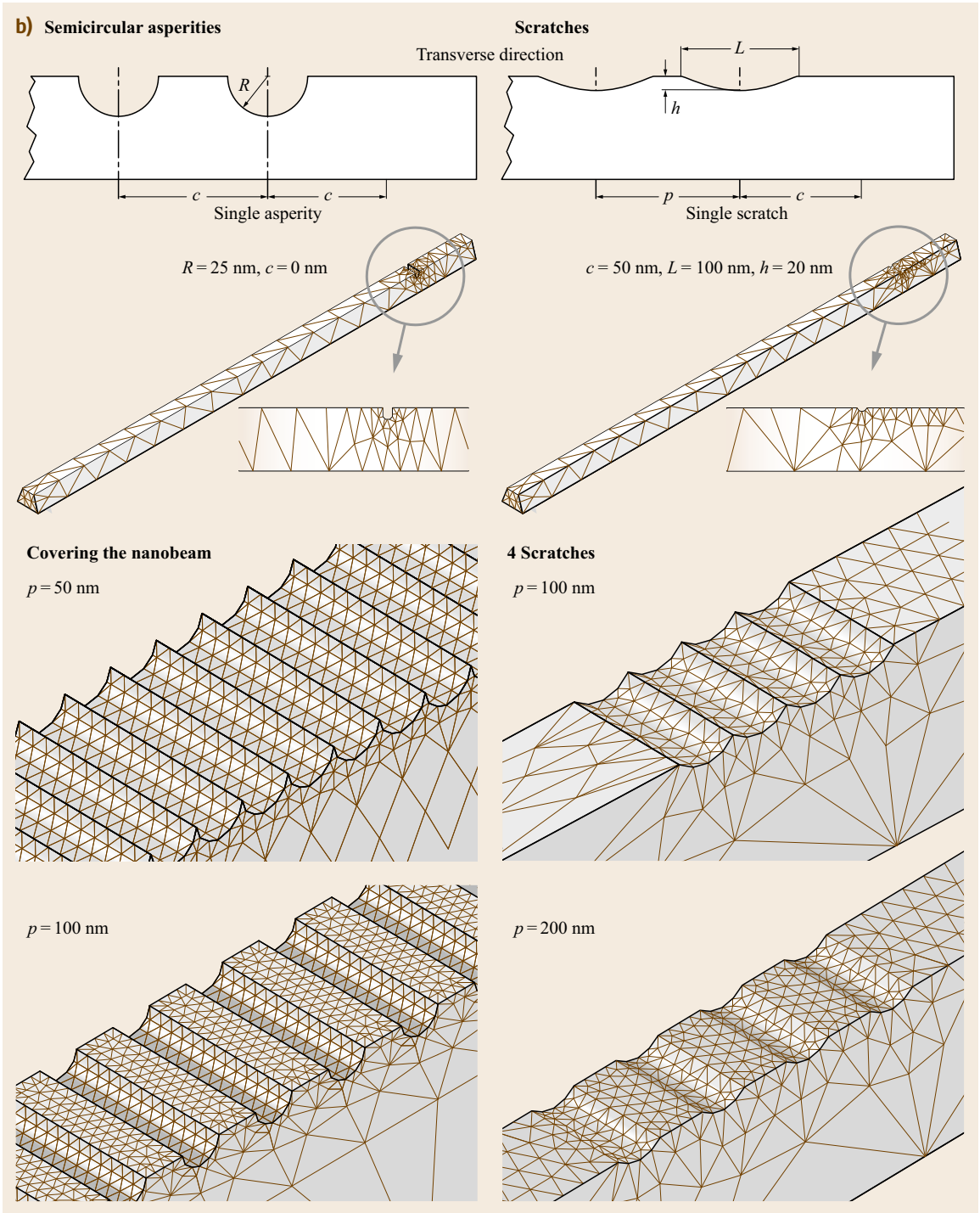


Fig. 32.26 (continued)

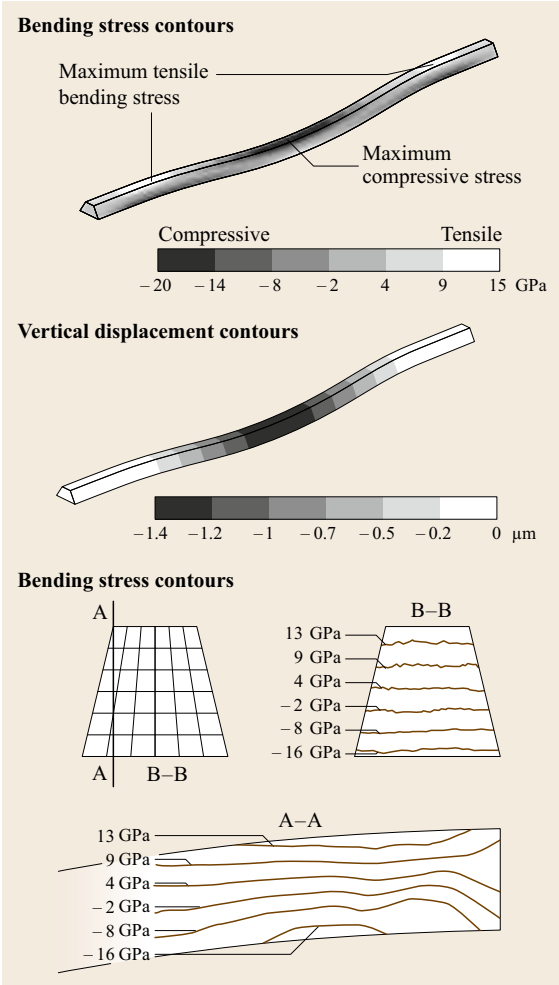


Fig. 32.27 Bending stress contours, vertical displacement contours, and bending stress contours after loading trapezoidal Si nanobeams ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, $\nu = 0.28$) at $70\,\mu\text{N}$ load (after [32.72])

the variation of maximum bending stresses as a function of asperity shape and the number of asperities. The maximum bending stresses increase as the asperity number increases for both semicircular and grooved asperities. This can be attributed to the fact that as asperity number increases, the moment of inertia decreases for that cross-section. Also the distance from the neutral axis increases because the neutral axis shifts downwards. Both these factors lead to the increase in the maximum bending stresses, and this effect is more pronounced in the case of a semicircular asperity as it exhibits a higher value of maximum bending stress

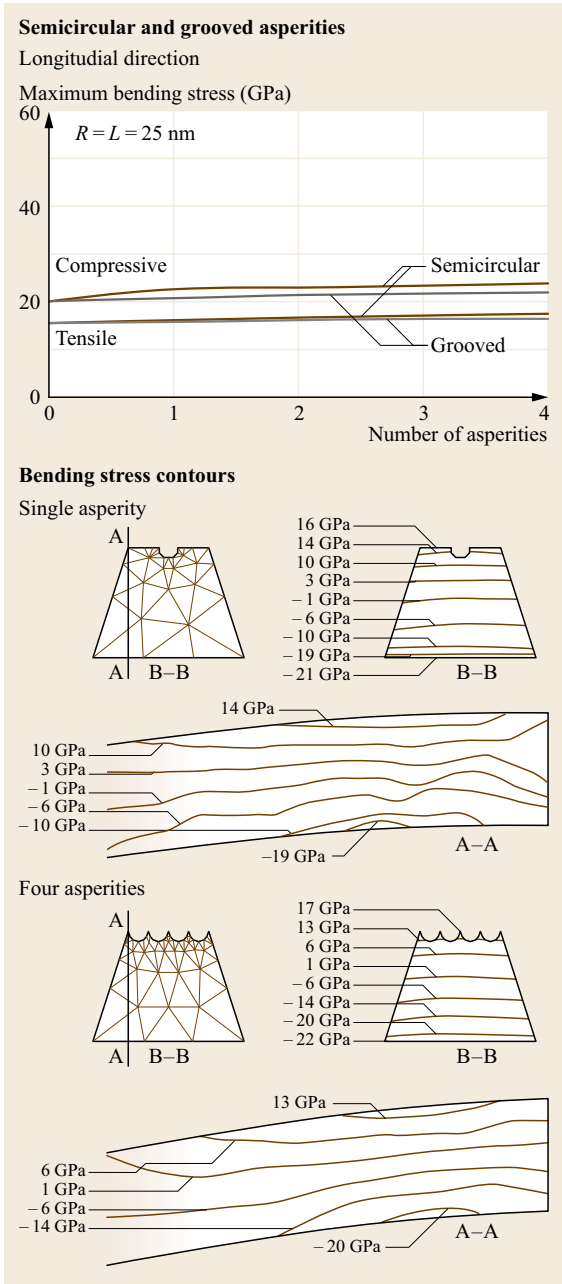


Fig. 32.28 Effect of longitudinal semicircular and grooved asperities in different numbers on maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, $\nu = 0.28$, load $= 70\,\mu\text{N}$). Bending stress contours obtained in the beam with semicircular single asperity and four adjacent asperities of $R = 25\,\text{nm}$ (after [32.72])

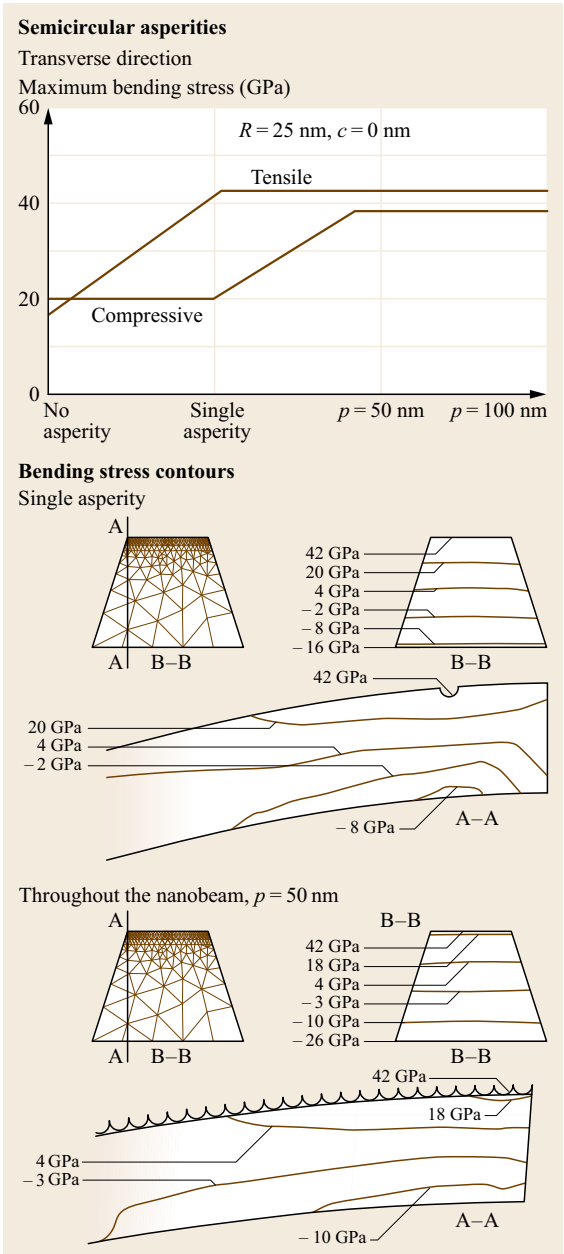


Fig. 32.29 Effect of transverse semicircular asperities located at different pitch values on the maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, $\nu = 0.28$, load $= 70\,\mu\text{N}$). Bending stress contours obtained in the beam with semicircular single asperity and semicircular asperities throughout the nanobeam surface at $p = 50\,\text{nm}$ (after [32.72])

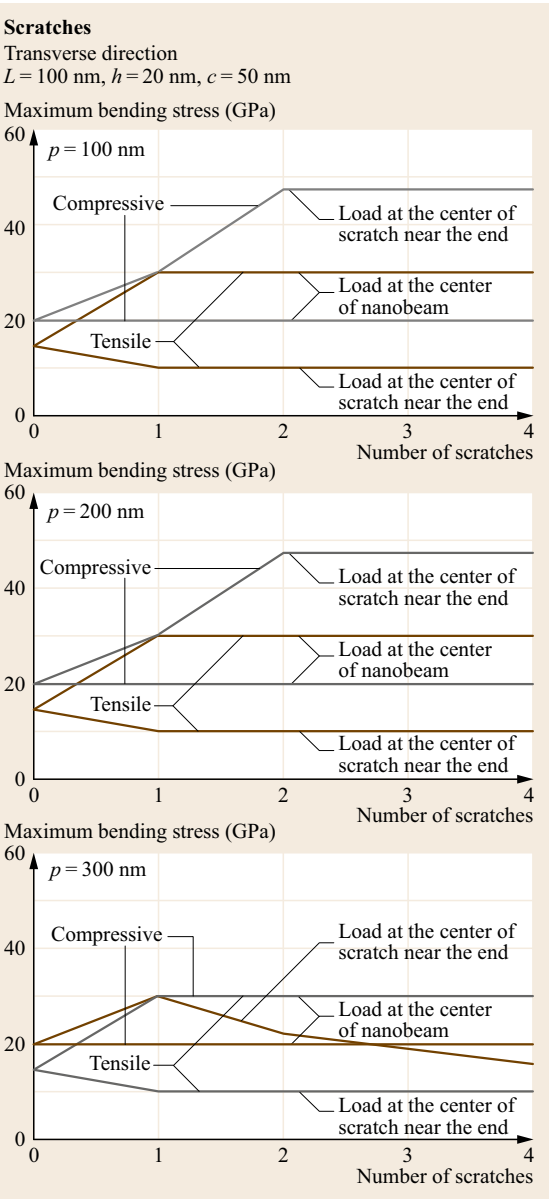


Fig. 32.30 Effect of number of scratches along with the variation in the pitch on the maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, $\nu = 0.28$, load $= 70\,\mu\text{N}$). Also shown is the effect of load when applied at the center of the beam and at the center of the scratch near the end (after [32.72])

than that in a grooved asperity. Figure 32.28 shows the stress contours obtained at a section of the beam from the front and from the side for both cases when we have a single semicircular asperity and when four adjacent semicircular asperities are present. Trends are similar to that observed earlier for a smooth nanobeam (Fig. 32.27).

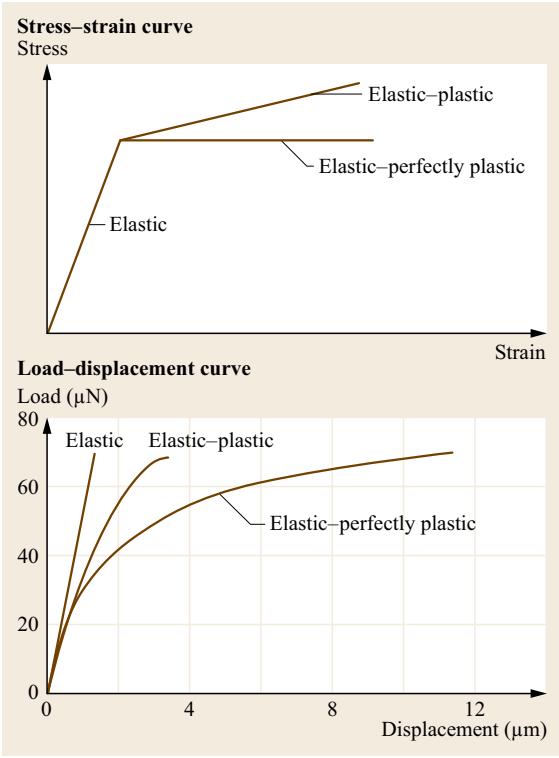


Fig. 32.31 Schematic representation of stress-strain curves and load-displacement curves for material when it is elastic, elastic-plastic, or elastic-perfectly plastic for a Si nanobeam ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, tangent modulus in plastic range $= 0.5\,\text{E}$, $\nu = 0.28$) (after [32.72])

32.3.3 Effect of Roughness in the Transverse Direction and Scratches

We analyze semicircular asperities when placed along the transverse direction followed by the effect of scratches on the maximum bending stresses in varying numbers and different pitch [32.72]. In the analysis of semicircular transverse asperities three cases were considered, which included a single asperity, asperities throughout the nanobeam surface separated by pitch equal to 50 nm and pitch equal to 100 nm. In all of these cases, the c value was kept equal to 0 nm. Figure 32.29 shows that the value of maximum tensile stress is 42 GPa, which is much larger than the maximum tensile stress value with no asperity of 16 GPa or when the semicircular asperity is present in the longitudinal direction. It is also observed that the maximum tensile stress does not vary with the number of asperities or the pitch, while the maximum compressive stress does increase dramatically for the asperities present throughout the beam surface from its value when a single asperity is present. Maximum tensile stress occurs at the ends. An increase in p does not add any asperities at the ends whereas asperities are added in the central region where compressive stresses are maximum. The semicircular asperities present at the center cause the local perturbation in the stress distribution at the center of the asperity where load is being applied leading to a high value of maximum compressive stress [32.105]. Figure 32.29 also shows the stress contours obtained at a section of the beam from front and from the side for both cases when there is a single semicircular asperity and when asperities are present throughout the beam surface at a pitch equal to 50 nm. Trends are similar to that observed earlier for a smooth nanobeam (Fig. 32.27).

In the study pertaining to scratches, the number of scratches are varied along with the variation in the pitch. Furthermore the load is applied at the center of the beam and at the center of the scratch near the end for all the cases. In all of these cases, the c value was kept equal to 50 nm, and the L value was equal to 100 nm with the h value being 20 nm. Figure 32.30 shows that the value of maximum tensile stress remains almost the same with the number of scratches for both types of loading, that is, when load is applied at the center of the beam and at the center of the scratch near the end. This is because the maximum tensile stress occurs at the beam ends, no matter where the load gets applied. But the presence of a scratch does increase the maximum tensile stress as compared to its value for a smooth nanobeam, although the number of scratches no longer matters as the maximum tensile stress occurring at the nanobeam end is unaffected by the presence of more scratches beyond the first scratch in the direction towards the center. The value of the tensile stress is much lower when the load is applied at the center of the scratch, and it can be explained as follows. The negative bending moment at the end near the load applied decreases with load offset after two-thirds of the length of the beam [32.106]. Since this negative bending moment is responsible for tensile stresses, their behavior with the load offset is the same as the negative bending moment. Also, the value of maximum compressive stress when load is applied at the center of the nanobeam remains almost the same, as the center geometry is unchanged due to the number of scratches, and hence the maximum compressive stress occurring below the load at the center is same. On the other hand when the load is applied at the center of the scratch we observe that the maximum compressive stress increases dramatically because the local perturbation in the stress distribution at the center of the scratch where load is being applied leads to a high value of maximum compressive

Table 32.5 Stresses and displacements for materials which are elastic, elastic-plastic or elastic-perfectly plastic (load = 70 μN, $w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6$ μm, $R = 25$ nm, $E = 169$ GPa, tangent modulus in plastic range $= 0.5E$, $\nu = 0.28$)

	Elastic		Elastic–plastic		Elastic–perfectly plastic	
	Smooth nanobeam	Single semicircular longitudinal asperity	Smooth nanobeam	Single semicircular longitudinal asperity	Smooth nanobeam	Single semicircular longitudinal asperity
Maximum von Mises stress (GPa)	18.2	19.3	13.5	15.2	7.8	9.1
Maximum displacement (μm)	1.34	1.40	3.35	3.65	11.5	12.3

stress [32.105]. It increases further with the number of scratches and then levels off. This can be attributed to the fact that when there is another scratch present close to the scratch near the end, the stress concentration is more, as the effect of local perturbation in the stress distribution is more significant. However, this effect is insignificant when more than two scratches are present.

Now we address the effect of pitch on the maximum compressive stress when the load is applied at the center of the scratch near the end. When the pitch is up to a value of 200 nm the maximum compressive stress increases with the number of scratches, as discussed earlier. On the other hand, when the pitch value goes beyond 225 nm this effect is reversed. This is because the presence of another scratch no longer affects the local perturbation in the stress distribution at the scratch near the end. Instead more scratches at a fair distance distribute the maximum compressive stress at the scratch near the end, and the stress starts going down. Such observations of maximum bending stresses can help in identifying the number of asperities and scratches allowed, separated by an optimum distance from each other.

32.3.4 Effect on Stresses and Displacements for Materials Which are Elastic, Elastic–Plastic or Elastic–Perfectly Plastic

This section deals with the beam modeled as elastic, elastic-plastic and elastic-perfectly plastic to observe the variation in the stresses and displacements from an elastic model used so far [32.72]. Figure 32.31 shows the typical stress-strain curves for the three types of deformation regimes and their corresponding load-displacement curves obtained from the model of an Si nanobeam, which are found to exhibit same trends.

Table 32.5 shows the comparison of the maximum von Mises stress and maximum displacements for both a smooth nanobeam and a nanobeam with a defined roughness, which is a single semicircular longitudinal asperity of R value equal to 25 nm for the three different models. It is observed that the maximum value of stress is obtained at a given load for elastic material whereas the displacement is maximum for elastic-perfectly plastic material. Also the pattern that the maximum bending stress value increases for a rough nanobeam still holds true in the other models as well.

32.4 Summary

Mechanical properties of nanostructures at relevant scales are required to design realistic MEMS/NEMS and BioMEMS/BioNEMS devices. Most mechanical properties are scale dependent. Micro/nanomechanical properties, hardness, elastic modulus, and scratch resistance of bulk materials of undoped single-crystal silicon (Si) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au were presented. It was found that the SiC film exhibits higher hardness, elastic modulus, and scratch resistance as compared to other materials.

Bending tests have been performed on the Si and SiO_2 nanobeams and Ni-P, Au, PPMA and PMMA, PS and PS/clay microbeams using an AFM and a depth-sensing nanoindenter respectively. The bending tests were used to evaluate elastic modulus, bending strength (fracture stress), fracture toughness (K_{IC}), and fatigue strength of the beam materials. The Si and SiO_2 nanobeams exhibited an elastic linear response with sudden brittle fracture. The notched Ni-P beam showed linear deformation behavior followed by abrupt failure. The Au beam showed elastic-plastic deformation

behavior. Measured elastic modulus values for Si(110) and SiO_2 were comparable to bulk values. Measured bending strength for Si and SiO_2 were twice as large as the values reported on larger-scale specimens. This indicates that bending strength shows a specimen-size dependence. Measured fracture toughness values for Si and SiO_2 were also comparable to values obtained on larger specimens. At stress amplitudes less than 15% of their bending strength and at mean stresses of less than 30% of the bending strength, Si and SiO_2 displayed an apparent endurance life of greater than 30 000 cycles. SEM observations of the fracture surfaces revealed a cleavage type of fracture for both materials when subjected to bending as well as fatigue.

The hardness, elastic modulus, and creep behavior of the PPMA, PMMA, PS and a PS/clay nanocomposite microbeams were evaluated using nanoindentation in continuous stiffness mode. The H, E, and creep resistance decreased in the following order: PS/clay, PMMA, PS, PPMA. The scratch resistance was observed to decrease in the following order: PMMA, PS/clay, PS, PPMA.

The elastic modulus of the double-anchored beams obtained from normal beam bending was lower than the E measured from nanoindentation due to stress concentration at the beam ends. The test environment affected the mechanical properties as well. After 36 hours of soaking in DI water, the hardness and elastic modulus of PPMA was affected, while the properties of the other three polymers were unaffected due to their low water absorption properties. Similarly, for tests conducted at human body temperature (37.5 °C), only PPMA exhibited a decrease in the hardness and modulus. The other materials did not exhibit any change because while this temperature is within the glass transition temperature range of PMMA, it is well below the T_g of the other polymers. By performing bending experiments at higher imposed loads, permanent deformation can be attained such that the yield and breaking strength can be evaluated. Adding nanoclay filler improved bending strength. It also modified the fracture behavior of the polymer microbeams. The unfilled polymers deformed in a ductile manner while the PS/clay nanocomposite exhibited minimal yielding followed by brittle failure.

Lateral bending has been demonstrated to provide a measure for the elastic, plastic and fracture properties of cantilever beams.

The AFM and nanoindenters used in this study can be satisfactorily used to evaluate the mechanical properties of micro/nanoscale structures for use in MEMS/NEMS.

FEM simulations were used to predict stress and deformation in nanostructures. FEM has been used to analyze the effect of the type of surface roughness and scratches on stresses and deformation of nanostructures. We found that roughness affects the maximum bending stresses. The maximum bending stresses increased as the asperity number increased for both semicircular and grooved asperities in the longitudinal direction. When the semicircular asperity was present in the transverse direction, the value of the maximum tensile stress was much larger than the maximum tensile stress value with no asperity or when the semicircular asperity was present in the longitudinal direction. This observation suggests that an asperity in the transverse direction is more detrimental. The presence of scratches increased the maximum tensile stress. The maximum tensile stress remained almost the same with the number of scratches for two types of loading, that is, when load was applied at the center of the beam or at the center of the scratch near the end, although the value of the tensile stress was much lower when the load was applied at the center of the scratch. This means that the load applied at the ends was less damaging. This analysis shows that FEM simulations can be useful to designers for development of the most suitable geometry for nanostructures.

32.A Appendix: Fabrication Procedure for the Double-Anchored and Cantilever Beams

The starting materials for the polymer microbeams are PPMA (molecular weight 250 000, Scientific Polymer Products), PMMA (molecular weight 75 000, Sigma-Aldrich), PS (melt flow index 4.0, Sigma-Aldrich) and PS/clay solutions in anisole (Acros Organics). The clay additive in PS/clay is Cloisite 20A surface modified natural montmorillonite (Southern Clay Products, Inc.) with a thickness of approximately 1 nm and lateral dimensions of 70–150 nm. The surface modification of the clay additive improves the dispersion of the nanoparticles in the polymer matrix, thus improving the properties of the composite. To prepare the nanocomposite, the clay was first dispersed in the PS matrix by melt compounding at a concentration of 10 wt% (clay/PS). The composite was then dissolved in anisole and sonicated for at least eight hours to dissolve the polymer and redisperse the particles.

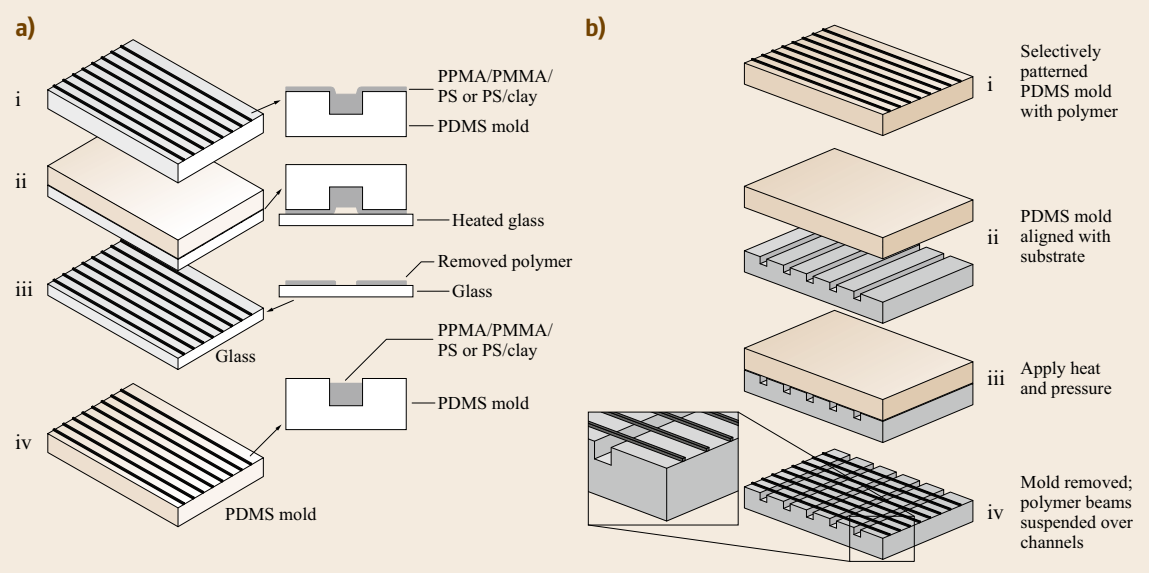


Fig. 32.32a,b Schematic of the micromolding process used in polymer double-anchored beam fabrication. **(a)** Selective patterning of PDMS mold: (i) mold is spin coated with a polymer layer; (ii) mold is inverted and brought into contact with a heated glass plate; (iii) mold is removed from the plate, transferring the surface polymer onto the glass; and (iv) mold is left with polymer only in the recessed features. **(b)** Stamping of the polymer beams: (i) selectively patterned mold; (ii) mold is inverted and aligned with the photolithographically patterned substrate; (iii) mold and substrate are brought into contact under heat and pressure; and (iv) polymer is transferred onto the photoresist and mold is removed (after [32.65])

Double-anchored polymer microbeams were fabricated using a soft lithography-based micromolding process along with standard photolithography [32.65]. The process involves selectively filling a poly(dimethyl siloxane) (PDMS) mold with the polymer of interest followed by transfer of the resulting structures to a prefabricated substrate. This substrate is a silicon wafer with a layer of SU-8 25 negative tone photoresist (MicroChem Corp.) patterned by photolithography to create 25 μm-wide channels, which is the resulting length of the suspended beams. (The second number 25 in the SU-8 designation relates to a viscosity appropriate for a given thickness.)

A patterned PDMS mold with the desired polymer beam geometry was fabricated from a photoresist master. Briefly, a layer of SU-8 5 photoresist was spin coated on silicon, and photolithography was used to define 5 μm-wide photoresist features separated by 45 μm gaps. A 10:1 ratio of T-2 PDMS translucent base and curing agent (Dow Corning) was mixed thoroughly and poured over the photoresist master to transfer the pattern into the PDMS. The mold was then placed in a vacuum desiccator to remove bubbles. The sample was removed from the vacuum periodically, and a razor blade was used to remove surface bubbles. After the bubbles were completely removed, the PDMS mold was allowed to cure at room temperature for 48 h before removing it from the wafer.

Next, the PDMS mold was selectively coated with the polymer solution to form the microbeams and then transferred to the substrate. Figure 32.32 is a schematic of the fabrication process used for making the polymer microbeams. As shown in Fig. 32.32a, the polymer solutions were spin coated on the PDMS mold for one minute. A 10% solution of polymer was spin coated on a mold with 5.3 μm-deep features at 3000 rpm for fabrication of the beams in the bending experiments. After spin coating, the mold was brought into contact with a heated glass plate to promote adhesion of the contacting polymer materials. This process removed the polymer material from the raised surfaces of the mold, resulting in the polymer remaining only in the recessed portions of the PDMS mold. The glass plate was heated to 175 °C for all four materials. As shown in Fig. 32.32b, the selectively coated mold was then aligned with the photolithographically patterned silicon substrate so that the beams of interest (PPMA, PMMA, PS, PS/clay) ran perpendicular to the channels defined in the photoresist. The substrate was then heated and pressure was applied to the top of the mold to transfer the material onto the substrate. The transfer temperatures for PPMA, PS, PS/clay and PMMA were 95, 125,

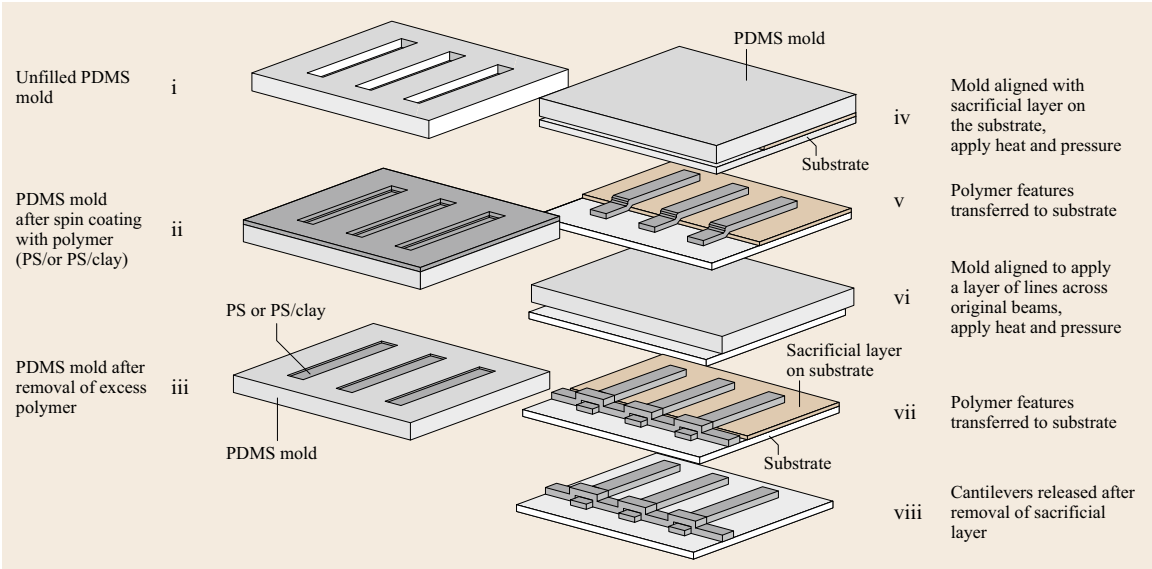


Fig. 32.33 Schematic of the polymer cantilever beam fabrication process: (i) PDMS mold; (ii) mold spin coated with the polymer of interest; (iii) selectively filled mold after removal of surface polymer; (iv) mold aligned with sacrificial layer; (v) polymer features transfer to the substrate using heat and pressure; (vi) mold aligned for a second time to add another layer of lines across original beams for reinforcement; (vii) polymer features transfer to the substrate using heat and pressure; and (viii) cantilever structures after removal of the sacrificial layer (after [32.65])

125, and 175 °C respectively, and the transfer pressure for all materials was around 0.21 MPa.

After removal of the mold, two types of polymer beams were transferred onto the wafer sample. The first type is supported beams, which was used to determine the hardness, elastic modulus, and creep response. The second type is suspended beams, on which the bending experiments were performed. The double-anchored beam samples studied by *Palacio* et al. [32.65] were 3–5 μm thick, 5 μm wide and nominally 25 μm long.

The process for fabrication of the polymer cantilever for lateral bending tests is shown in Fig. 32.33 [32.65]. A PDMS mold was first cast from a photolithographically patterned SU-8/silicon master. The resulting molds consisted of 50 μm-wide channels that were approximately 27 μm deep. The PDMS mold was then coated with the polymer of interest (PS or PS/clay). The polymers were spin coated on the mold at 3000 rpm for one minute at concentrations of 15 and 10% for PS and PS/clay respectively. The polymer on the raised surface of the mold was removed by contacting the surface with a glass slide heated to 180 °C. The mold was then inverted and manually aligned with a silicon substrate coated with patches of polyvinyl alcohol (PVA), which acts as the sacrificial layer. The PVA was patterned using photolithography and reactive ion etching with an oxygen plasma. The substrate was heated and pressure was applied to the top side of the mold to transfer the polymer from the recessed features of the mold onto the substrate such that a 350 μm portion of it is attached to the PVA and the remainder is attached to the bare silicon surface. Transfer temperatures for PS and PS/clay were 150 and 175 °C respectively. The process was repeated to apply another layer of beams across the length of the original layer. This was performed to provide reinforcement as it was observed that the one-layer design was not robust. The cantilever beam samples studied by *Palacio* et al. [32.65] were 12–27 μm thick, 60–80 μm wide and 350 μm long.

References

32.1 I. Fujimasa: *Micromachines: A New Era in Mechanical Engineering* (Oxford Univ. Press, Oxford 1996)

32.2 W.S. Trimmer (Ed.): *Micromachines and MEMS, Classic and Seminal Papers to 1990* (IEEE Press, New York 1997)

32.3 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

32.4 G.T.A. Kovacs: *Micromachined Transducers Sourcebook* (WCB McGraw-Hill, Boston 1998)

32.5 S.D. Senturia: *Microsystem Design* (Kluwer, Boston 2000)
32.6 M. Elwenspoek, R. Wiegerink: *Mechanical Microsensors* (Springer, Berlin 2001)
32.7 M. Gad-el-Hak: *The MEMS Handbook* (CRC, Boca Raton 2002)
32.8 T.R. Hsu: *MEMS and Microsystems: Design and Manufacture* (McGraw-Hill, Boston 2002)
32.9 A. Hierlemann: *Integrated Chemical Microsensor Systems in CMOS Technology* (Springer, Berlin 2005)
32.10 M. Madou: *Fundamentals of Microfabrication and Nanotechnology*, 3rd edn. (CRC, Boca Raton 2011)
32.11 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Switzerland 2016)
32.12 K.E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)
32.13 G. Timp (Ed.): *Nanotechnology* (Springer, New York 1999)
32.14 M.S. Dresselhaus, G. Dresselhaus, P. Avouris: *Carbon Nanotubes – Synthesis, Structure, Properties and Applications* (Springer, Berlin 2001)
32.15 E.A. Rietman: *Molecular Engineering of Nanosystems* (Springer, New York 2001)
32.16 W.A. Goddard, D.W. Brenner, S.E. Lyshevski, G.J. Iafrate (Eds.): *Handbook of Nanoscience, Engineering, and Technology* (CRC, Boca Raton 2002)
32.17 H.S. Nalwa (Ed.): *Nanostructures Materials and Nanotechnology* (Academic, San Diego 2002)
32.18 C.P. Poole, F.J. Owens: *Introduction to Nanotechnology* (Wiley, Hoboken 2003)
32.19 A. Manz, H. Becker (Eds.): *Microsystem Technology in Chemistry and Life Sciences*, Topics in Current Chemistry, Vol. 194 (Springer, Heidelberg 1998)
32.20 J. Cheng, L.J. Kricka (Eds.): *Biochip Technology* (Harwood Academic, Philadelphia 2001)
32.21 M.J. Heller, A. Guttman (Eds.): *Integrated Microfabricated Biodevices* (Marcel Dekker, New York 2001)
32.22 C.L.P. San, E.P.H. Yap (Eds.): *Frontiers in Human Genetics* (World Scientific, Singapore 2001)
32.23 C.H. Mastrangelo, H. Becker (Eds.): *Microfluidics and BioMEMS*, Proc. SPIE, Vol. 4560, 2001)
32.24 H. Becker, L.E. Lacascio: Polymer microfluidic devices, Talanta **56**, 267–287 (2002)
32.25 D.J. Beebe, G.A. Mensing, G.M. Walker: Physics and applications of microfluidics in biology, Annu. Rev. Biomed. Eng. **4**, 261–286 (2002)
32.26 A. van der Berg (Ed.): *Lab-on-a-Chip: Chemistry in Miniaturized Synthesis and Analysis Systems* (Elsevier, Amsterdam 2003)
32.27 J.V. Zoval, M.J. Madou: Centrifuge-based fluidic platforms, Proc. IEEE **92**, 140–153 (2000)
32.28 R. Raiteri, M. Grattarola, H. Butt, P. Skladal: Micromechanical cantilever-based biosensors, Sens. Actuators B **79**, 115–126 (2001)
32.29 W.C. Tang, A.P. Lee: Defense applications of MEMS, MRS Bulletin **26**, 318–319 (2001)
32.30 M.R. Taylor, P. Nguyen, J. Ching, K.E. Peterson: Simulation of microfluidic pumping in a genomic DNA blood-processing cassette, J. Micromech. Microeng. **13**, 201–208 (2003)
32.31 K. Park (Ed.): *Controlled Drug Delivery: Challenges and Strategies* (American Chemical Society, Washington 1997)
32.32 R.S. Shawgo, A.C.R. Grayson, Y. Li, M.J. Cima: BioMEMS for drug delivery, Curr. Opin. Solid State Mater. Sci. **6**, 329–334 (2002)
32.33 P.A. Öberg, T. Togawa, F.A. Spelman: *Sensors in Medicine and Health Care* (Wiley, New York 2004)
32.34 P. Decuzzi, S. Lee, B. Bhushan, M. Ferrari: A theoretical model for the margination of particles with blood vessels, Ann. Biomed. Eng. **33**, 179–190 (2005)
32.35 S.N. Bhatia, C.S. Chen: Tissue engineering at the micro-scale, Biomed. Microdevices **2**, 131–144 (1999)
32.36 R.P. Lanza, R. Langer, J. Vacanti (Eds.): *Principles of Tissue Engineering* (Academic, San Diego 2000)
32.37 E. Leclerc, K.S. Furukawa, F. Miyata, T. Sakai, T. Ushida, T. Fujii: Fabrication of microstructures in photosensitive biodegradable polymers for tissue engineering applications, Biomaterials **25**, 4683–4690 (2004)
32.38 T.H. Schulte, R.L. Bardell, B.H. Weigl: Microfluidic technologies in clinical diagnostics, Clin. Chim. Acta **321**, 1–10 (2002)
32.39 B. Bhushan: Macro- and microtribology of MEMS materials. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 1515–1548
32.40 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)
32.41 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)
32.42 S. Johansson, J.A. Schweitz, L. Tenerz, J. Tiren: Fracture testing of silicon microelements in-situ in a scanning electron microscope, J. Appl. Phys. **63**, 4799–4803 (1988)
32.43 F. Ericson, J.A. Schweitz: Micromechanical fracture strength of silicon, J. Appl. Phys. **68**, 5840–5844 (1990)
32.44 E. Obermeier: Mechanical and thermophysical properties of thin film materials for MEMS: Techniques and devices. In: *Micromech. Struct. Mater. Res. Symp. Proc.*, Vol. 444 (Materials Research Society, Pittsburgh 1996) pp. 39–57
32.45 C.J. Wilson, A. Ormeggi, M. Narbutovskih: Fracture testing of silicon microcantilever beams, J. Appl. Phys. **79**, 2386–2393 (1996)
32.46 W.N. Sharpe Jr., B. Yuan, R.L. Edwards: A new technique for measuring the mechanical properties of thin films, J. Microelectromech. Syst. **6**, 193–199 (1997)
32.47 K. Sato, T. Yoshioka, T. Anso, M. Shikida, T. Kawabata: Tensile testing of silicon film having different crystallographic orientations carried out on a silicon chip, Sens. Actuators A **70**, 148–152 (1998)
32.48 S. Greek, F. Ericson, S. Johansson, M. Fürtsch, A. Rump: Mechanical characterization of thick polysilicon films: Young's modulus and fracture strength evaluated with microstructures, J. Micromech. Microeng. **9**, 245–251 (1999)

32.49 D.A. LaVan, T.E. Buchheit: Strength of polysilicon for MEMS devices, Proc. SPIE **3880**, 40–44 (1999)

32.50 E. Mazza, J. Dual: Mechanical behavior of a μm-sized single crystal silicon structure with sharp notches, J. Mech. Phys. Solids **47**, 1795–1821 (1999)

32.51 T. Yi, C.J. Kim: Measurement of mechanical properties for MEMS materials, Meas. Sci. Technol. **10**, 706–716 (1999)

32.52 H. Kahn, M.A. Huff, A.H. Heuer: Heating effects on the Young's modulus of films sputtered onto micromachined resonators. In: *Microelectromech. Struct. Mater. Res. Symp. Proc., Vol. 518* (Materials Research Society, Pittsburgh 1998) pp. 33–38

32.53 S. Johansson, F. Ericson, J.A. Schweitz: Influence of surface-coatings on elasticity, residual-stresses, and fracture properties of silicon microelements, J. Appl. Phys. **65**, 122–128 (1989)

32.54 R. Ballarini, R.L. Mullen, Y. Yin, H. Kahn, S. Stemmer, A.H. Heuer: The fracture toughness of polysilicon microdevices: A first report, J. Mater. Res. **12**, 915–922 (1997)

32.55 H. Kahn, R. Ballarini, R.L. Mullen, A.H. Heuer: Electrostatically actuated failure of microfabricated polysilicon fracture mechanics specimens, Proc. R. Soc. Lond. Ser. A **455**, 3807–3823 (1999)

32.56 A.M. Fitzgerald, R.H. Dauskardt, T.W. Kenny: Fracture toughness and crack growth phenomena of plasma-etched single crystal silicon, Sens. Actuators A **83**, 194–199 (2000)

32.57 T. Tsuchiya, A. Inoue, J. Sakata: Tensile testing of insulating thin films: Humidity effect on tensile strength of SiO_2 films, Sens. Actuators A **82**, 286–290 (2000)

32.58 J.A. Connally, S.B. Brown: Micromechanical fatigue testing, Exp. Mech. **33**, 81–90 (1993)

32.59 K. Komai, K. Minoshima, S. Inoue: Fracture and fatigue behavior of single-crystal silicon microelements and nanoscopic AFM damage evaluation, Microsyst. Technol. **5**, 30–37 (1998)

32.60 T. Namazu, Y. Isono, T. Tanaka: Evaluation of size effect on mechanical properties of single-crystal silicon by nanoscale bending test using AFM, J. Microelectromech. Syst. **9**, 450–459 (2000)

32.61 S. Sundararajan, B. Bhushan: Development of AFM-based techniques to measure mechanical properties of nanoscale structures, Sens. Actuators A **101**, 338–351 (2002)

32.62 X. Li, B. Bhushan: Fatigue studies of nanoscale structures for MEMS/NEMS applications using nanoindentation techniques, Surf. Coat. Technol. **163/164**, 521–526 (2003)

32.63 X. Li, B. Bhushan, K. Takashima, C.W. Baek, Y.K. Kim: Mechanical characterization of micro-/nanoscale structures for MEMS/NEMS applications using nanoindentation techniques, Ultramicroscopy **97**, 481–494 (2003)

32.64 G. Wei, B. Bhushan, N. Ferrell, D. Hansford: Microfabrication and nanomechanical characterization of polymer MEMS for biological applications, J. Vac. Sci. Technol. A **23**, 811–819 (2005)

32.65 M. Palacio, B. Bhushan, N. Ferrell, D. Hansford: Nanomechanical characterization of polymer beam structures for bioMEMS applications, Sens. Actuators A **135**, 637–650 (2007)

32.66 T. Hsu, N. Sun: Residual stresses/strains analysis of MEMS. In: *Proc. Int. Conf. Model. Simul. Microsyst. Semicond. Sens. Actuators*, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) pp. 82–87

32.67 A. Kolpekwar, C. Kellen, R.D. Blanton: Fault model generation for MEMS. In: *Proc. Int. Conf. Model. Simul. Microsyst. Semicond. Sens. Actuators*, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) pp. 111–116

32.68 H.A. Rueda, M.E. Law: Modeling of strain in boron-doped silicon cantilevers. In: *Proc. Int. Conf. Model. Simul. Microsyst. Semicond. Sens. Actuators*, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) pp. 94–99

32.69 M. Heinzelmann, M. Petzold: FEM analysis of microbeam bending experiments using ultra-micro indentation, Comput. Mater. Sci. **3**, 169–176 (1994)

32.70 C.J. Wilson, P.A. Beck: Fracture testing of bulk silicon microcantilever beams subjected to a side load, J. Microelectromech. Syst. **5**, 142–150 (1996)

32.71 B. Bhushan, G.B. Agrawal: Stress analysis of nanostructures using a finite element method, Nanotechnology **13**, 515–523 (2002)

32.72 B. Bhushan, G.B. Agrawal: Finite element analysis of nanostructures with roughness and scratches, Ultramicroscopy **97**, 495–507 (2003)

32.73 K.E. Petersen: Silicon as a mechanical material, Proc. IEEE **70**, 420–457 (1982)

32.74 B. Bhushan, S. Sundararajan, X. Li, C.A. Zorman, M. Mehregany: Micro-/nanotribological studies of single-crystal silicon and polysilicon and SiC films for use in MEMS devices. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 407–430

32.75 S. Sundararajan, B. Bhushan: Micro-/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear **217**, 251–261 (1998)

32.76 X. Li, B. Bhushan: Micro-/nanomechanical characterization of ceramic films for microdevices, Thin Solid Films **340**, 210–217 (1999)

32.77 H. Becker, C. Gärtner: Polymer microfabrication methods for microfluidic analytical applications, Electrophoresis **21**, 12–26 (2000)

32.78 J.C. McDonald, D.C. Duffy, J.R. Anderson, D.T. Chiu, H. Wu, O.J.A. Schueller, G.M. Whitesides: Fabrication of microfluidic systems in poly(dimethylsiloxane), Electrophoresis **21**, 27–40 (2000)

32.79 M. Palacio, B. Bhushan, N. Ferrell, D. Hansford: Adhesion properties of polymer/silicon interfaces for biological micro-/nanoelectromechanical applications, J. Vac. Sci. Technol. A **25**, 1275–1284 (2007)

32.80 B. Ellis: *Polymers: A Property Database* (CRC, Boca Raton 2000), http://www.polymersdatabase.com/

32.81 J. Brandrup, E.H. Immergut, E.A. Grulke (Eds.): *Polymer Handbook*, 4th edn. (Wiley, New York

1999)
32.82 J.E. Mark: *Polymers Data Handbook* (Oxford Univ. Press, New York 1999)
32.83 B. Bhushan, X. Li: Nanomechanical characterization of solid surfaces and thin films, Int. Mater. Rev. **48**, 125–164 (2003)
32.84 B.R. Lawn, A.G. Evans, D.B. Marshall: Elastic/plastic indentation damage in ceramics: The median/radial system, J. Am. Ceram. Soc. **63**, 574 (1980)
32.85 S. Sundararajan, B. Bhushan, T. Namazu, Y. Isono: Mechanical property measurements of nanoscale structures using an atomic force microscope, Ultramicroscopy **91**, 111–118 (2002)
32.86 W.C. Young, R.G. Budynas: *Roark's Formulas for Stress and Strain* (McGraw-Hill, New York 2002)
32.87 R.W. Hertzberg: *Deformation and Fracture Mechanics of Engineering Materials*, 3rd edn. (Wiley, New York 1989) pp. 277–278
32.88 D.E. Aspnes: *Properties of Silicon*, EMIS Datarev., Vol. 4 (INSPEC Institution of Electrical Engineers, London 1988)
32.89 C.T.-C. Nguyen, R.T. Howe: An integrated CMOS micromechanical resonator high-Q oscillator, IEEE J. Solid-State Circuits **34**, 440–455 (1999)
32.90 L.J. Hornbeck: A digital light processing update – status and future applications. In: *Proc. Soc. Photo-Opt. Eng., Projection Displ. V*, Vol. 3634 (1999) pp. 158–170
32.91 M. Tanaka: Fracture toughness and crack morphology in indentation fracture of brittle materials, J. Mater. Sci. **31**, 749 (1996)
32.92 B. Bhushan, S. Venkatesan: Mechanical and tribological properties of silicon for micromechanical applications: A review, Adv. Inf. Storage Syst. **5**, 211–239 (1993)
32.93 B. Bhushan, B.K. Gupta: *Handbook of Tribology: Materials, Coatings, and Surface Treatments* (McGraw-Hill, New York 1991), reprint with corrections (Krieger, Malabar 1997)
32.94 T. Tsuchiya, O. Tabata, J. Sakata, Y. Taga: Specimen size effect on tensile strength of surface-micromachined polycrystalline silicon thin films, J. Microelectromech. Syst. **7**, 106–113 (1998)
32.95 T. Yi, L. Li, C.J. Kim: Microscale material testing of single crystalline silicon: Process effects on surface morphology and tensile strength, Sens. Actuators A **83**, 172–178 (2000)
32.96 I.H. Loh, M.S. Sheu, A.B. Fischer: Biocompatible polymer surfaces. In: *Desk Reference of Functional Polymers: Syntheses and Applications*, ed. by R. Arshady (American Chemical Society, Washington 1997)
32.97 D.B. Holt, P.R. Gauger, A.W. Kusterbech, F.S. Ligler: Fabrication of a capillary immunosensor in polymethyl methacrylate, Biosens. Bioelectron. **17**, 95–103 (2002)
32.98 F.W.J. Billmeyer: *Textbook of Polymer Science* (Wiley, New York 1984)
32.99 Anonymous: Rohm and Haas General Information on PMMA, Philadelphia, Pennsylvania
32.100 T.G. van Kooten, H.T. Spijker, H.H. Busscher: Plasma-treated polystyrene surfaces: Model surface for studying cell-biomaterial interactions, Biomaterials **25**, 1735–1747 (2004)
32.101 M. Alexandre, P. Dubois: Polymer-layered silicate nanocomposites: Preparation, properties and uses of a new class of materials, Mater. Sci. Eng. **28**, 1–63 (2000)
32.102 S.S. Ray, M. Okamoto: Polymer/layered silicate nanocomposites: A review from preparation to processing, Prog. Polym. Sci. **28**, 1539–1641 (2003)
32.103 R.H. Boundy, R.F. Boyer (Eds.): *Styrene, Its Polymers, Copolymers and Derivatives* (Reinhold, New York 1952)
32.104 C.A. Harper (Ed.): *Modern Plastics Encyclopedia* (McGraw-Hill, New York 2000)
32.105 S.P. Timoshenko, J.N. Goodier: *Theory of Elasticity*, 3rd edn. (McGraw-Hill, New York 1970)
32.106 J.E. Shigley, L.D. Mitchell: *Mechanical Engineering Design*, 4th edn. (McGraw-Hill, New York 1993)

Part G

Molecula

Part G Molecularly-Thick Films for Lubrication

33 Nanotribology of Ultrathin and Hard Amorphous Carbon Films
Bharat Bhushan, Columbus, USA

34 Self-Assembled Monolayers for Nanotribology and Surface Protection
Bharat Bhushan, Columbus, USA

35 Nanoscale Boundary Lubrication Studies
Bharat Bhushan, Columbus, USA

33. Nanotribology of Ultrathin and Hard Amorphous Carbon Films

Bharat Bhushan

One of the best materials to use in applications that require very low wear and reduced friction is diamond, especially in the form of a diamond coating. Unfortunately, true diamond coatings can be deposited only at high temperatures and on selected substrates, and they require surface finishing. However, the mechanical, thermal, and optical properties of hard, amorphous carbon – commonly known as a diamond-like carbon (DLC) coating – can be tailored to be similar to those of diamond. It can also be deposited at a wide range of thicknesses using a variety of deposition processes on various substrates at or near room temperature. The coatings reproduce the topography of the substrate, removing the need for surface finishing. Various physical and chemical characterization techniques have been used to study the structure, stoichiometry, and mechanical and tribological properties. The friction and wear properties of some DLC coatings make them very attractive for some tribological applications. The most significant current industrial applications of DLC coatings are in magnetic storage devices, micro/nanoelectromechanical systems, and various other consumer products including razor blades, sunglasses, and optical windows.

In this chapter, the state-of-the-art in the chemical, mechanical, and tribological characterization of ultrathin amorphous carbon coatings as thin as 5 nm is presented.

33.1 **Description of Commonly Used Deposition Techniques** 1145
33.1.1 Filtered Cathodic Arc Deposition Technique 1145
33.1.2 Ion-Beam Deposition Technique 1146
33.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique 1148
33.1.4 Sputtering Deposition Technique 1148
33.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique 1148

33.2 **Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties** 1149
33.2.1 EELS and Raman Spectroscopy 1149
33.2.2 Hydrogen Concentrations 1150
33.2.3 Physical Properties 1153
33.2.4 Summary 1154

33.3 **Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques** 1154
33.3.1 Micromechanical Characterization 1154
33.3.2 Microscratch and Microwear Studies 1161
33.3.3 Macroscale Tribological Characterization 1168
33.3.4 Coating Continuity Analysis 1173

33.4 **Conclusion** 1174

References 1175

Carbon exists in both crystalline and amorphous forms and exhibits both metallic and nonmetallic characteristics [33.1–3]. Crystalline carbon includes graphite, diamond, and a family of fullerenes (Fig. 33.1). Graphite and diamond are infinite periodic network solids with a planar structure, whereas the fullerenes are a molecular form of pure carbon with a finite network and a nonplanar structure. Graphite has a hexagonal, layered structure with weak interlayer bonding forces, and exhibits excellent lubrication properties. The graphite crystal may be visualized as infinite parallel layers of hexagons stacked 0.34 nm apart with a 0.1415 nm interatomic distance between the carbon atoms in the basal plane. The atoms lying in the basal planes are trigonally coordinated and closely packed with strong σ (covalent) bonds to its three carbon neighbors using the hybrid sp^2 orbitals. The fourth electron lies in a p_z orbital lying normal to the σ bonding plane. It forms a weak π bond by overlapping side-to-side with a p_z orbital of an adjacent atom to which carbon is attached by a σ bond. The layers (basal planes) themselves are relatively far apart and the forces that bond them are weak van der

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_33

Waals forces. These layers can align themselves parallel to the direction of the relative motion, and slide over one another with relative ease, providing low friction. Strong interatomic bonding and packing in each layer is thought to help reduce wear. Operating environment has a significant influence on the lubrication properties, i. e., low friction and low wear, of graphite. It lubricates better in a humid environment than a dry one which results from adsorption of water vapor and other gases from the environment. This further weakens the interlayer bonding forces, resulting in easy shear and transfer of the crystallite platelets to the mating surface. Thus, transfer plays an important role in controlling friction and wear. Graphite oxidizes at high operating temperatures, and can be used up to about 430 °C.

One of the fullerene molecules is C_{60}, commonly known as a Buckyball. Since the C_{60} molecules are very stable and do not require additional atoms to satisfy chemical bonding requirements, they are expected to have low adhesion to the mating surface and low surface energy. Since C_{60} molecules with a perfect spherical symmetry are weakly bonded to other molecules, C_{60} clusters get detached readily, similar to other layered-lattice structures. They either transfer to the mating surface by mechanical compaction, or are present as a loose wear particle that may roll like a tiny ball bearing in a sliding contact, resulting in low friction and wear. The wear particles are expected to be harder than as-deposited C_{60} molecules because of their phase transformation at the high asperity contact pressures present in a sliding interface. The low surface energy, spherical shape of C_{60} molecules, weak intermolecular bonding, and high load-bearing capacity offer potential for various mechanical and tribological applications. Sublimed C_{60} coatings and fullerene particles as an additive to mineral oils and greases, have been reported to be good solid lubricants comparable to graphite and MoS_2 [33.4–6].

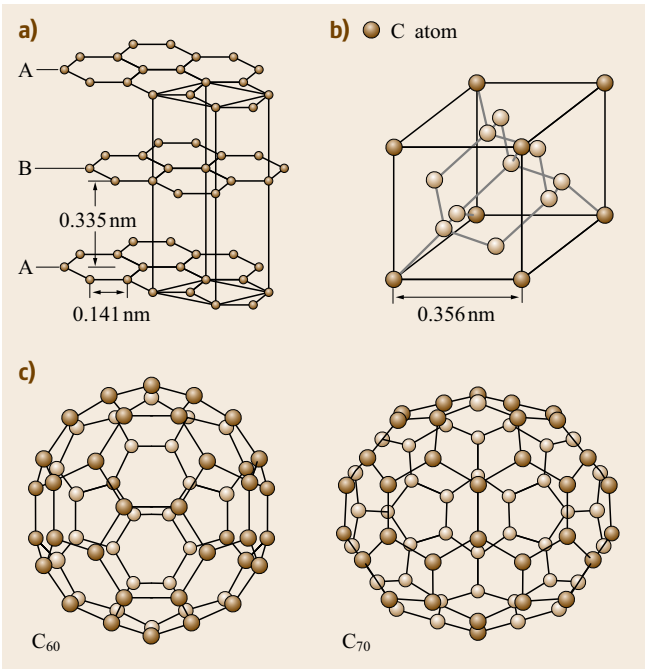


Fig. 33.1a–c The structure of three known forms of crystalline carbon. **(a)** Hexagonal structure of graphite. **(b)** Modified face-centered cubic (fcc) structure (two interpenetrating fcc lattices displaced by one one-quarter of the cube diagonal) of diamond (each atom is bonded to four others that form the corners of the pyramidal structure called a tetrahedron). **(c)** The structure of the two most common forms of fullerenes – soccer ball C_{60} and rugby ball C_{70} molecules

Diamond crystallizes in a modified face-centered cubic (fcc) structure with an interatomic distance of 0.154 nm. The diamond cubic lattice consists of two interpenetrating fcc lattices displaced by one-quarter of the cube diagonal. Each carbon atom is tetrahedrally coordinated, making strong σ (covalent) bonds to its four carbon neighbors using the hybrid sp^3 atomic orbitals. This accounts for its hardness (80–104 GPa) and thermal conductivity (900–2100 W/mK, on the order of five times that of copper), the highest of any known solid, and a high electrical resistivity, optical transmission, and large optical bandgap. It is relatively chemically inert, and it exhibits poor adhesion with other solids with consequent low friction and wear. Its high thermal conductivity allows dissipation of frictional heat during sliding and protects the interface. The dangling carbon bonds on the surface react with the environment to form hydrocarbons, which act as good lubrication films. These are some of the reasons for the low friction and wear of diamond. Diamond and its coatings find many industrial applications: tribological applications (low friction and wear), optical applications (exceptional optical transmission, high abrasion resistance), and thermal management or heat sink applications (high thermal conductivity). Diamond can be used to high temperatures and its starts to become graphite at about 1000 °C in ambient air and at about 1400 °C in vacuum. Diamond is an attractive material for cutting tools, as an abrasive for grinding wheels and lapping compounds, and other extreme-wear applications.

Natural diamond, particularly in large sizes, is very expensive and its coatings, a low-cost alternative, are attractive. True diamond coatings are deposited by chemical vapor deposition (CVD) processes at high substrate temperatures (on the order of 800 °C). They adhere best on a silicon substrate and require an interlayer for other substrates. A major roadblock to the widespread use of

true diamond films in tribological, optical, and thermal management applications is surface roughness. Growth of the diamond phase on a nondiamond substrate is initiated by nucleation either at randomly seeded sites or at thermally favored sites due to statistical thermal fluctuation at the substrate surface. On the basis of growth temperature and pressure conditions, favored crystal orientations dominate the competitive growth process. As a result, the grown films are polycrystalline in nature with relatively large grain size ($> 1\,\mu m$), and terminate in very rough surfaces with RMS (root mean square) roughnesses ranging from a few tenths of a micron to tens of microns. Techniques for polishing these films have been developed. It has been reported that the laser-polished films exhibit friction and wear properties almost comparable to that of bulk polished diamond [33.7, 8].

Amorphous carbon has no long-range order and the short-range order of carbon atoms can have one or more of three bonding configurations – sp^3 (diamond), sp^2 (graphite), or sp^1 (with two electrons forming strong σ bonds and the remaining two electrons left in orthogonal p_y and p_z orbitals to form weak π bonds). Short-range order controls the properties of amorphous materials and coatings. Hard amorphous carbon (a-C) coatings are commonly known as diamond-like carbon or DLC (implying high hardness) coatings. They are a class of coatings that are mostly metastable amorphous materials, but include a micro- or nanocrystalline phase. The coatings are random networks of covalently bonded carbon in hybridized tetragonal (sp^3) and trigonal (sp^2) local coordination, with some of the bonds terminated by hydrogen. These coatings have been deposited successfully by a variety of vacuum deposition techniques on a variety of substrates at or near room temperature. The coatings generally reproduce substrate topography and do not require any postfinishing. However, these coatings mostly adhere best on silicon substrates. Best adhesion is obtained on substrates that form carbides, e.g., Si, Fe, and Ti. On the basis of depth profile analyses using Auger and XPS (x-ray photoelectron spectroscopy) of DLC coatings deposited on silicon substrates, it has been reported that a substantial amount of silicon carbide (on the order of 5–10 nm in thickness) is present at the carbon-silicon interface for the coatings with good adhesion and high hardness (e.g., [33.9]). For good adhesion of DLC coatings to other substrates, in most cases, an interlayer of silicon is required, except for cathodic arc-deposited coatings.

There is significant interest in DLC coatings because of their unique combination of desirable properties. These properties include high hardness and wear resistance, chemical inertness to both acids and alkalis, lack of magnetic response, and an optical band gap ranging from zero to a few eV, depending upon the deposition conditions. These are used in a wide range of applications including tribological, optical, electronic,

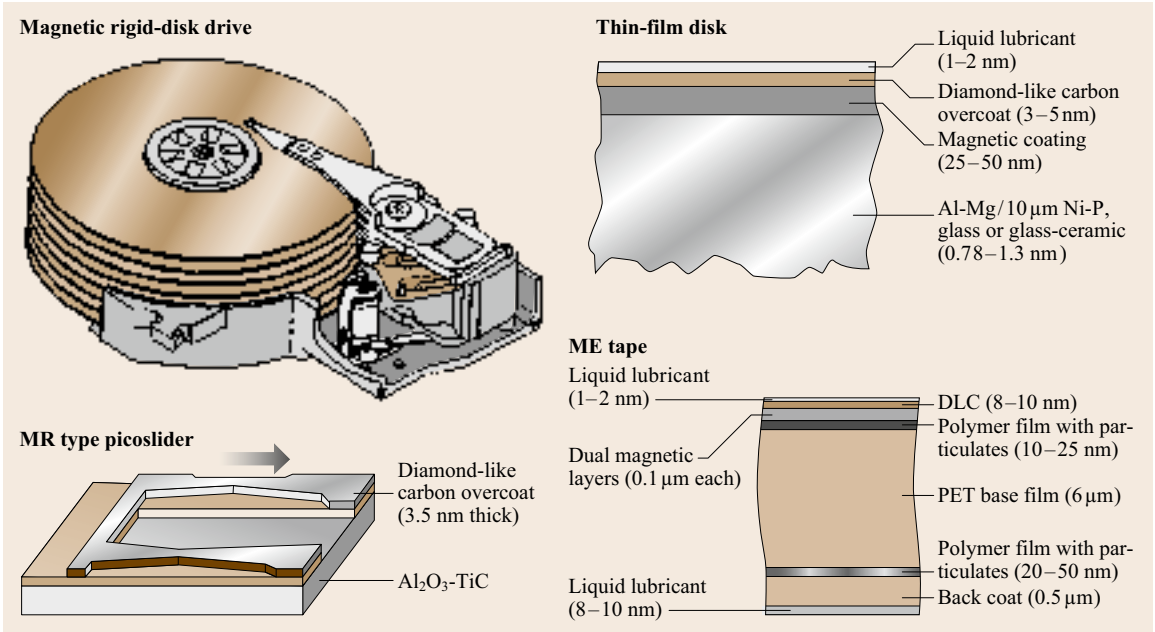


Fig. 33.2 Schematic of a magnetic rigid disk drive and MR-type picoslider and cross-sectional schematics of a magnetic thin-film rigid disk and a metal evaporated (ME) tape

and biomedical applications [33.1, 10, 11]. The high hardness, good friction and wear properties, versatility in deposition and substrates, and no requirements for postfinishing make them very attractive for tribological applications. The largest industrial application of the family of amorphous carbon coatings are typically deposited by DC/RF (direct current/radio frequency) magnetron sputtering (SP), plasma-enhanced chemical vapor deposition, or ion-beam (IB) deposition techniques in magnetic storage devices [33.12–20]. These coatings are employed to protect against wear and corrosion, in magnetic coatings on thin-film rigid disks and metal evaporated (ME) tapes, and in the thin-film head structure of a read/write disk head (Fig. 33.2). To maintain low physical spacing between the magnetic element of a magneto resistive-type (MR) read/write head and the magnetic layer of a media, thicknesses ranging from 3 to 10 nm are employed. Mechanical properties affect friction and wear, and these need to be optimized. In 1998, Gillette introduced Mach 3 razor blades with ultrathin DLC coatings, which is a large industrial application. DLC coatings are also used in other commercial applications such as cutting tools, glass windows of supermarket laser barcode scanners, other optical windows, and sunglasses. These coatings are also being used in micro- and nanoelec-

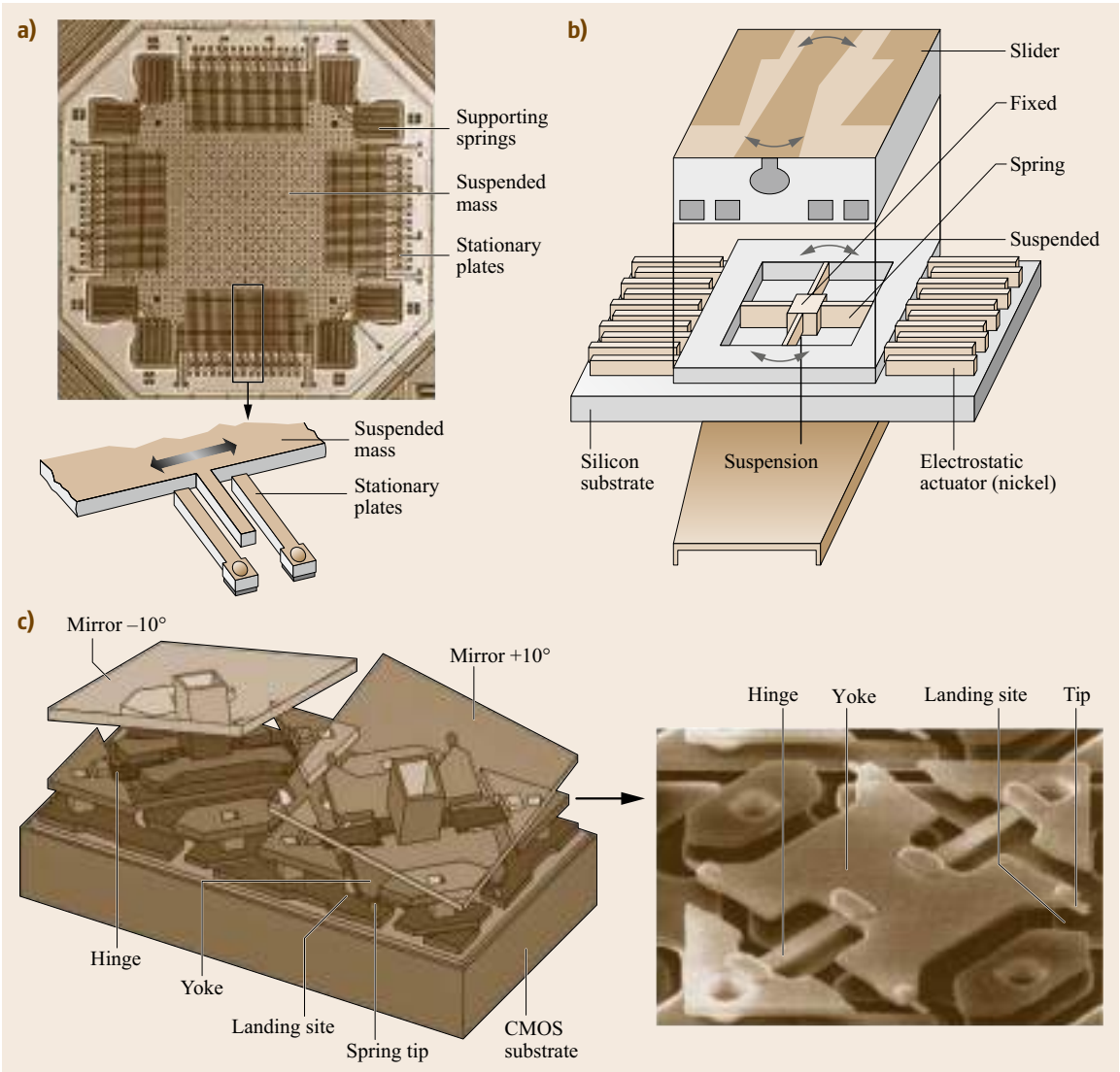


Fig. 33.3a–c Schematics of **(a)** a capacitive-type silicon accelerometer for automotive sensory applications, **(b)** digital micrometer devices for high-projection displays, and **(c)** a polysilicon rotary microactuator for magnetic disk drives

tromechanical systems (MEMS/NEMS) components (Fig. 33.3) [33.21–25].

In this chapter, an overview of the chemical, mechanical, and tribological characterization of ultrathin amorphous carbon coatings, commonly referred to as diamond-like carbon, is presented. First, an overview of the most commonly used deposition techniques is presented, followed by typical chemical and mechanical characterization data and typical tribological data both from coupon-level testing and functional testing.

33.1 Description of Commonly Used Deposition Techniques

The first hard amorphous carbon coatings were deposited by a beam of carbon ions produced in an argon plasma on room-temperature substrates as reported by *Aisenberg* and *Chabot* [33.26]. Subsequent confirmation by *Spencer* et al. [33.27] led to explosive growth of this field. Following on from the first work, several alternative techniques have been developed. The amorphous carbon coatings have been prepared by a variety of deposition techniques and precursors, including evaporation, DC, RF or ion beam sputtering, RF or DC plasma-enhanced chemical vapor deposition (PECVD), electron cyclotron resonance chemical vapor deposition (ECR-CVD), direct ion beam deposition, pulsed laser vaporization and vacuum arc, from a variety of carbon-bearing solids or gaseous source materials [33.1, 28]. Coatings with both graphitic and diamond-like properties have been produced. Evaporation and ion-plating techniques have been used to produce coatings with graphitic properties (low hardness, high electrical conductivity, very low friction, etc.) and all techniques have been used to produce coatings with diamond-like properties.

Structure and properties of a coating are dependent upon the deposition technique and the deposition parameters. High-energy surface bombardment has been used to produce harder and denser coatings. Decreasing sp^3/sp^2 fractions in order have been reported for cathodic arc deposition, pulsed laser vaporization, direct ion beam deposition, plasma-enhanced chemical vapor deposition, ion beam sputtering, and DC/RF sputtering [33.12, 29, 30]. A common feature to these techniques is that the deposition is energetic, i. e., carbon species arrive with an energy significantly greater than that represented by the substrate temperature. The resultant coatings are amorphous in structure, with hydrogen content up to 50%, and display a high degree of sp^3 character. From the results of previous investigations, it has been proposed that deposition of sp^3-bonded carbon requires the depositing species to have kinetic energies on the order of 100 eV or higher, well above those obtained in thermal processes like evaporation (0–0.1 eV). The species must then be quenched into the metastable configuration via rapid energy removal. Excess energy, such as that provided by substrate heating, is detrimental to the achievement of a high sp^3 fraction. In general, a high fraction of the sp^3-bonded carbon atoms in an amorphous network results in a higher hardness [33.30–37]. The mechanical and tribological properties of a carbon coating depend on the sp^3/sp^2-bonded carbon ratio, the amount of hydrogen in the coating, and adhesion of the coating to the substrate, which are influenced by the precursor material, kinetic energy of the carbon species prior to deposition, deposition rate, substrate temperature, substrate biasing, and the substrate itself [33.30, 34, 36, 38–47]. The kinetic energies and deposition rates involved in selected deposition processes used in the deposition of DLC coatings are compared in Table 33.1 [33.1, 29].

In the studies by *Gupta* and *Bhushan* [33.12, 48], *Li* and *Bhushan* [33.49, 50] and *Sundararajan* and *Bhushan* [33.51], DLC coatings, ranging in thickness typically from 3.5 to 20 nm, were deposited on single-crystal silicon, magnetic Ni-Zn ferrite, and Al_2O_3-TiC substrates (surface roughness $\approx 1-3$ nm RMS) by filtered cathodic arc (FCA) deposition, (direct) ion beam deposition (IBD), electron cyclotron resonance chemical vapor deposition (ECR-CVD), plasma-enhanced chemical vapor deposition (PECVD), and DC/RF planar, magnetron sputtering (SP) deposition techniques [33.52]. Herein, we will limit the presentation of data to coatings deposited by the FCA, IBD, ECR-CVD, and SP deposition techniques.

33.1.1 Filtered Cathodic Arc Deposition Technique

In the filtered cathodic arc deposition of carbon coatings [33.30, 53–60], a vacuum arc plasma source is used to form a carbon film. In the FCA technique used by Bhushan and coworkers (e.g., [33.12]), energetic carbon ions are produced by a vacuum arc discharge between a planar graphite cathode and a grounded anode (Fig. 33.4a). The cathode is a 6 mm diameter high-density graphite disk mounted on a water-cooled copper block. The arc is driven at an arc current of 200 A, arc duration of 5 ms, and arc repetition rate of 1 Hz. The plasma beam is guided by a magnetic field that transports current between the electrodes to form

Table 33.1 Summary of most commonly used deposition techniques and the kinetic energy of depositing species and deposition rates

Deposition technique	Process	Kinetic energy (eV)	Deposition rate (nm/s)
Filtered cathodic arc (FCA)	Energetic carbon ions produced by a vacuum arc discharge between a graphite cathode and a grounded anode	100–2500	0.1–1
Direct ion beam (IB)	Carbon ions produced from methane gas in an ion source and accelerated toward a substrate	50–500	0.1–1
Plasma-enhanced chemical vapor deposition (PECVD)	Hydrocarbon species produced by plasma decomposition of a hydrocarbon gas (such as acetylene) are accelerated toward a DC-biased substrate	1–30	1–10
Electron cyclotron resonance plasma chemical vapor deposition (ECR-CVD)	Hydrocarbon ions produced by the plasma decomposition of ethylene gas in the presence of a plasma at the electron cyclotron resonance condition are accelerated toward a RF-biased substrate	1–50	1–10
DC/RF sputtering	Sputtering of graphite target by argon ion plasma	1–10	1–10

tiny, rapidly moving spots on the cathode surface. The source is coupled to a 90° bent magnetic filter to remove the macroparticles produced concurrently with the plasma in the cathode spots. The ion current density at the substrate is in the range of 10–50 mA/cm^2. The base pressure is less than 10^{-4} Pa. Compared to electron beam evaporation with auxiliary discharge, much higher plasma density is achieved with the aid of powerful arc discharge. In this process, the cathodic material suffers a complicated transition from the solid phase to an expanding, nonequilibrium plasma via liquid and dense, equilibrium nonideal plasma phases [33.59]. The carbon ions in the vacuum arc plasma have a direct kinetic energy of 20–30 eV. The high voltage pulses are applied to the substrate mounted on a water-cooled sample holder; the ions are accelerated through a sheath and arrive at the substrate with an additional energy given by the potential difference between the plasma and the substrate. The substrate holder is pulse biased to a negative voltage up to −2 kV with a pulse duration of 1 μs. The negative biasing of −2 kV corresponds to 2 keV kinetic energy of the carbon ions. The use of a pulse bias instead of a DC bias has the advantages of applying a much higher voltage and building a surface potential on a nonconducting film. The energy of the ions is varied during the deposition. For the first 10% of the deposition the substrates are pulsed biased to −2 keV with a pulse duty cycle of 25%, i. e., for 25% of the time the energy is 2 keV, for the remaining 75% of the time it is 20 eV which is the *natural* energy of carbon ions in a vacuum discharge. For the last 90% of the deposition the pulsed bias voltage is reduced to −200 eV with a pulse bias duty cycle of 25%, i. e., the energy is 200 eV for 25% and 20 eV for 75% of the deposition. The high energy at the beginning leads to a good intermixing and good adhesion of the films whereas the lower energy at the later stage leads to hard films. Under the conditions described, the deposition rate at the substrate is about 0.1 nm/s which is slow. Compared with most gaseous plasmas, the cathodic arc plasma is almost fully ionized and as the ionized carbon atoms have high kinetic energy they help in achieving a high fraction of sp^3-bonded carbon ions which in turn results in a high hardness and higher interfacial adhesion. *Cuomo* et al. [33.43] have reported that based on electron energy loss spectroscopy (EELS) analysis, the sp^3-bonded carbon fraction of cathodic arc coating is 83% as compared with 38% for the ion-beam sputtered carbon. These coatings are reported to be *nonhydrogenated*.

This technique does not require an adhesion underlayer for nonsilicon substrates. However, adhesion of the DLC coatings on the electrically insulating substrate is poor as negative pulse biasing forms an electrical sheath which accelerates depositing ions to the substrate and enhances the adhesion of the coating to the substrate with associated ion implantation. It is difficult to build potential on an insulating substrate and lack of biasing results in poor adhesion.

33.1.2 Ion-Beam Deposition Technique

In direct ion beam deposition of a carbon coating [33.61–65], as used by Bhushan and coworkers (e.g., [33.12]) (deposition carried out by Ion Tech. Inc., Fort Collins, Colorado), the carbon coating is deposited from an accelerated carbon ion beam. The sample is precleaned by ion etching. For the case of nonsilicon substrates, a 2–3 nm thick amorphous silicon adhesion layer is deposited by ion beam sputtering using an ion beam of a mixture of methane and argon at 200 V. For the carbon deposition, the chamber is pumped to about 10^{-4} Pa and methane gas is fed through the cylindrical ion source and is ionized by energetic electrons produced by a hot-wire filament (Fig. 33.4b). Ionized species then pass through a grid with a bias voltage of

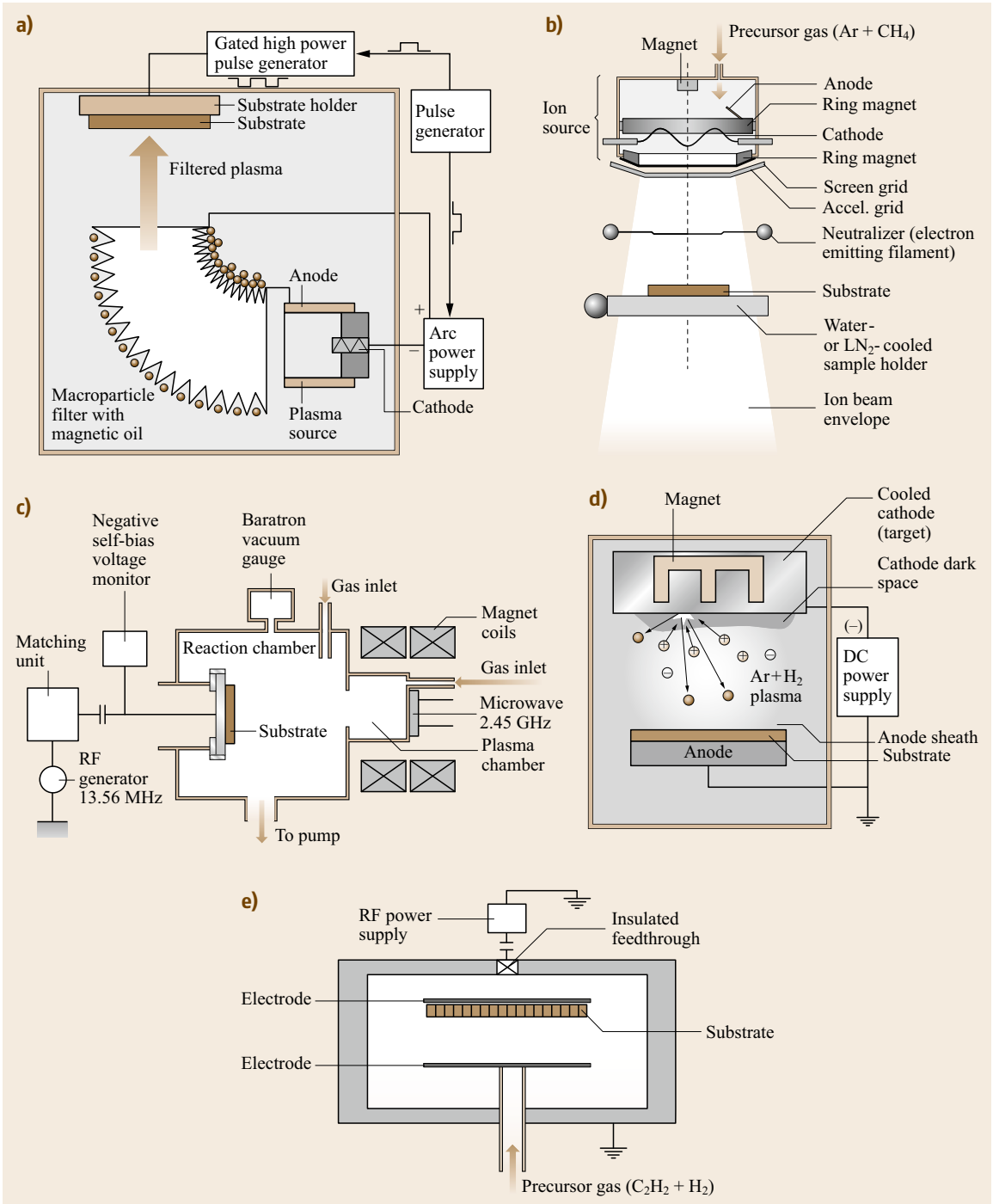


Fig. 33.4a–e Schematic diagrams of deposition by **(a)** filtered cathodic arc deposition, **(b)** ion beam deposition, **(c)** electron cyclotron resonance chemical vapor deposition (ECR-CVD), **(d)** DC planar magnetron sputtering, and **(e)** plasma-enhanced chemical vapor deposition (PECVD)

about 50 eV where they gain a high acceleration energy and reach a hot-wire filament, emitting thermionic electrons that neutralize the incoming ions. Discharging of ions is important when insulating ceramics are used as substrates. The species are then deposited on a water-cooled substrate. Operating conditions are adjusted to give an ion beam with an acceleration energy of about 200 eV and a current density of about 1 mA/cm^2. Under this operating condition, the deposition rate is about 0.1 nm/s which is slow. Incidentally, for deposition of tough and soft coatings, these are deposited at high acceleration energy of about 400 eV at a deposition rate of about 1 nm/s. The ion-beam deposited carbon coatings are reported to be hydrogenated (30–40 at.% hydrogen).

33.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique

The lack of electrodes and its ability to create high densities of charged and excited species at low pressures ($\leq 10^{-4}$ Torr) in ECR (electron cyclotron resonance) plasma, make it an attractive processing discharge for coating depositions [33.66]. In the ECR-CVD carbon-coating deposition process described by *Suzuki* and *Okada* [33.67] and used by *Li* and *Bhushan* [33.49, 50] and *Sundararajan* and *Bhushan* [33.51], microwave power is generated by a magnetron operating in continuous mode at a frequency of 2.45 GHz (Fig. 33.4c). The plasma chamber functions as a microwave cavity resonator. The magnetic coils arranged around the plasma chamber generate a magnetic field of 875 G necessary for electron cyclotron resonance conditions. The substrate is placed on a stage which is capacitively connected to a 13.56 MHz RF generator. The process gas is introduced into the plasma chamber and the hydrocarbon ions produced are accelerated by a negative self-bias voltage, which is generated by applying RF power to the substrate. Both the substrate stage and plasma chamber are water-cooled. The process gas used is 100% ethylene and its flow rate is held constant at 100 sccm; the microwave power is 100–900 W while the RF power is 30–120 W. The pressure during deposition is kept close to the optimum value of 5.5×10^{-3} Torr. Before the deposition, the substrates are cleaned using Ar ions generated in the ECR plasma chamber.

33.1.4 Sputtering Deposition Technique

In the DC planar magnetron sputtered carbon coating technique [33.13, 34, 38, 41, 68–72], a carbon coating is deposited by the sputtering of a graphite target with Ar ion plasma. In the glow discharge, positive ions from the plasma strike the target with sufficient energy to dislodge atoms by momentum transfer, which are intercepted by the substrate. First, a circa 5 nm thick amorphous silicon adhesion layer is deposited by sputtering if the deposition process is to be carried out on a nonsilicon surface. In the process used by *Bhushan* and coworkers (e.g., [33.12]), the coating is deposited by sputtering of the graphite target with Ar ion plasma, at 300 W power for a 200 mm diameter target and a pressure of about 0.5 Pa (6 mTorr) (Fig. 33.4d). Plasma is generated by applying a DC potential between the substrate and the target. *Bhushan* et al. [33.36] reported that the sputtered carbon coating contains about 35 at.% hydrogen. The hydrogen is derived from hydrocarbon contaminants present in the deposition chamber. To produce a hydrogenated carbon coating with a larger concentration of hydrogen, deposition is carried out in Ar and hydrogen plasma.

33.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique

In the RF-PECVD carbon coating deposition process as used by Bhushan and coworkers (e.g., [33.12]), the carbon coating is deposited by adsorption of most free radicals of hydrocarbon to the substrate and chemical bonding to other atoms on the surface. The hydrocarbon species are produced by the RF plasma decomposition of hydrocarbon precursors such as acetylene (C_2H_2) (Fig. 33.4e) [33.28, 70, 73–76]. Instead of requiring thermal energy, as in thermal CVD, the energetic electrons in the plasma (at pressures ranging from 1 to 5×10^2 Pa, typically less than 10 Pa) can activate almost any reaction among the gases in the glow discharge at relatively low substrate temperatures ranging from 100 to 600 °C (typically less than 300 °C). To deposit the coating on nonsilicon substrates, a circa 4 nm thick amorphous silicon adhesion layer, used to improve adhesion, is first deposited under similar conditions from a gas mixture of 1% silane in argon [33.77]. In the process used by *Bhushan* and coworkers [33.12], the plasma is sustained in a parallel-plate geometry by a capacitive discharge at 13.56 MHz, at a surface power density on the order of 100 mW/cm^2. The deposition is carried out at a flow rate on the order of 6 sccm and at a pressure on the order of 4 Pa (30 mTorr) on a cathode-mounted substrate maintained at a substrate temperature of 180 °C. The cathode bias is fixed at the order of −120 V with an external DC power supply attached to the substrate (powered electrode). The carbon coatings deposited by PECVD usually contain hydrogen up to 50% [33.36, 78].

33.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties

The chemical structure and properties of amorphous carbon coatings are a function of the conditions of deposition. It is important to understand the relationship between the chemical structure of amorphous carbon coatings and their properties in order to define useful deposition parameters. Amorphous carbon films are metastable phases formed when carbon particles are condensed on a substrate. The prevailing atomic arrangement in the DLC coatings is amorphous or quasiamorphous with small diamond (sp^3), graphite (sp^2), and other unidentifiable micro- or nanocrystallites. The coatings in dependence on the deposition process and its conditions contain varying amounts of sp^3/sp^2 ratio and hydrogen. The sp^3/sp^2 ratio of DLC coatings ranges typically from 50% to close to 100% with an increase in hardness with sp^3/sp^2 ratio. Most DLC coatings except those produced by a filtered cathodic arc contain from a few to about 50 at.% hydrogen. Sometimes hydrogen and nitrogen are deliberately added to produce hydrogenated (a-C:H) and nitrogenated amorphous carbon (a-C:N) coatings, respectively. Hydrogen helps to stabilize sp^3 sites (most of the carbon atoms attached to hydrogen have a tetrahedral structure). Therefore, the sp^3/sp^2 ratio for hydrogenated carbon is higher [33.31]. The optimal sp^3/sp^2 ratio in a random covalent network composed of sp^3 and sp^2 carbon sites (N_{sp^2} and N_{sp^3}) and hydrogen is [33.31],

$$\frac{N_{sp^3}}{N_{sp^2}} = \frac{6X_H - 1}{8 - 13X_H} , \qquad (33.1)$$

where X_H is the hydrogen atom fraction. The hydrogenated carbon has a larger optical band gap, higher electrical resistivity (semiconductor), and has a lower optical absorption or high optical transmission. The hydrogenated coatings have a lower density probably because of reduction of cross-linking due to hydrogen incorporation. However, hardness decreases with an increase in hydrogen even though the proportion of sp^3 sites increases (that is, as the local bonding environment becomes more diamond-like) [33.79, 80]. It has been speculated that the high hydrogen content introduces frequent terminations in the otherwise strong three-dimensional network and that hydrogen increases the soft polymeric component of the structure more than it enhances the cross-linking sp^3 fraction.

A number of investigations have been performed to identify the microstructure of amorphous carbon films using a variety of techniques such as Raman spectroscopy, EELS, nuclear magnetic resonance, optical measurements, transmission electron microscopy, and x-ray photoelectron spectroscopy [33.34]. The structure of diamond-like amorphous carbon is amorphous or quasiamorphous with small graphitic (sp^2), tetrahedrally coordinated (sp^3), and other unidentifiable nanocrystallites (typically on the order of a couple of nm, randomly oriented) [33.34, 81, 82]. These studies indicate that the chemical structure and physical properties of the coatings are quite variable, depending on deposition techniques and film growth conditions. It is clear that both sp^2 and sp^3-bonded atomic sites are incorporated in diamond-like amorphous carbon coatings and that the physical and chemical properties of the coatings depend strongly on their chemical bonding and microstructure. Systematic studies have been conducted for chemical characterization and to investigate how the physical and chemical properties of amorphous carbon coatings vary as a function of deposition parameters (e.g., [33.34, 36, 41]). EELS and Raman spectroscopy are commonly used to characterize the chemical bonding and microstructure. Hydrogen concentration of the coatings is obtained by means of forward recoil spectrometry (FRS). A variety of physical properties of the coatings relevant to tribological performance are also measured.

To show conventional data obtained from characterization of typical amorphous carbon coatings and their relationships to physical properties, we present data in the following sections on several sputtered, one RF-PECVD amorphous carbon and one microwave-PECVD (MPECVD) diamond coating [33.34, 36, 41]. The sputtered coatings were DC magnetron sputtered at a chamber pressure of 10 mTorr, under sputtering power densities of 0.1 and 2.1 W/cm^2 in a pure Ar plasma, labeled W1 and W2, respectively. They were prepared at a power density of 2.1 W/cm^2 with various hydrogen fractions of 0.5, 1, 3, 5, 7, and 10% in Ar/H gas mixtures; these are labeled H1, H2, H3, H4, H5, and H6, respectively.

33.2.1 EELS and Raman Spectroscopy

EELS and Raman spectra were obtained of four sputtered (W1, W2, H1, and H3) and one PECVD carbon sample. Figure 33.5 shows the EELS spectra of these carbon coatings. EELS spectra for bulk diamond and polycrystalline graphite in an energy range up to 50 eV are also shown in Fig. 33.5. One prominent peak is seen at 35 eV in diamond, while two peaks are seen at 27 and 6.5 eV in graphite, which are called $(\pi + \sigma)$

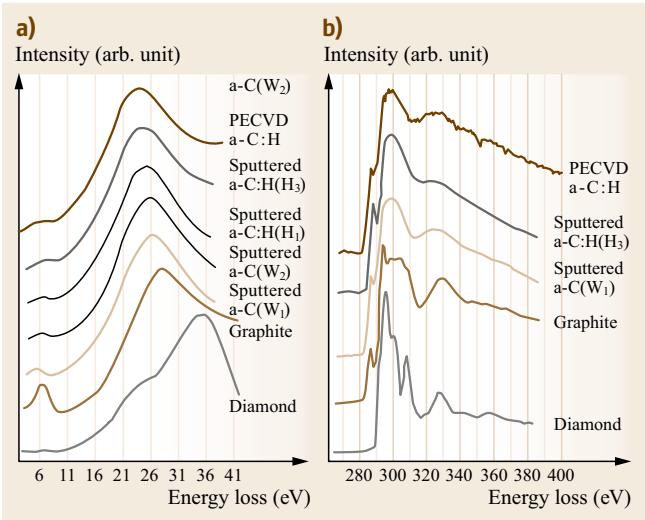


Fig. 33.5 **(a)** Low-energy and **(b)** high-energy EELS of DLC coatings produced by DC magnetron sputtering and RF-PECVD techniques. Data for bulk diamond and polycrystalline graphite are included for comparisons (after [33.36])

and (π) peaks, respectively. These peaks are produced by the energy loss of transmitted electrons to the plasmon oscillation of the valence electrons. The $\pi + \sigma$ peak in each coating is positioned at a lower energy region than that of graphite. The π peaks in the W series and PECVD samples are also seen at a lower energy region than that of the graphite. However, the π peaks in the H-series are comparable to or higher than that of graphite (Table 33.2). The plasmon oscillation frequency is proportional to the square root of the corresponding electron density to a first approximation. Therefore, the samples in the H-series most likely have a higher density of π electrons than the other samples.

Amorphous carbon coatings mainly contain a mixture of sp^2- and sp^3-bonds even though there is some evidence for the presence of sp-bonds as well [33.83]. The PECVD coatings and the H-series coatings in this study have nearly the same mass density as seen in Table 33.4, to be presented later, but the former have a lower concentration of hydrogen (18.1%) than the H-series (35–39%), as seen in Table 33.3, to be presented later. The relatively low energy position of the π peaks of PECVD coatings, compared to those of the H-series, indicates that the PECVD coatings contain a higher fraction of sp^3-bonds than the sputtered hydrogenated carbon coatings (H-series).

Figure 33.5b shows the EELS spectra associated with inner-shell (*K*-shell) ionization. Again, the spectra for diamond and polycrystalline graphite are included for comparison. Sharp peaks are observed at 285.5 and 292.5 eV in graphite, while no peak is seen at 285.5 eV in diamond. The general features of the *K*-shell EELS spectra for the sputtered and PECVD carbon samples resemble the features of graphite, but with the higher energy features smeared. The observation of the peak at 285.5 eV in the sputtered and PECVD coatings also indicates the presence of sp^2-bonded atomic sites in the coatings. All these spectra peak at 292.5 eV, similar to the spectra of graphite, but the peak in graphite is sharper.

Raman spectra from samples W1, W2, H1, and PECVD are shown in Fig. 33.6. Raman spectra could not be observed in specimens H2 and H3 due to high fluorescence signals. The Raman spectra of single-crystal diamond and polycrystalline graphite are also shown for comparison in Fig. 33.6. The results of the spectral fits are summarized in Table 33.2. We will focus on the G-band position, which has been shown to be related to the fraction of sp^3-bonded sites. Increasing the power density in the amorphous carbon coatings (W1 and W2) results in a higher G-band frequency, implying a smaller fraction of sp^3-bonding in W2 than in W1. This is consistent with the higher density of W1. H1 and PECVD still have lower G-band positions than W1, implying an even higher fraction of sp^3-bonding, which is presumably caused by the incorporation of H atoms into the lattice. The high hardness of H3 might be attributed to efficient sp^3 cross-linking of small, sp^2-ordered domains.

The Raman spectrum of a MPECVD diamond coating is shown in Fig. 33.6. The diamond Raman peak is at 1333 cm^{-1} with a line width of 7.9 cm^{-1}. There is a small broad peak around 1525 cm^{-1}, which is attributed to a small amount of a-C:H. This impurity peak is not intense enough to fit to separate G- and D-bands. The diamond peak frequency is very close to that of natural diamond (1332.5 cm^{-1}; Fig. 33.6), indicating that the coating is not under stress [33.84]. The large line width compared to that of natural diamond (2 cm^{-1}) indicates that the microcrystallites likely have a high concentration of defects [33.85].

33.2.2 Hydrogen Concentrations

FRS analysis of six sputtered (W2, H2, H3, H4, H5, and H6) coatings, one PECVD coating, and one diamond coating was performed. Figure 33.7 shows an overlay of the spectra from the six sputtered samples. Similar spectra were obtained from the PECVD and the diamond films. Table 33.3 shows the H and C fractions as well as the amount of impurities (Ar and O) in the films in atomic %. Most apparent is the large fraction of H in the sputtered films. Regardless of how much H_2 is in the Ar sputtering gas, the H content of the coatings is about the same, $\approx$ 35 at.%. Interestingly, there is still $\approx$ 10% H present in the coating sputtered in pure Ar

Table 33.2 Experimental results from EELS and Raman spectroscopy (after [33.36])

Sample	EELS peak position		Raman peak position		Raman FWHM[a]		I_D/I_G[d]
	π (eV)	$\pi+\sigma$ (eV)	G-band[b] (cm^{-1})	D-band[c] (cm^{-1})	G-band (cm^{-1})	D-band (cm^{-1})	
Sputtered a-C coating (W1)	5.0	24.6	1541	1368	105	254	2.0
Sputtered a-C coating (W2)	6.1	24.7	1560	1379	147	394	5.3
Sputtered a-C:H coating (H1)	6.3	23.3	1542	1334	95	187	1.6
Sputtered a-C:H coating (H3)	6.7	22.4	e	e	e	e	e
PECVD a-C:H coating	5.8	24.0	1533	1341	157	427	1.5
Diamond coating	–	–	1525[f]	1333[g]	–	8[g]	–
Graphite (for reference)	6.4	27.0	1580	1358	37	47	0.7
Diamond (for reference)	–	37.0	–	1332[g]	–	2[g]	–

[a] Full width at half-maximum
[b] Peak associated with sp^2 *graphite* carbon
[c] Peak associated with sp^2 *disordered* carbon (not sp^3-bonded carbon)
[d] Intensity ratio of the D-band to the G-band
[e] Fluorescence
[f] Includes D- and G-band, signal too weak to analyze
[g] Peak position and width for diamond phonon

Table 33.3 Experimental results of FRS analysis (after [33.36])

Sample	Ar/H ratio	C (at.% ± 0.5)	H (at.% ± 0.5)	Ar (at.% ± 0.5)	O (at.% ± 0.5)
Sputtered a-C coating (W2)	100/0	90.5	9.3	0.2	–
Sputtered a-C:H coating (H2)	99/1	63.9	35.5	0.6	–
Sputtered a-C:H coating (H3)	97/3	56.1	36.5	–	7.4
Sputtered a-C:H coating (H4)	95/5	53.4	39.4	–	7.2
Sputtered a-C:H coating (H5)	93/7	58.2	35.4	0.2	6.2
Sputtered a-C:H coating (H6)	90/10	57.3	35.5	–	7.2
PECVD a-C:H coating	99.5% CH_4	81.9	18.1	–	–
Diamond coating	H_2-1 mol % CH_4	94.0	6.0	–	–

Table 33.4 Experimental results of physical properties (after [33.36])

Sample	Mass density (g/cm^3)	Nanohardness (GPa)	Elastic modulus (GPa)	Electrical resistivity (Ω cm)	Compressive residual stress (GPa)
Sputtered a-C coating (W1)	2.1	15	141	1300	0.55
Sputtered a-C:H coating (W2)	1.8	14	136	0.61	0.57
Sputtered a-C:H coating (H1)	–	14	96	–	> 2
Sputtered a-C:H coating (H3)	1.7	7	35	$> 10^6$	0.3
PECVD a-C:H coating	1.6–1.8	33–35	≈ 200	$> 10^6$	1.5–3.0
Diamond coating	–	40–75	370–430	–	–
Graphite (for reference)	2.267	Soft	9–15	5×10^{-5}[a], 4×10^{-3}[b]	0
Diamond (for reference)	3.515	70–102	900–1050	10^7-10^{20}	0

[a] Parallel to layer planes
[b] Perpendicular to layer planes

(W2). It is also interesting to note that Ar is present only in coatings grown under low (< 1%) H content in the Ar carrier gas. The presence of O in the coatings, combined with the fact that the coatings were prepared approximately 9 months before the FRS analysis, caused suspicion that they had absorbed water vapor, and that this may be the cause for the H peak in specimen W2.

All samples were annealed for 24 h at 250 °C in a flowing He furnace and then reanalyzed. Surprisingly, the H content of all coatings measured increased slightly, even though the O content decreased, and W2 still had a substantial amount of H_2. This slight increase in H concentration is not understood. However, since the H concentration did not decrease with the oxygen

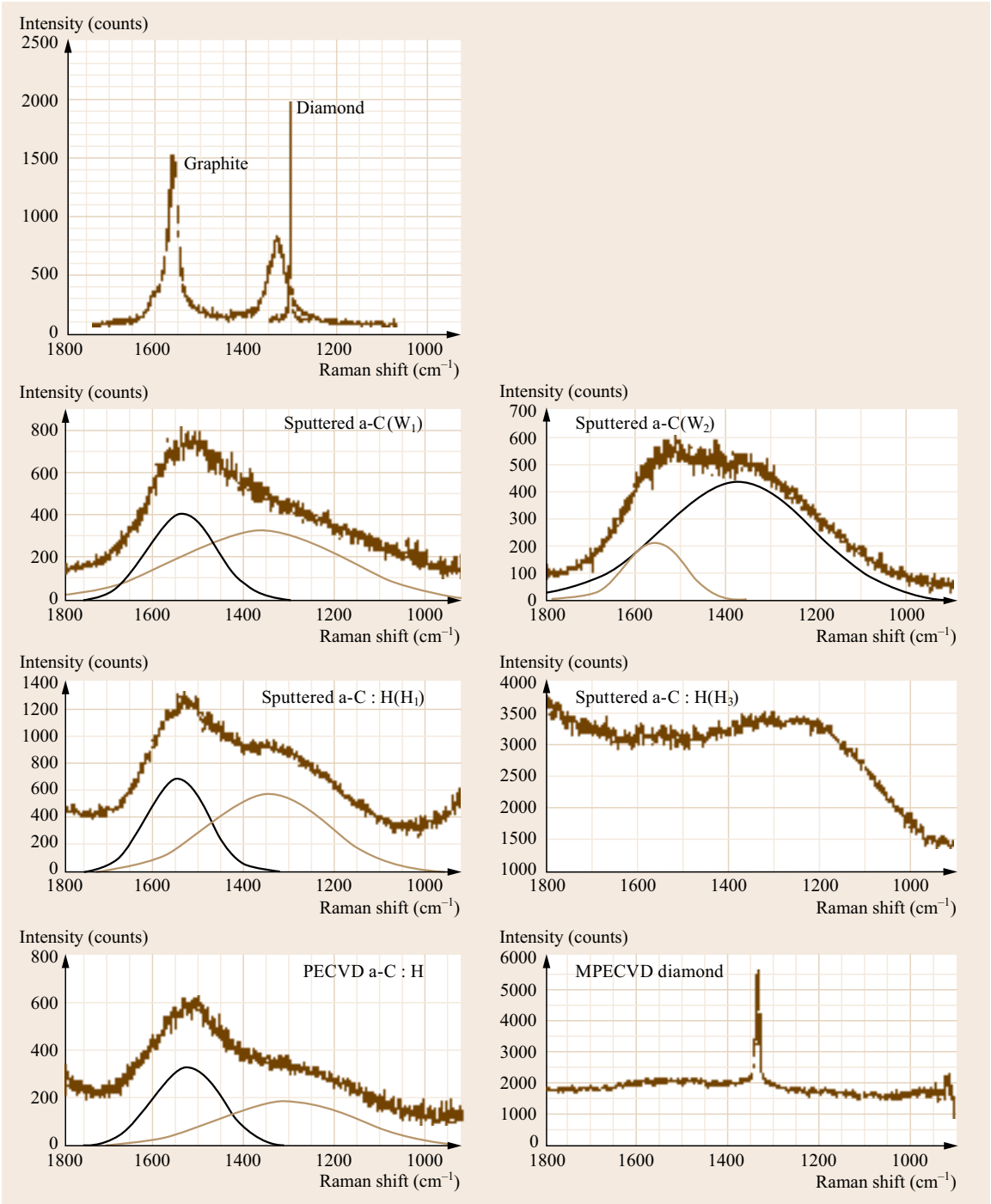


Fig. 33.6 Raman spectra of DLC coatings produced by DC magnetron sputtering and RF-PECVD techniques and a diamond film produced by the MPECVD technique. Data for bulk diamond and microcrystalline graphite are included for comparison (after [33.36])

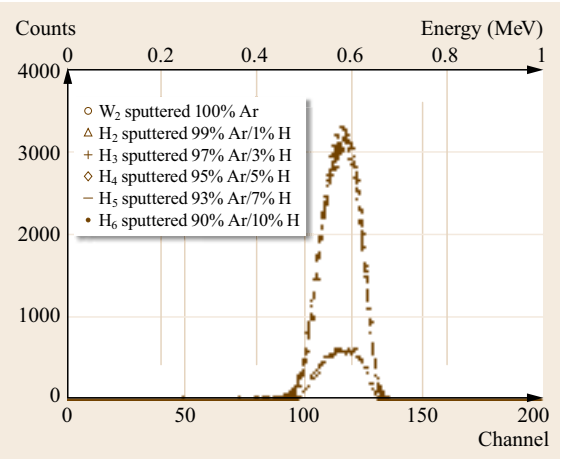


Fig. 33.7 FRS spectra of six DLC coatings produced by DC magnetron sputtering (after [33.36])

as a result of annealing, it suggests that the high H concentration is not due to adsorbed water vapor. The PECVD film has more H ($\approx 18\%$) than the sputtered films initially, but after annealing it has the same fraction as specimen W2, the film sputtered in pure Ar. The diamond film has the smallest amounts of hydrogen, as seen in Table 33.3.

33.2.3 Physical Properties

Physical properties of the four sputtered (W1, W2, H1, and H3) coatings, one PECVD coating, one diamond coating, and bulk diamond and graphite are presented in Table 33.4. The hydrogenated carbon and the diamond coatings have very high resistivity compared to unhydrogenated carbon coatings. It appears that unhydrogenated carbon coatings have a higher density than the hydrogenated carbon coatings, although both groups are less dense than graphite. The density depends on deposition technique and parameters. It appears that unhydrogenated sputtered coatings deposited at low power exhibit the highest density. Nanohardness of hydrogenated carbon is somewhat lower than that of the unhydrogenated carbon coatings. PECVD coatings are significantly harder than sputtered coatings. The nanohardness and modulus of elasticity of the diamond coatings are very high compared to that of DLC coatings even though the hydrogen content is similar. The compressive residual stresses of the PECVD coatings are substantially higher than those of sputtered coatings, which is consistent with the hardness results.

Figure 33.8a shows the effect of hydrogen in the plasma on the residual stresses and the nanohardness for sputtered coatings W2 and H1 to H6. The coatings

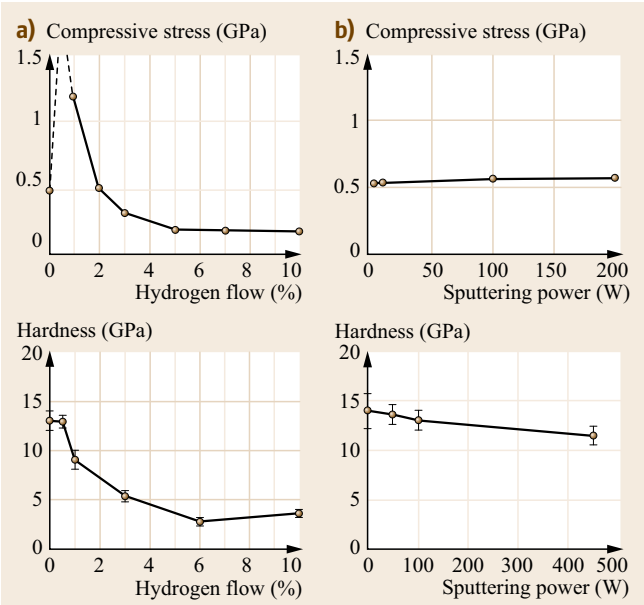


Fig. 33.8a,b Residual compressive stresses and nanohardness **(a)** as a function of hydrogen flow rate; sputtering power is 100 W and target diameter is 75 mm (power density $= 2.1\,\mathrm{W/cm^2}$), and **(b)** as a function of sputtering power over a 75 mm diameter target with no hydrogen added in the plasma (after [33.41])

made with H_2 flow between 0.5 and 1.0% delaminate very quickly, even when only a few tens of nm thick. In pure Ar and at H_2 flows greater than 1%, the coatings appear to be more adhesive. The tendency of some coatings to delaminate can be caused by intrinsic stress in the coating, which is measured by substrate bending. All of the coatings in the figure are in compressive stress. The maximum stress occurs between 0 and 1% H_2 flow, but the stress cannot be quantified in this range because the coatings instantly delaminate upon exposure to air. At higher hydrogen concentrations the stress gradually diminishes. A generally decreasing trend is observed in the hardness of the coatings as hydrogen content increased. The hardness decreases slightly, going from 0% H_2 to 0.5% H_2, and then decreases sharply. These results are probably lower than the true values because of local delamination around the indentation point. This is especially likely for the 0.5% and 1.0% coatings where delamination is visually apparent, but may also be true to a lesser extent for the other coatings. Such an adjustment would bring the hardness profile into closer correlation with the stress profile. *Weissmantel* et al. [33.69] and *Scharff* et al. [33.86] observed a downturn in hardness for high bias and low pressure of hydrocarbon gas in ion-plated carbon coating and therefore presumably low hydrogen content in support of the above contention.

Figure 33.8b shows the effect of sputtering power (with no hydrogen added in the plasma) on the residual stresses and nanohardness for various sputtered coatings. As the power decreases, compressive stress does not seem to change while nanohardness slowly increases. The rate of change becomes more rapid at very low power levels.

Addition of H_2 during sputtering of carbon coatings increases H concentration in the coating. Hydrogen causes the character of the C-C bonds to shift from sp^2 to sp^3, and the rising number of C-H bonds which ultimately relieves stress and produces a softer *polymer-like* material. Low-power deposition, like the presence of hydrogen, appears to stabilize the formation of sp^3 C-C bonds, increasing hardness. These coatings have lower stress and better adhesion. An increase in temperature during deposition at high power density results in graphitization of the coating material, responsible for a decrease in hardness with an increase in power density. Unfortunately, low power also means impractically low deposition rates.

33.2.4 Summary

On the basis of EELS and Raman data, all DLC coatings have both sp^2- and sp^3-bonding. The sp^2/sp^3 bonding ratio depends on deposition techniques and parameters. The DLC coatings deposited by sputtering and PECVD, contain significant concentrations of hydrogen while the diamond coating contains only small amounts of hydrogen impurity. Sputtered coatings with no deliberate addition of hydrogen in the plasma contain a significant amount of hydrogen. Regardless of how much hydrogen is in the Ar sputtering gas, the hydrogen content of the coatings increases initially with no further increase thereafter.

Hydrogen flow and sputtering power density affect the mechanical properties of these coatings. Maximum compressive residual stress and hardness occur between 0% and 1% hydrogen flow, resulting in rapid delamination. Low sputtering power moderately increases hardness while relieving residual stress.

33.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques

33.3.1 Micromechanical Characterization

Common mechanical characterizations include measurement of hardness and elastic modulus, fracture toughness, fatigue life, and scratch and wear testing. Nanoindentation and atomic force microscopy (AFM) are used for mechanical characterization of ultrathin films.

Hardness and elastic modulus are calculated from the load displacement data obtained by nanoindentation at loads ranging typically from 0.2 to 10 mN using a commercially available nanoindenter [33.23, 87]. This instrument monitors and records the dynamic load and displacement of the three-sided pyramidal diamond (Berkovich) indenter during indentation. For fracture toughness measurement of ultrathin films ranging from 100 nm to a few μm, a nanoindentation-based technique is used in which through-thickness cracking in the coating is detected from a discontinuity observed in the load-displacement curve and energy released during the cracking is obtained from the curve [33.88–90]. On the basis of energy released, fracture mechanics analysis is then used to calculate fracture toughness. An indenter with a cube-corner tip geometry is preferred because the through-thickness cracking of hard films can be accomplished at lower loads. In fatigue measurement, a conical diamond indenter having a tip radius of about one micron is used and load cycles of a sinusoidal shape are applied [33.91, 92]. The fatigue behavior of coatings is studied by monitoring the change in contact stiffness which is sensitive to the damage formation.

Hardness and Elastic Modulus

For materials that undergo plastic deformation, high hardness and elastic modulus are generally needed for low friction and wear, whereas for brittle materials, high fracture toughness is needed [33.1–3, 21]. DLC coatings used for many applications are hard and brittle and values of hardness and fracture toughness need to be optimized.

Representative load-displacement plots of indentations made at 0.2 mN peak indentation load on 100 nm thick DLC coatings deposited by the four deposition techniques on single-crystal silicon substrate are compared in Fig. 33.9. Indentation depths at peak load range from about 18−26 nm, smaller than that of the coating thickness. Many of the coatings exhibit discontinuity or pop-in marks in the loading curve, which indicate a sudden penetration of the tip into the sample. A nonuniform penetration of the tip into a thin coating possibly results due to either formation of cracks in the coating, formation of cracks at the coating-substrate interface, or

debonding or delamination of the coating from the substrate.

The hardness and elastic modulus values at a peak load of 0.2 mN on the various coatings and single-crystal silicon substrate are summarized in Table 33.5 and Fig. 33.10 [33.48, 50, 90, 91]. Typical values for the peak and residual indentation depths ranged from 18 to 26 nm and 6 to 12 nm, respectively. The FCA coatings exhibit the highest hardness of 24 GPa and elastic modulus of 280 GPa of the various coatings, followed by the ECR-CVD, IB, and SP coatings. Hardness and elastic modulus have been known to vary over a wide range with sp^3-to-sp^2 bonding ratio, which depends on the kinetic energy of the carbon species and amount of hydrogen [33.6, 31, 48, 93, 94]. The high hardness and elastic modulus of the FCA coatings are attributed to the high kinetic energy of carbon species involved in the FCA deposition [33.12, 48]. *Anders* et al. [33.58] also reported a high hardness, measured by nanoindentation, of about 45 GPa for cathodic arc carbon coatings. They observed a change in hardness from 25 to 45 GPa with pulsed bias voltage and bias duty cycle. The high hardness of cathodic arc carbon was attributed to the high percentage (more than 50%) of sp^3 bonding. *Savvides* and *Bell* [33.95] reported an increase in hardness from 12 to 30 GPa and in elastic modulus from 62 to 213 GPa with an increase of sp^3-to-sp^2 bonding ratio, from 3 to 6, for a-C:H coating deposited by low-energy ion-assisted unbalanced magnetron sputtering of a graphite target in an Ar-H_2 mixture.

Bhushan et al. [33.36] reported hardnesses of about 15 and 35 GPa and elastic moduli of about 140 and

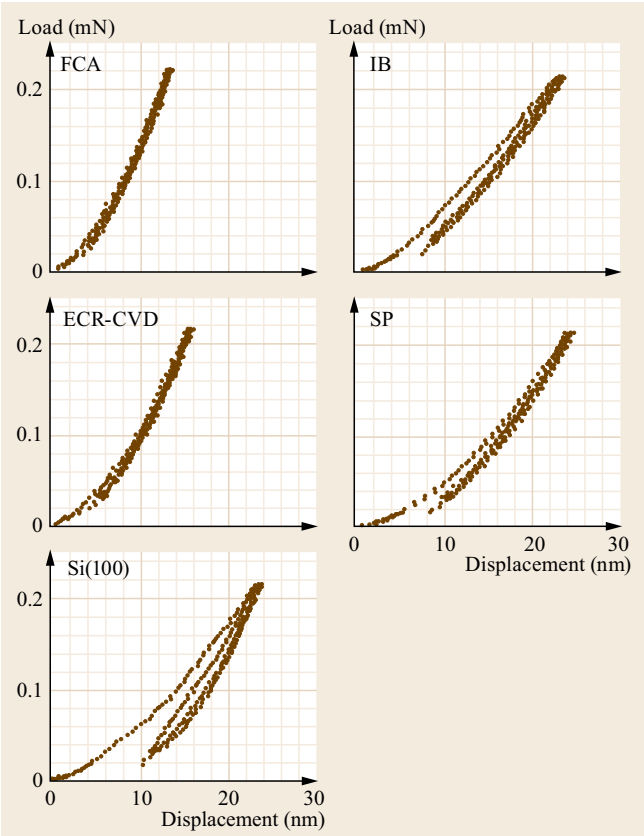


Fig. 33.9 Load versus displacement plots for various 100 nm thick amorphous carbon coatings on a single-crystal silicon substrate and a bare substrate

Table 33.5 Hardness, elastic modulus, fracture toughness, fatigue life, critical load during scratch, coefficient of friction during accelerated wear testing and residual stresses of various DLC coatings on single-crystal silicon substrate

Coating	Hardness[a] [33.49] (GPa)	Elastic modulus[a] [33.49] (GPa)	Fracture toughness[a] [33.90] (MPa $m^{1/2}$)	Fatigue life[b] N_f[d] [33.91] $\times 10^4$	Critical load during scratch[b] [33.49] (mN)	Coefficient of friction during accelerated wear testing[b] [33.49]	Compressive residual stress[c] [33.48] (GPa)
Cathodic arc carbon coating (a-C)	24	280	11.8	2.0	3.8	0.18	12.5
Ion-beam carbon coating (a-C:H)	19	140	4.3	0.8	2.3	0.18	1.5
ECR-CVD carbon coating (a-C:H)	22	180	6.4	1.2	5.3	0.22	0.6
DC sputtered carbon coating (a-C:H)	15	140	2.8	0.2	1.1	0.32	2.0
Bulk graphite (for comparison)	Very soft	9–15	–	–	–	–	–
Diamond (for comparison)	80–104	900–1050	–	–	–	–	–
Si(100) substrate	11	220	0.75	–	0.6	0.55	0.02

[a] Measured on 100 nm thick coatings
[b] Measured on 20 nm thick coatings
[c] Measured on 400 nm thick coatings
[d] N_f was obtained for a mean load of 10 μN and a load amplitude of 8 μN

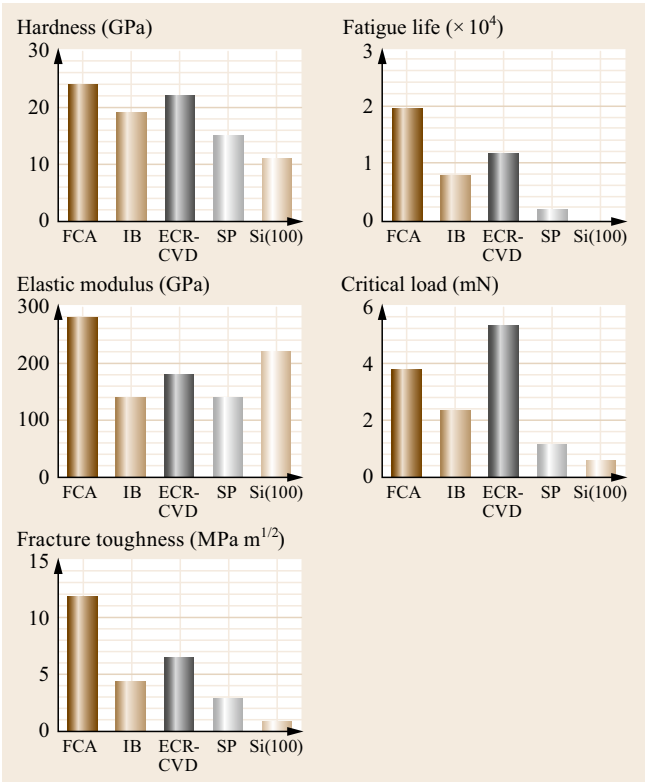


Fig. 33.10 Bar charts summarizing the data of various coatings and single-crystal silicon substrate. Hardness, elastic modulus, and fracture toughness were measured on 100 nm thick coatings, and fatigue life and critical load during scratch were measured on 20 nm thick coatings

200 GPa, measured by nanoindentation, for a-C:H coatings deposited by DC magnetron sputtering and RF-plasma-enhanced chemical vapor deposition techniques, respectively. The high hardness of RF-PECVD a-C:H coatings is attributed to a higher concentration of sp^3 bonding than in a sputtered hydrogenated a-C:H coating. Hydrogen is believed to play a crucial role in the bonding configuration of carbon atoms by helping to stabilize tetrahedra coordination (sp^3 bonding) of carbon species. *Jansen* et al. [33.79] suggested that the incorporation of hydrogen efficiently passivates the dangling bonds and saturates the graphitic bonding to some extent. However, a large concentration of hydrogen in the plasma in sputter deposition is undesirable. *Cho* et al. [33.34] and *Rubin* et al. [33.41] observed that hardness decreased from 15 to 3 GPa with increased hydrogen content. *Bhushan* and *Doerner* [33.96] reported a hardness of about 10–20 GPa and an elastic modulus of about 170 GPa, measured by nanoindentation, for 100 nm thick DC-magnetron sputtered a-C:H on the silicon substrate.

Residual stresses measured using a well-known curvature measurement technique are also presented in Table 33.5. The DLC coatings are under significant compressive internal stresses. Very high compressive stresses in FCA coatings are believed to be partly responsible for their high hardness. However, high stresses result in coating delamination and buckling. For this reason, the coatings thicker than about 1 μm have a tendency to delaminate from the substrates.

Fracture Toughness

Representative load-displacement curves of indentations on the 400 nm thick cathodic arc carbon coating on silicon at various peak loads are shown in Fig. 33.11. Steps are found in all curves as shown by the arrows in Fig. 33.11a. In the 30 mN SEM (scanning electron microscope) micrograph, in addition to several radial cracks, ring-like through-thickness cracking is observed with small lips of material overhanging the edge of indentation. The step at about 23 mN in the loading curves of indentations made at 30 and 100 mN peak indentation loads results from the ring-like through-thickness cracking. The step at 175 mN in the loading curve of the indentation made at 200 mN peak indentation load is caused by spalling and second ring-like through-thickness cracking.

On the basis of work by *Li* et al. [33.88], the fracture process progresses in three stages: (1) first ring-like through-thickness cracks form around the indenter by high stresses in the contact area, (2) delamination and buckling occur around contact area at the coating-substrate interface by high lateral pressure, (3) second ring-like through-thickness cracks and spalling are generated by high bending stresses at the edges of the buckled coating (Fig. 33.12a). In the first stage, if the coating under the indenter is separated from the bulk coating via the first ring-like through-thickness cracking, a corresponding step will be present in the loading curve. If discontinuous cracks form and the coating under the indenter is not separated from the remaining coating, no step appears in the loading curve because the coating still supports the indenter and the indenter cannot suddenly advance into the material. In the second stage, for the coating used in the present study, the advances of the indenter during the radial cracking, delamination, and buckling are not big enough to form steps in the loading curve because the coating around the indenter still supports the indenter, but generate discontinuities which change the slope of the loading curve with increasing indentation loads. In the third stage, the stress concentration at the end of the interfacial crack cannot be relaxed by the propagation of the interfacial crack. With an increase in indentation depth, the height of the bulged coating increases. When the height

reaches a critical value, the bending stresses caused by the bulged coating around the indenter will result in the second ring-like through-thickness crack formation and spalling at the edge of the buckled coating as shown in Fig. 33.12a, which leads to a step in the loading curve. This is a single event and results in the separation of the part of the coating around the indenter from the bulk coating via cracking through coatings. The step in the loading curve is only from coating cracks and not from interfacial or substrate cracks.

The area under the load-displacement curve is the work performed by the indenter during elastic-plastic deformation of the coating-substrate system. The strain energy release in the first/second ring-like cracking and spalling can be calculated from the corresponding steps in the loading curve. Figure 33.12b shows a modeled load-displacement curve. In the figure, OACD is the loading curve and DE is the unloading curve. The first ring-like through-thickness crack should be considered. It should be emphasized that the edge of the buckled coating is far from the indenter, therefore, it does not matter if the indentation depth exceeds the coating thickness or if deformation of the substrate occurs around the indenter when we measure fracture toughness of the coating from the released energy during the second ring-like through-thickness cracking (spalling). Suppose that the second ring-like through-thickness cracking occurs at AC. Now, let us consider the loading curve OAC. If the second ring-like through-thickness crack does not occur, it can be understood that OA will be extended to OB to reach the same displacement as OC. This means that the crack formation changes the loading curve OAB into OAC. For point B, the elastic-plastic energy stored in the coating-substrate system should be OBF. For point C, the elastic-plastic energy stored in the coating-substrate system should be OACF. Therefore, the energy difference before and after the crack generation is the area of ABC, i. e., this energy stored in ABC will be released as strain energy to create the ring-like through-thickness crack. According to the theoretical analysis by *Li* et al. [33.88], the fracture toughness of thin films can be written as

$$K_{\mathrm{Ic}} = \left[\left(\frac{E}{(1-\nu^2)\,2\pi C_{\mathrm{R}}}\right)\left(\frac{U}{t}\right)\right]^{1/2}, \qquad (33.2)$$

where E is the elastic modulus, ν is the Poisson's ratio, $2\pi C_{\mathrm{R}}$ is the crack length in the coating plane, t is the coating thickness, and U is the strain energy difference before andafter cracking.

Using (33.2), the fracture toughness of the coatings are calculated. The loading curve is extrapolated along the tangential direction of the loading curve from the starting point of the step-up to reach the same displacement as the step. The area between the extrapolated line and the step is the estimated strain energy difference before and after cracking. C_{R} is measured from SEM micrographs or AFM images of indentations. The second ring-like crack is where the spalling occurs. For example, for the 400 nm thick cathodic arc carbon coating data presented in Fig. 33.11, U of 7.1 nNm is assessed from the steps in Fig. 33.11a at the peak indentation loads of 200 mN. For C_{R} of 7.0 µm from Fig. 33.11b, $E = 300$ GPa measured using the nanoindenter and an assumed value of 0.25 for ν, the fracture toughness value is calculated as 10.9 MPa $\sqrt{m}$ [33.88, 89]. The fracture toughness and related data for various 100 nm thick DLC coatings are presented in Fig. 33.10 and Table 33.5.

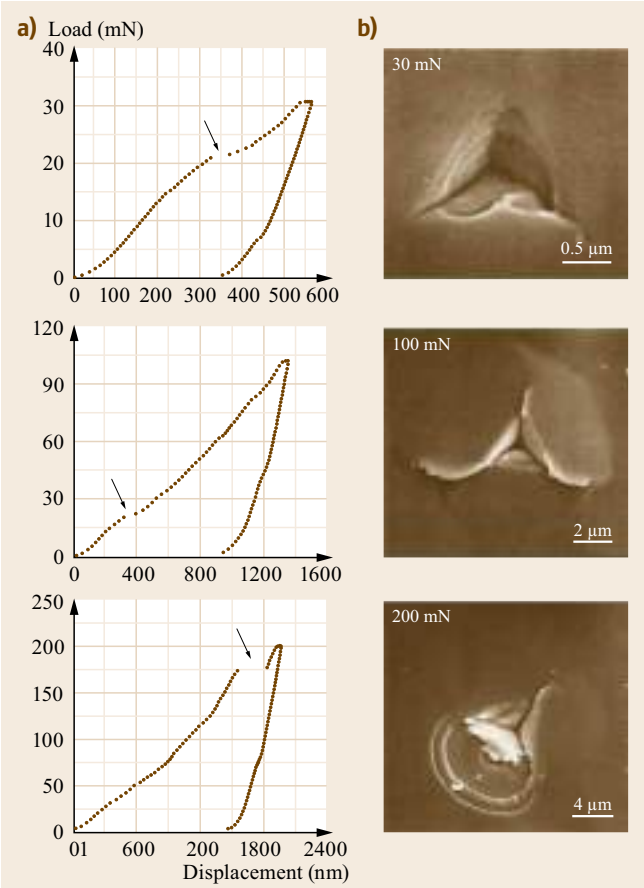


Fig. 33.11 **(a)** Load-displacement curves of indentations made at 30, 100, and 200 mN peak indentation loads using the cube corner indenter, and **(b)** the SEM micrographs of indentations on the 400 nm thick cathodic arc carbon coating on silicon. The arrows indicate steps during the loading portion of the load-displacement curve (after [33.88])

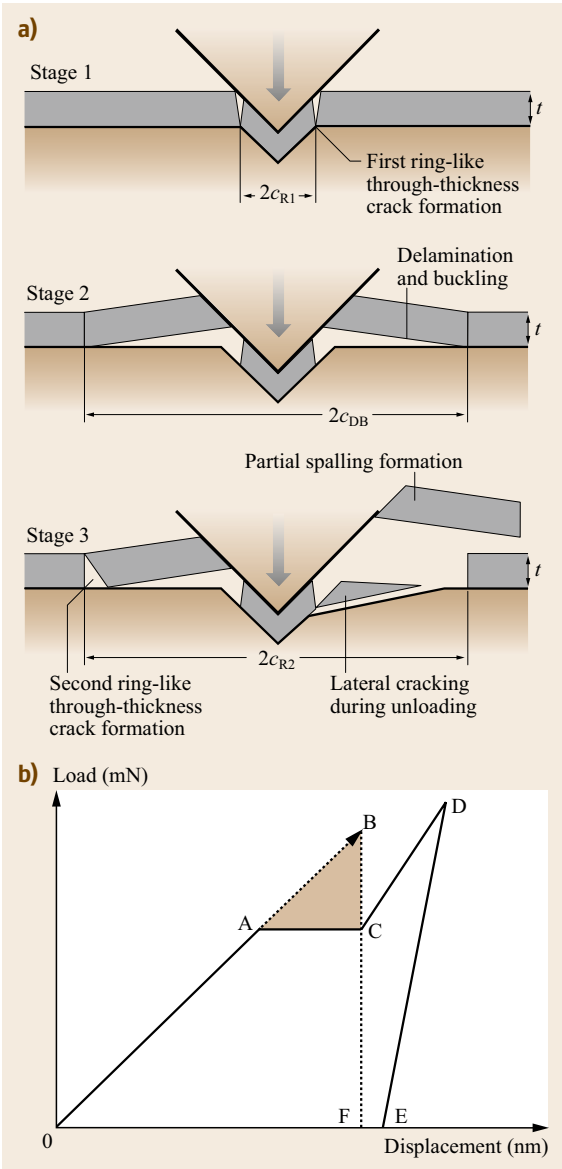


Fig. 33.12 **(a)** Schematic of various stages in nanoindentation fracture for the coatings-substrate system, and **(b)** schematic of a load-displacement curve, showing a step during the loading cycle and associated energy release

Nanofatigue

A delayed fracture resulting from extended service is called fatigue [33.97]. Fatigue fracturing progresses through a material via changes within the material at the tip of a crack, where there is a high stress intensity. There are several situations: cyclic fatigue, stress corrosion, and static fatigue. Cyclic fatigue results from cyclic loading of machine components. In a low-flying slider in a magnetic head-disk interface, isolated asperity contacts occur during use and the fatigue failure occurs in the multilayered thin-film structure of the magnetic disk [33.13]. In many MEMS components, impact occurs and the failure mode is cyclic fatigue. Asperity contacts can be simulated using a sharp diamond tip in an oscillating contact with the component.

Figure 33.13 shows the schematic of a fatigue test on a coating-substrate system using a continuous stiffness measurement (CSM) technique. Load cycles are applied to the coating, resulting in a cyclic stress; P is the cyclic load, P_{mean} is the mean load, P_o is the oscillation load amplitude, and ω is the oscillation frequency. The following results can be obtained: (1) endurance limit; i. e., the maximum load below which there is no coating failure for a preset number of cycles; (2) number of cycles at which the coating failure occurs; and (3) changes in contact stiffness measured by using the unloading slope of each cycle which can be used to monitor the propagation of interfacial cracks during the cyclic fatigue process.

Figure 33.14a shows the contact stiffness as a function of the number of cycles for 20 nm thick FCA coatings cyclically deformed by various oscillation load amplitudes with a mean load of 10 μN at a frequency of 45 Hz. At 4 μN load amplitude, no change in contact stiffness was found for all coatings. This indicates that 4 μN load amplitude is not high enough to damage the coatings. At 6 μN load amplitude, an abrupt decrease in contact stiffness was found at a certain number of cycles for each coating, indicating that fatigue damage has occurred. With increasing load amplitude, the number of cycles to failure, N_f, decreases for all coatings. Load amplitude versus N_f, a so-called S–N curve, is plotted in Fig. 33.14b. The critical load amplitude, below which no fatigue damage occurs (an endurance limit), was identified for each coating. This critical load amplitude together with mean load is of critical importance to the design of head-disk interfaces or MEMS/NEMS device interfaces.

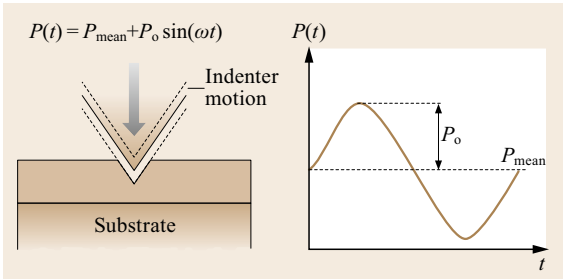


Fig. 33.13 Schematic of a fatigue test on a coating-substrate system using the continuous stiffness measurement technique

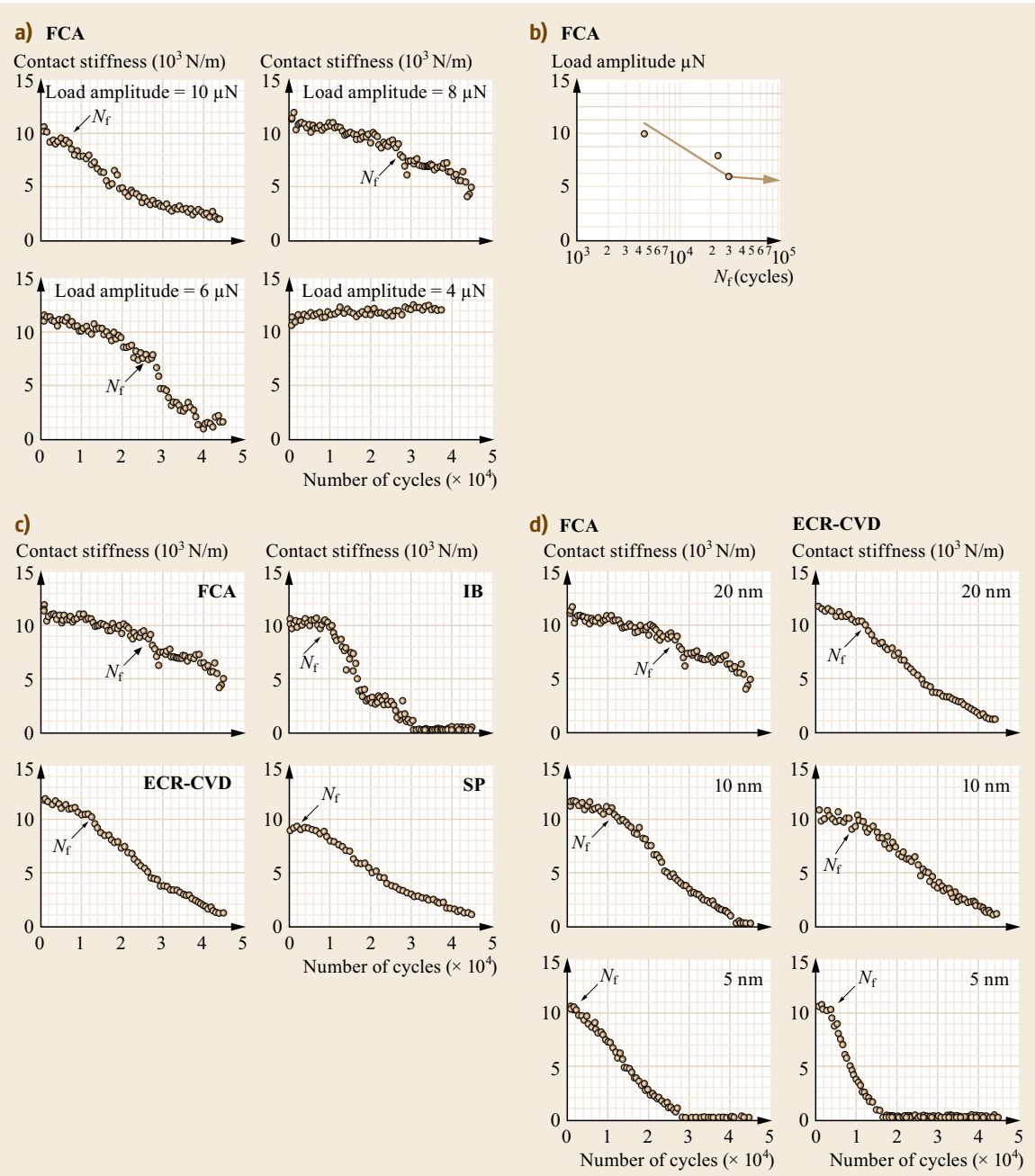


Fig. 33.14 **(a)** Contact stiffness as a function of the number of cycles for 20 nm thick FCA coatings cyclically deformed by various oscillation load amplitudes with a mean load of 10 μN at a frequency of 45 Hz; **(b)** plot of load amplitude versus N_f; **(c)** contact stiffness as a function of the number of cycles for four different 20 nm thick coatings; and **(d)** contact stiffness as a function of the number of cycles for two coatings of different thicknesses

To compare the fatigue lives of the different coatings studied, the contact stiffness as a function of the number of cycles for 20 nm thick FCA, IB, ECR-CVD, and SP coatings cyclically deformed by an oscillation load amplitude of 8 μN with a mean load of 10 μN at a frequency of 45 Hz is shown in Fig. 33.14c. The FCA coating has the longest N_f, followed by ECR-CVD, IB, and SP coatings. In addition, after N_f, the contact

stiffness of the FCA coating shows a slower decrease than the other coatings. This indicates that after N_f, the FCA coating had less damage than others. The fatigue behavior of FCA and ECR-CVD coatings of different thicknesses is compared in Fig. 33.14d. For both coatings, N_f decreases with decreasing coating thickness. At 10 nm, FCA and ECR-CVD have almost the same fatigue life. At 5 nm, the ECR-CVD coating shows a slightly longer fatigue life than the FCA coating. This indicates that even for nanometer thick DLC coatings their microstructure and residual stresses are not uniform across the thickness direction. Thinner coatings are more influenced by interfacial stresses than thicker coatings.

Figure 33.15a shows the high magnification SEM images of 20 nm thick FCA coatings before, at, and after N_f. In the SEM images, the net-like structure is the gold film coated on the DLC coating, which should be ignored in analyzing the indentation fatigue damage. Before N_f, no delamination or buckling was found except the residual indentation mark at magnifications up to 1 200 000 × using SEM. This suggests that only plastic deformation occurred before N_f. At N_f, the coating around the indenter bulged upwards, indicating delamination and buckling. Therefore, it is believed that the decrease in contact stiffness at N_f results from the delamination and buckling of the coating from the substrate. After N_f, the buckled coating was broken down around the edge of the buckled area, forming a ring-like crack. The remaining coating overhung at the edge of the buckled area. It is noted that the indentation size increases with increasing number of cycles. This indicates that deformation, delamination and buckling, and ring-like crack formation occurred over a period of time.

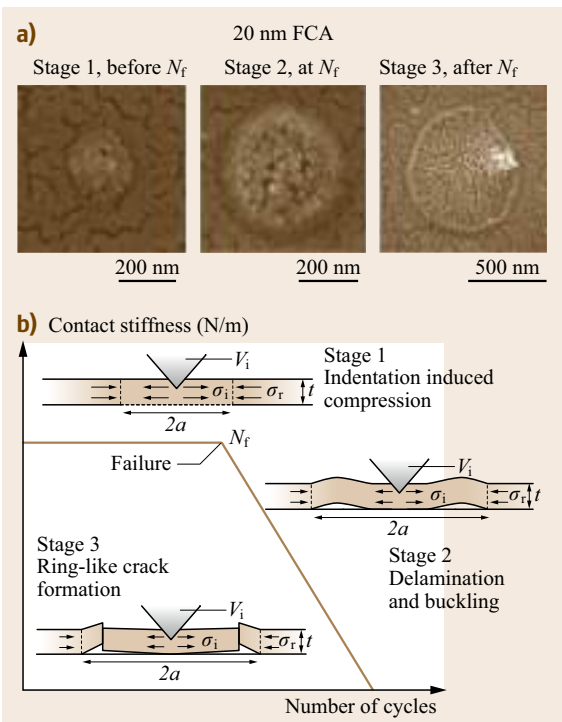


Fig. 33.15 **(a)** High magnification SEM images of a coating before, at, and after N_f, and **(b)** schematic of various stages in the indentation fatigue damage for a coating-substrate system (after [33.91])

The schematic in Fig. 33.15b shows various stages in the indentation fatigue damage for a coating-substrate system. On the basis of this study, three stages appear to exist in the indentation fatigue damage: (1) indentation-induced compression; (2) delamination and buckling; (3) ring-like crack formation at the edge of the buckled coating. Residual stresses are often induced in coatings by the deposition process. The model shown in Fig. 33.15b considers a coating with a uniform biaxial residual compression σ_r. In the first stage, indentation induces elastic/plastic deformation, exerting an outward acting pressure on the coating around the indenter. Interfacial defects like voids and impurities act as original cracks. These cracks propagate and link up as the indentation compressive stress increases. At this stage, the coating, which is under the indenter and above the interfacial crack (with a crack length of $2a$), still maintains a solid contact with the substrate; the substrate still fully supports the coating. Therefore, this interfacial crack does not lead to an abrupt decrease in contact stiffness, but gives rise to a slight decrease in contact stiffness as shown in Fig. 33.14. The coating above the interfacial crack is treated as a rigidly clamped disk. We assume that the crack radius, a, is large compared with the coating thickness t. Since the coating thickness ranges from 20 to 5 nm, this assumption is easily satisfied in this study (radius of the delaminated and buckled area, shown in Fig. 33.15a, is on the order of 100 nm). The compressive stress caused by indentation is given as [33.98]

$$\sigma_i = \frac{EV_i}{2\pi t a^2 (1-\nu)} , \qquad (33.3)$$

where ν and E are the Poisson's ratio and elastic modulus of the coating, V_i is the indentation volume, t is the coating thickness, and a is the crack radius. With increasing number of cycles, the indentation volume V_i increases. Therefore, the indentation compressive stress σ_i increases accordingly. In the second stage, buckling occurs during the unloading segment of the fatigue testing cycle when the sum of indentation compressive stress σ_i and the residual stress σ_r exceed the critical buckling stress σ_b for the delaminated circular section

as given by [33.99]

$$\sigma_b = \frac{\mu^2 E}{12\,(1-\nu^2)}\left(\frac{t}{a}\right)^2 , \qquad (33.4)$$

where the constant μ equals 42.67 for a circular clamped plate with a constrained center point and 14.68 when the center is unconstrained. The buckled coating acts as a cantilever. In this case, the indenter indents a cantilever rather than a coating-substrate system. This ultrathin coating cantilever has much less contact stiffness than the coating-substrate system. Therefore, the contact stiffness shows an abrupt decrease at N_f. In the third stage, with increasing number of cycles, the delaminated and buckled size increases, resulting in a further decrease in contact stiffness since the cantilever beam length increases. Conversely, a high bending stress acts at the edge of the buckled coating. The larger the buckled size, the higher is the bending stress. The cyclically bending stress causes fatigue damage at the end of the buckled coating, forming a ring-like crack. The coating under the indenter is separated from the bulk coating (caused by the ring-like crack at the edge of the buckled coating) and the substrate (caused by the delamination and buckling in the second stage). Therefore, the coating under the indenter is not constrained but is free to move with the indenter during fatigue testing. At this point, the sharp nature of the indenter is lost because the coating under the indenter gets stuck on the indenter. The indentation fatigue experiment results in the contact of a relatively huge blunt tip with the substrate. This results in a low contact stiffness value.

Compressive residual stresses result in delamination and buckling. A coating with higher adhesion strength and a less compressive residual stress is required for a higher fatigue life. Interfacial defects should be avoided in the coating deposition process. We know that the ring-like crack formation occurs in the coating. Formation of fatigue cracks in the coating depends upon the hardness and fracture toughness. Cracks are more difficult to form and propagate in the coating with higher strength and fracture toughness.

It is now accepted that long fatigue life in a coating-substrate almost always involves *living with cracks*, that the threshold or limit condition is associated with the nonpropagation of existing cracks or defects, even though these cracks may be undetectable [33.97]. For all coatings studied, at 4 μN, contact stiffness does not change much. This indicates that delamination and buckling did not occur within the number of cycles tested in this study. This is probably because the indentation-induced compressive stress was not high enough to allow the cracks to propagate and link up under the indenter or the sum of indentation compressive stress σ_i and the residual stress σ_r did not exceed the critical buckling stress σ_b.

Figure 33.10 and Table 33.5 summarize the hardness, elastic modulus, fracture toughness, and fatigue life of all coatings studied. A good correlation exists between fatigue life and other mechanical properties. Higher mechanical properties result in a longer fatigue life. The mechanical properties of DLC coatings are controlled by the sp^3-to-sp^2 ratio. The sp^3-bonded carbon exhibits the outstanding properties of diamond [33.52]. A higher deposition kinetic energy will result in a larger fraction of sp^3-bonded carbon in an amorphous network. Thus, the higher kinetic energy for the FCA could be responsible for its better carbon structure and higher mechanical properties [33.49–51, 100]. Higher adhesion strength between the FCA coating and substrate makes the FCA coating more difficult to delaminate from the substrate.

33.3.2 Microscratch and Microwear Studies

For microscratch studies, a conical diamond indenter (e.g., having a tip radius of about one micron and included angle of 60°) is drawn over the sample surface, and the load is increased steadily (typically 2–25 mN), until substantial damage occurs. The coefficient of friction is monitored during scratching. Scratch-induced damage of a coating, specifically fracture or delamination, can be monitored by in situ friction force measurements and by optical and SEM imaging of the scratches after tests. A gradual increase in friction is associated with plowing and an abrupt increase in friction is associated with fracture or catastrophic failure [33.101]. The load, corresponding to the abrupt increase in friction or an increase in friction above a certain value (typically 2× the initial value), provides a measure of scratch resistance or adhesive strength of a coating and is called *critical load*. Depth of scratches with increasing scratch length or normal load is measured using an AFM, typically with an area of 10 × 10 μm [33.2, 3, 49, 102].

Microscratch and microwear studies are also conducted using an AFM [33.23, 51, 100, 103, 104]. A square-pyramidal diamond tip (tip radius ≈ 100 nm) or a three-sided pyramidal diamond (Berkovich) tip, with an apex angle of 60° and a tip radius of about 100 nm, mounted on a platinum-coated, rectangular stainless steel cantilever of stiffness of about 40 N/m, is scanned orthogonal to the long axis of the cantilever to generate scratch and wear marks. During the scratch test, the normal load is either kept constant or is increased (typically from 0 to 100 μN), until damage occurs. Topography images of the scratch are obtained in situ with the AFM at a low load. By scanning the

sample during scratching, wear experiments can be conducted. Wear at a constant load is monitored as a function of number of cycles. Normal loads ranging from 10–80 μN are typically used.

Microscratch

Scratch tests conducted with a sharp diamond tip simulate a sharp asperity contact. In a scratch test, the cracking or delamination of a hard coating is signaled by a sudden increase in the coefficient of friction [33.23]. The load associated with this event is called the *critical load*.

Wu [33.105], *Bhushan* et al. [33.71], *Gupta* and *Bhushan* [33.12, 48], and *Li* and *Bhushan* [33.49, 50, 102] have used a nanoindenter to perform microscratch studies (mechanical durability) of various carbon coatings. The coefficient of friction profiles as a function of increasing normal load and AFM surface height maps of regions over scratches at the respective critical loads (indicated by the arrows in the friction profiles and AFM images), made on the various coatings of different thicknesses and the single-crystal silicon substrate using a conical tip are compared in Figs. 33.16 and 33.17. *Bhushan* and *Koinkar* [33.103], *Koinkar* and *Bhushan* [33.104], *Bhushan* [33.23], and *Sundararajan* and *Bhushan* [33.51, 100] have used an AFM to perform microscratch studies. Data for various coatings of different thicknesses and the silicon substrate using a Berkovich tip are compared in Figs. 33.18 and 33.19. Critical loads for various coatings tested using the nanoindenter and AFM are summarized in Fig. 33.20. The selected data for 20 nm thick coatings obtained using the nanoindenter, are also presented in Fig. 33.10 and Table 33.5.

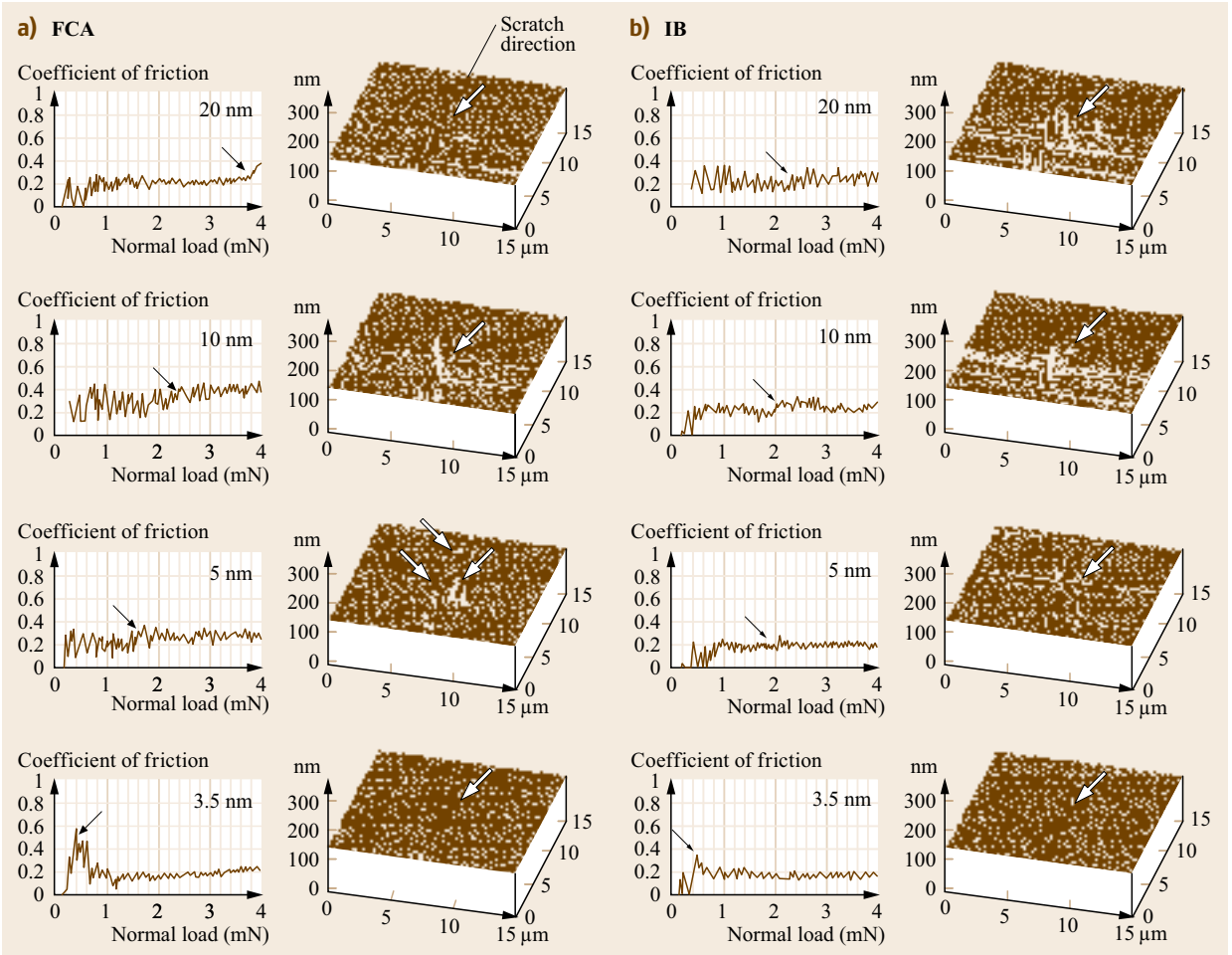


Fig. 33.16a–d Coefficient of friction profiles as a function of normal load and corresponding AFM surface height maps of regions over scratches at the respective critical loads (indicated by the *arrows* in the friction profiles and AFM images) for coatings of different thicknesses deposited by various deposition techniques: **(a)** FCA, **(b)** IB, **(c)** ECR-CVD, **(d)** SP

It can be seen that there exists a well-defined critical load for each coating. The AFM images clearly show that below the critical loads the coatings were plowed by the scratch tip, associated with the plastic flow of materials. At and after the critical loads, debris (chips) or buckling were observed on the sides of scratches. Delamination or buckling can be observed around or after the critical loads which suggests that the damage starts from delamination and buckling. For the 3.5 and 5 nm thick FCA coatings, before the critical loads small debris are observed on the sides of scratches. This suggests that the thinner FCA coatings are not so durable. It is obvious that for a given deposition method, the critical loads increase with increasing coating thickness. This indicates that the critical load is determined not only by the adhesion strength to the substrate but also by the coating thickness. We note that the debris generated on the thicker coatings are larger than that generated on the thinner coatings. For a thicker coating, which is more difficult to be broken, the broken coating chips (debris) are larger than those of the thinner coatings. The difference in the residual stresses of the coatings of different thicknesses could also affect debris size. The AFM image shows that the silicon substrate was damaged by plowing, associated with the plastic flow of materials. At and after the critical load, small and uniform debris are observed and the amount of debris increases with increasing normal load.

Since, at the critical load, the damage mechanism appears to be the onset of plowing, higher hardness and fracture toughness of a coating will therefore result in higher load required for deformation and hence higher critical load. Figure 33.21 shows critical loads of the various coatings, obtained with AFM tests, as a function of the coating hardness and fracture toughness (from Table 33.5). It can be seen that, in general, higher coating hardness and fracture toughness result

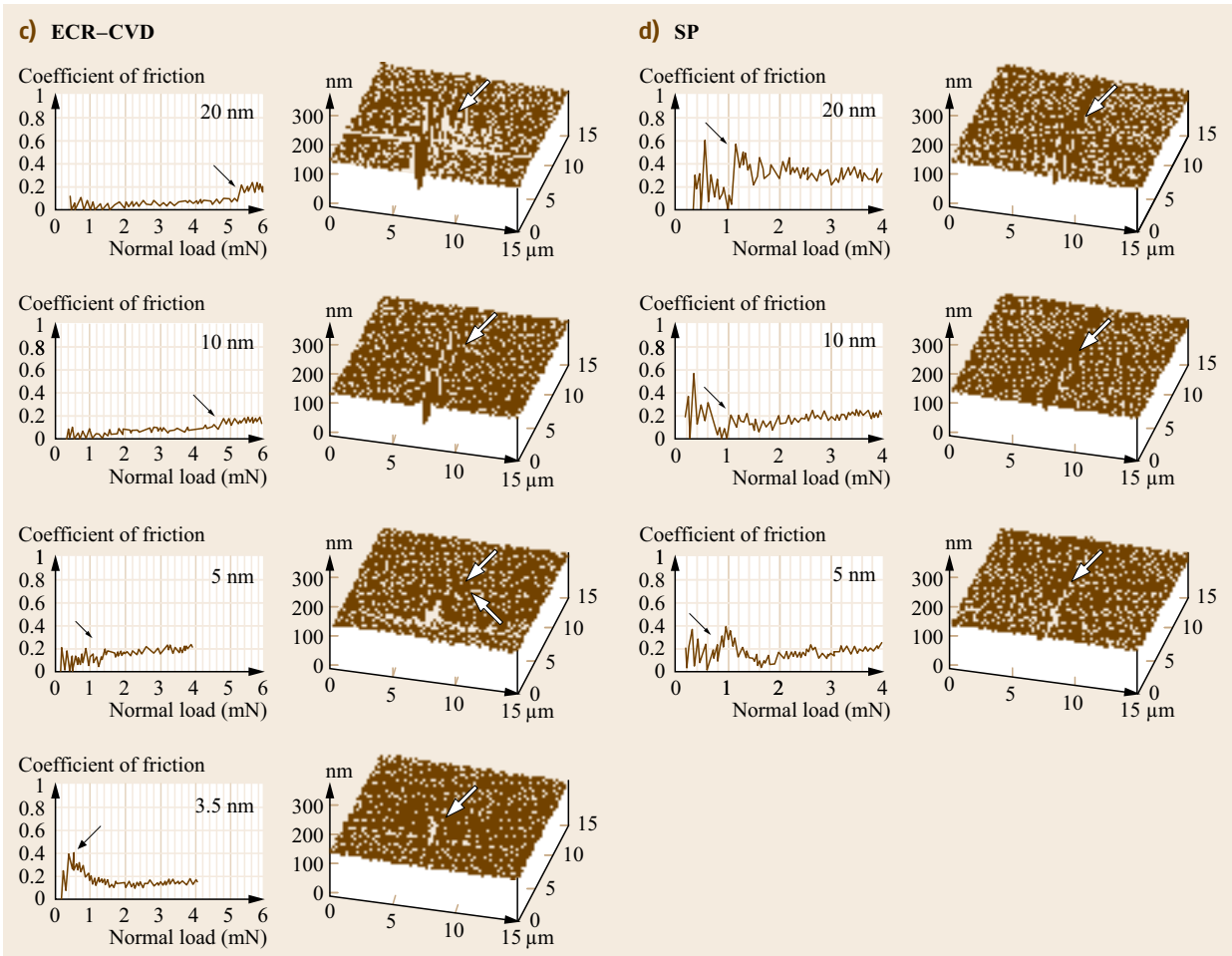


Fig. 33.16 (continued)

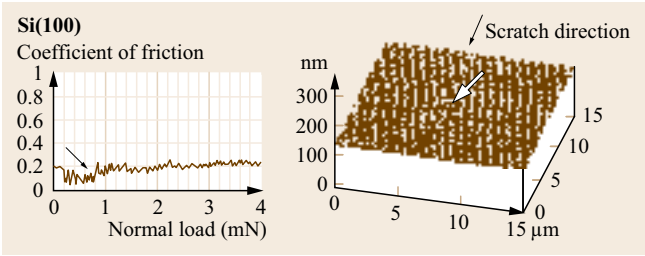


Fig. 33.17 Coefficient of friction profiles as a function of normal load and corresponding AFM surface height maps of regions over scratches at the respective critical loads (indicated by the *arrows* in the friction profiles and AFM images) for Si(100)

in higher critical load. The only exceptions are the FCA coatings at 5 and 3.5 nm coating thickness, which show the lowest critical loads despite their high hardness and fracture toughness. The brittleness of the thinner FCA coatings may be one reason for their low critical loads. The mechanical properties of coatings that are less than 10 nm thick are unknown. The FCA process may result in coatings with low hardness at such low thickness due to differences in coating stoichiometry and structure as compared to the coatings of higher thickness. Also, at these thicknesses stresses at the coating-substrate interface may affect adhesion and load-carrying capacity of the coatings.

On the basis of experimental results, a schematic of scratch damage mechanisms of the DLC coatings used in this study is shown in Fig. 33.22. Below the critical load, if a coating has a good combination of strength and fracture toughness, plowing associated with the plastic flow of materials is responsible for the damage of the coating (Fig. 33.22a). Whereas, if a coating has a lower fracture toughness, cracking could occur during plowing, associated with the formation of small debris (Fig. 33.22b). When normal load increases up to the critical load, delamination or buckling will occur at the coating-substrate interface (Fig. 33.22c). Further increase in normal load will result in breakdown of a coating via through-coating thickness cracking, as shown in Fig. 33.22d. Therefore, adhesion strength plays a crucial role in the determination of critical load. If a coating has stronger adhesive strength with the substrate, the coating is more difficult to be delaminated, which will result in a higher critical load. The interfacial and residual stresses of a coating could greatly affect the delamination and buckling [33.1]. The coating with higher interfacial and residual stresses is more easily delaminated and buckled, which will result in

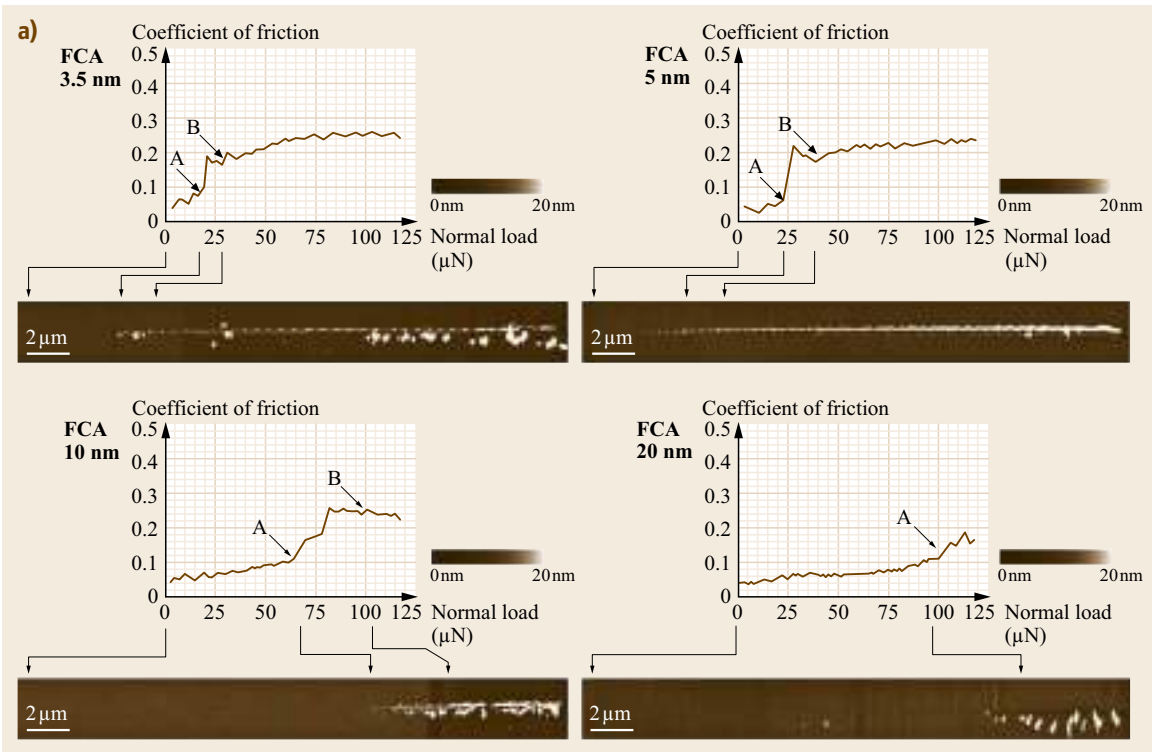


Fig. 33.18a–c Coefficient of friction profiles during scratch as a function of normal load and corresponding AFM surface height maps for **(a)** FCA, **(b)** ECR-CVD, and **(c)** SP coatings (after [33.100])

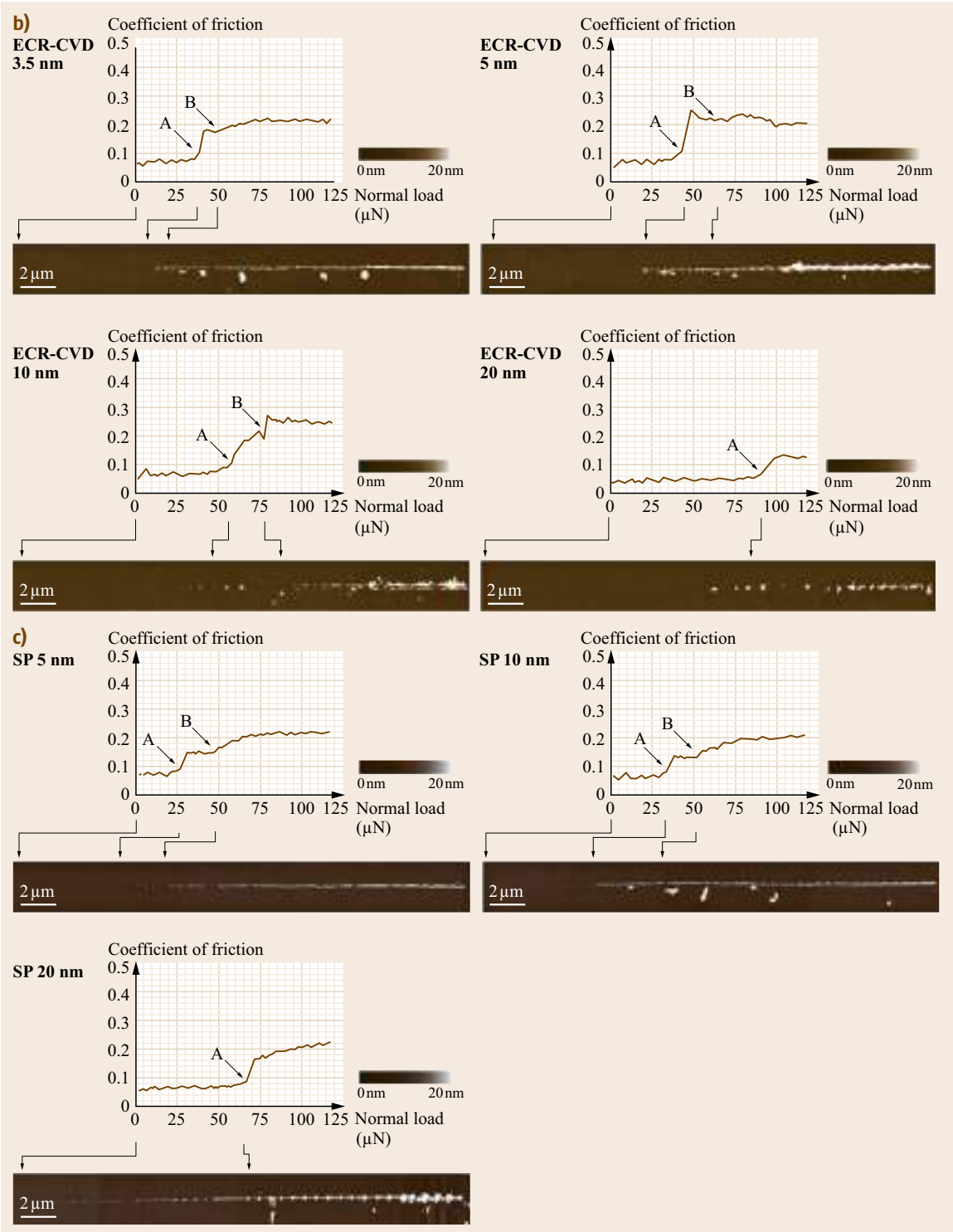


Fig. 33.18 (continued)

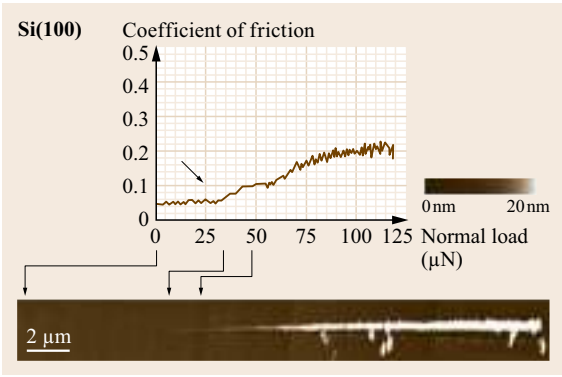


Fig. 33.19 Coefficient of friction profiles during scratch as a function of normal load and corresponding AFM surface height maps for Si(100) (after [33.100])

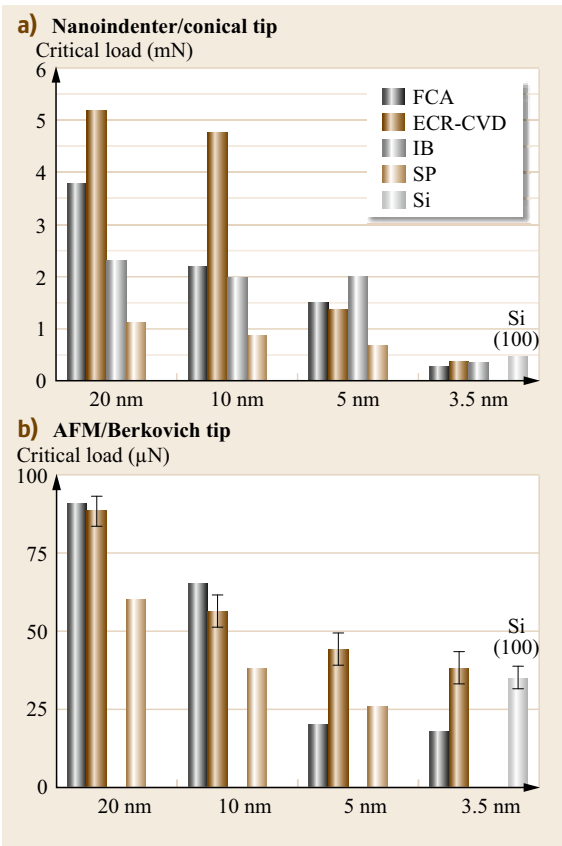


Fig. 33.20a,b Critical loads estimated from the coefficient of friction profiles from **(a)** nanoindenter and **(b)** AFM tests for various coatings of different thicknesses and the Si(100) substrate

a low critical load. It has been reported earlier that the FCA coatings have higher residual stresses than the other coatings [33.48]. Interfacial stresses play a more important role when a coating gets thinner. A large mismatch in elastic modulus between the FCA coatings and the silicon substrate may cause large interfacial stresses. This may be why the thinner FCA coatings show relatively low critical loads compared with the thicker FCA coatings, even though the FCA coatings have higher hardness and elastic modulus. The brittleness of the thinner FCA coatings may be another reason for the lower critical loads. The strength and fracture toughness of a coating also affect the critical load. Higher strength and fracture toughness will make the coating more difficult to be broken after delamination and buckling. The high scratch resistance/adhesion of FCA coatings is attributed to an atomic intermixing at the coating-substrate interface because of high kinetic energy (2 keV) plasma formed during the cathodic arc deposition process [33.58]. The atomic intermixing at the interface provides a graded compositional transi-

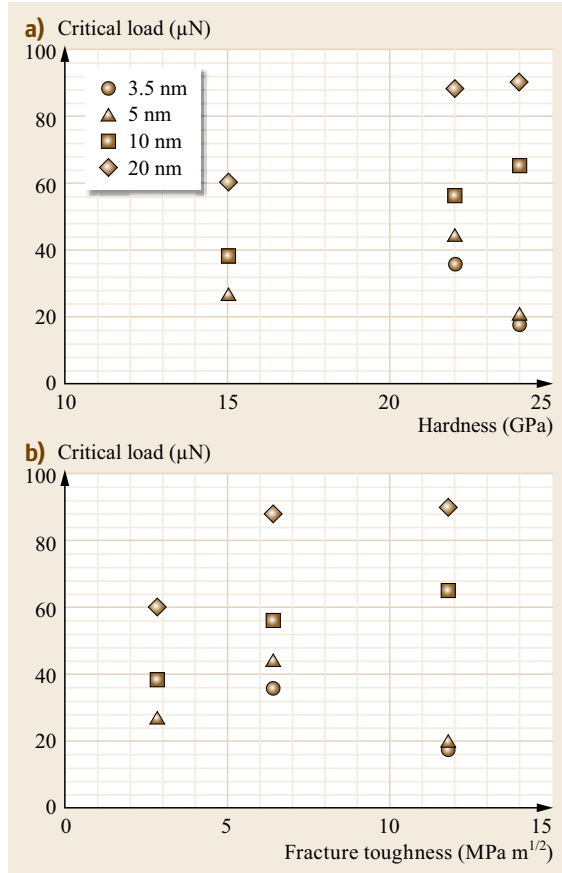


Fig. 33.21a,b Measured critical loads estimated from the coefficient of friction profiles from AFM tests as a function of **(a)** coating hardness and **(b)** fracture toughness. Coating hardness and fracture toughness values were obtained using a nanoindenter on 100 nm thick coatings (Table 33.5)

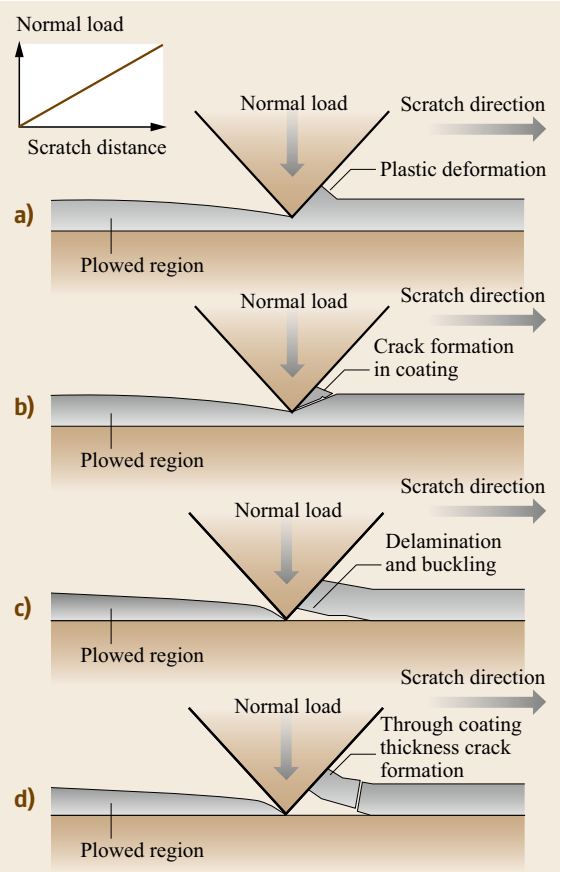


Fig. 33.22a–d Schematic of scratch damage mechanisms of the DLC coatings, **(a)** plowing associated with the plastic flow of materials, **(b)** plowing associated with the formation of small debris, **(c)** delamination and buckling at the critical load, and **(d)** breakdown via through-coating thickness cracking at and after the critical load (after [33.49])

tion between the coating and the substrate materials. In all other coatings used in this study, the kinetic energy of the plasma was insufficient for atomic intermixing.

Gupta and *Bhushan* [33.12, 48] and *Li* and *Bhushan* [33.49, 50] measured scratch resistance of DLC coatings deposited on Al_2O_3-TiC, Ni-Zn ferrite, and single-crystal silicon substrates. For good adhesion of a DLC coating to other substrates, in most cases, an interlayer of silicon is required except for cathodic arc deposited coatings. The best adhesion with cathodic arc carbon coating is obtained on electrically conducting substrates such as Al_2O_3-TiC and silicon as compared to Ni-Zn ferrite.

Microwear

Microwear studies can be conducted using an AFM [33.23]. For microwear studies, a three-sided pyramidal single-crystal natural diamond tip with an apex angle of about 80° and a tip radius of about 100 nm is used at relatively high loads of 1–150 μN. The diamond tip is mounted on a stainless steel cantilever beam with a normal stiffness of about 30 N/m. The sample is generally scanned in a direction orthogonal to the long axis of the cantilever beam (typically at a rate of 0.5 Hz). The tip is mounted on the beam such that one of its edges is orthogonal to the beam axis. For wear studies, typically an area of 2 μm × 2 μm is scanned for a selected number of cycles.

Microwear studies of various types of DLC coatings have been conducted [33.51, 103, 104]. Figure 33.23a shows a wear mark on uncoated Si(100). Wear occurs uniformly and material is removed layer by layer via plowing from the first cycle, resulting in the constant friction force seen during the wear (Fig. 33.24a). Figure 33.23b shows AFM images of the wear marks on all 10 nm coatings. It is seen that coatings wear nonuniformly. Coating failure is sudden and accompanied by a sudden rise in the friction force (Fig. 33.24b). Figure 33.24 shows the wear depth of the Si(100) substrate and various DLC coatings at two different loads. FCA and ECR-CVD, 20 nm thick coatings show excellent wear resistance up to 80 μN, the load that is required for the IB 20 nm coating to fail. In these tests, *failure* of a coating results when the wear depth exceeds the quoted coating thickness. The SP 20 nm coating fails at the much lower load of 35 μN. At 60 μN, the coating hardly provides any protection. Moving onto the 10 nm coatings, ECR-CVD coating requires about 45 cycles at 60 μN to fail as compared to IB and FCA which fail at 45 μN. The FCA coating exhibits slight roughening in the wear track after the first few cycles, which leads to an increase in the friction force. The SP coating continues to exhibit poor resistance, failing at 20 μN. For the 5 nm coatings, the load required for the coatings to fail continue to decrease. But IB and ECR-CVD still provide adequate protection as compared to bare Si(100) in that order, failing at 35 μN compared to FCA at 25 μN and SP at 20 μN. Almost all the 20, 10, and 5 nm coatings provide better wear resistance than bare silicon. At 3.5 nm, the FCA coating provides no wear resistance, failing almost instantly at 20 μN. The IB and ECR-CVD coating show good wear resistance at 20 μN compared to bare Si(100). But IB lasts only about 10 cycles and ECR-CVD about 3 cycles at 25 μN.

The wear tests highlight the differences in the coatings more vividly than the scratch tests. At higher thicknesses (20 and 10 nm), the ECR-CVD and FCA coating appear to show the best wear resistance. This is probably due to higher hardness of the coatings (see Table 33.5). At 5 nm, the IB coating appears to be the

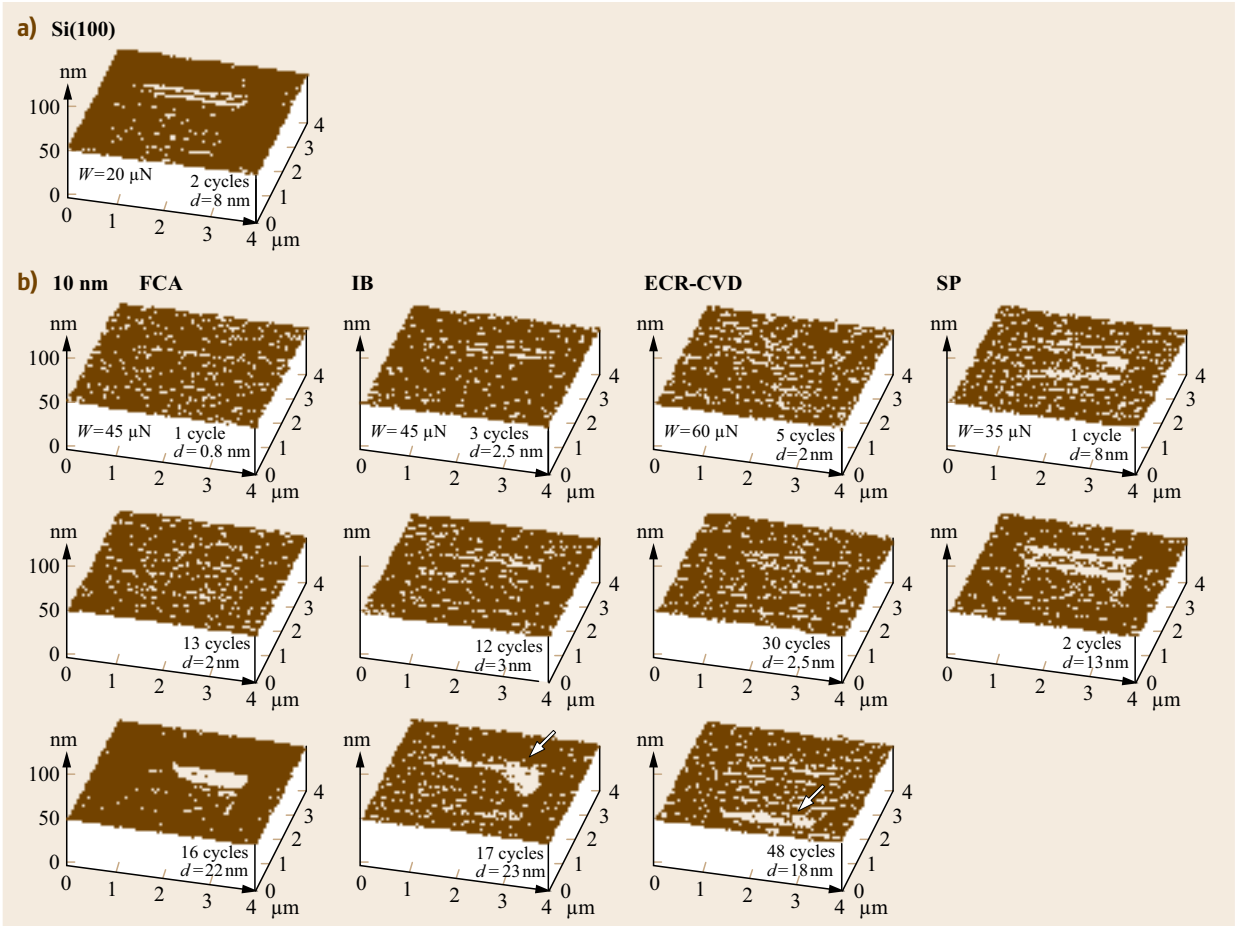


Fig. 33.23a,b AFM images of wear marks on **(a)** bare Si(100), and **(b)** all 10 nm thick DLC coatings (after [33.51])

best. FCA coatings show poorer wear resistance with decreasing coating thickness. This suggests that the trends in hardness seen in Table 33.5 no longer hold at low thicknesses. SP coatings showed consistently poor wear resistance at all thicknesses. The IB 3.5 nm coating does provide reasonable wear protection at low loads.

33.3.3 Macroscale Tribological Characterization

So far data on mechanical characterization and microscratch and microwear studies using a nanoindenter and an AFM have been presented. Mechanical properties affect tribological performance of the coatings and microwear studies simulate a single asperity contact which helps in the development of fundamental understanding of the wear process. These studies are useful in screening various candidates as well as to understand the relationships between deposition conditions and properties of various samples. As a next step, macroscale friction and wear tests need to be conducted to measure tribological performance of the coatings.

Macroscale accelerated friction and wear tests to screen a large number of candidates and functional tests on selected candidates have been conducted. An accelerated test is designed to accelerate the wear process such that it does not change the failure mechanism. The accelerated friction and wear tests are generally conducted using a ball-on-flat tribometer under reciprocating motion [33.71]. Typically, a diamond tip with 20 μm tip radius or a sapphire ball with 3 mm diameter and surface finish of about 2 nm RMS is slid against the coated substrates at selected loads. Coefficient of friction is monitored during the tests.

Functional tests are conducted using an actual machine under close to actual operating conditions for which coatings are developed. Generally, the tests are accelerated somewhat to fail the interface in a short time.

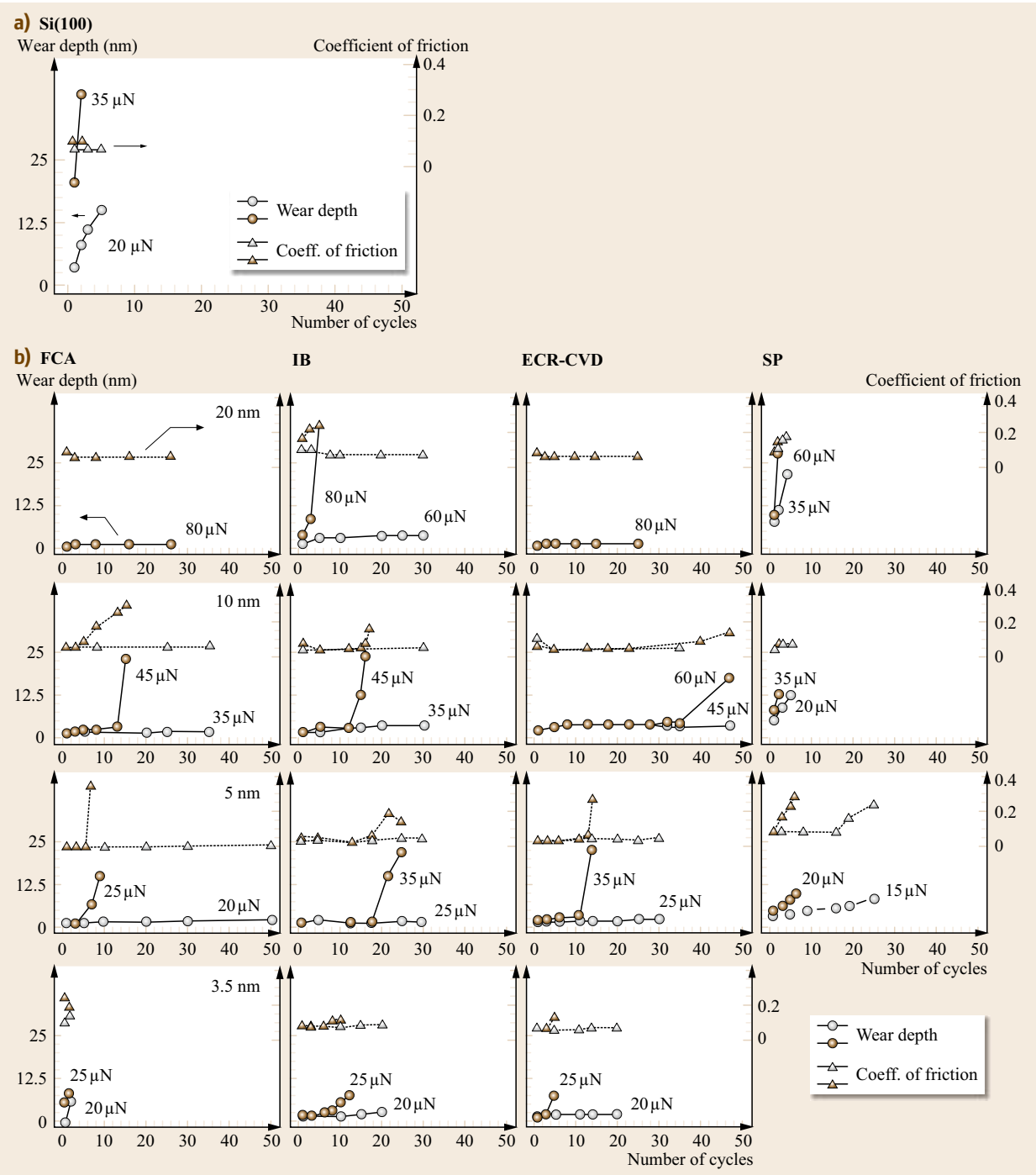


Fig. 33.24a,b Wear data of **(a)** bare Si(100) and **(b)** all DLC coatings. Coating thickness is constant along each row in **(b)**. Both wear depth and coefficient of friction during wear for a given cycle are plotted (after [33.51])

Accelerated Friction and Wear Tests

Li and *Bhushan* [33.49] conducted accelerated friction and wear tests on DLC coatings deposited by various deposition techniques, using a ball-on-flat tribometer. Average values of coefficient of friction are presented in Table 33.5. The optical micrographs of wear tracks and

debris formed on all samples when slid against a sapphire ball after sliding distance of 5 m are presented in Fig. 33.25. The normal load used for the 20 and 10 nm thick coatings is 200 mN and the normal load used for the 5 and 3.5 nm thick coatings and silicon substrate is 150 mN.

Among the 20 nm thick coatings, the SP coating exhibits a higher coefficient of friction of about 0.3 than the other coatings which have comparable values of coefficient of friction of about 0.2. The optical micrographs show that the SP coating has a larger wear track and more debris than the IB coatings. No wear track and debris were found on the 20 nm thick FCA and ECR-CVD coatings. The optical micrographs of 10 nm thick coatings show that the SP coating was severely damaged showing a large wear track with scratches and lots of debris. The FCA and ECR-CVD coatings show smaller wear tracks and less debris than the IB coatings.

For the 5 nm thick coatings, the wear tracks and debris of the IB and ECR-CVD coatings are comparable. The bad wear resistance of the 5 nm thick FCA coating is in good agreement with the low scratch critical load, which may be due to the higher interfacial and residual stresses as well as brittleness of the coating.

At 3.5 nm, all coatings exhibit wear. The FCA coating provides no wear resistance, failing instantly like the silicon substrate. Large block-like debris is observed on the sides of the wear track of the FCA coating. This indicates that large region delamination and buckling occurred during sliding, resulting in large block-like debris. These block-like debris, in turn, scratched the coating, making the coating damage even more severe. The IB and ECR-CVD coatings are able to provide some protection against wear at 3.5 nm.

In order to better evaluate the wear resistance of various coatings, based on the optical examination of the wear tracks and debris after tests, the wear damage index bar chart of the various coatings of different thicknesses and an uncoated silicon substrate is presented in Fig. 33.26. Among the 20 and 10 nm thick coatings, the SP coatings show the worst damage, followed by IB and FCA/ECR-CVD. At 5 nm, the FCA and SP coatings show the worst damage, followed by IB, and ECR-CVD coatings. All 3.5 nm thick coatings show the same heavy damage as the uncoated silicon substrate.

The wear damage mechanisms of thick and thin DLC coatings studied are believed to be as illustrated in Fig. 33.27. The deformation zone and Herztian and wear fatigue cracks formed beneath the surface during the early stages of sliding extend within the coating upon subsequent sliding [33.1]. Formation of fatigue cracks depend upon the hardness and subsequent cycles. These are controlled by the sp^3-to-sp^2 ratio. For thicker coatings, the cracks generally do not penetrate the coating. While for a thinner coating, the cracks easily propagate down to the interface with the aid of the interfacial stresses, and they get diverted along the interface just enough to cause local delamination of the coating. When this happens, the coating experiences excessive plowing. At this point, the coating fails catastrophically, resulting in a sudden rise in the coeffi-

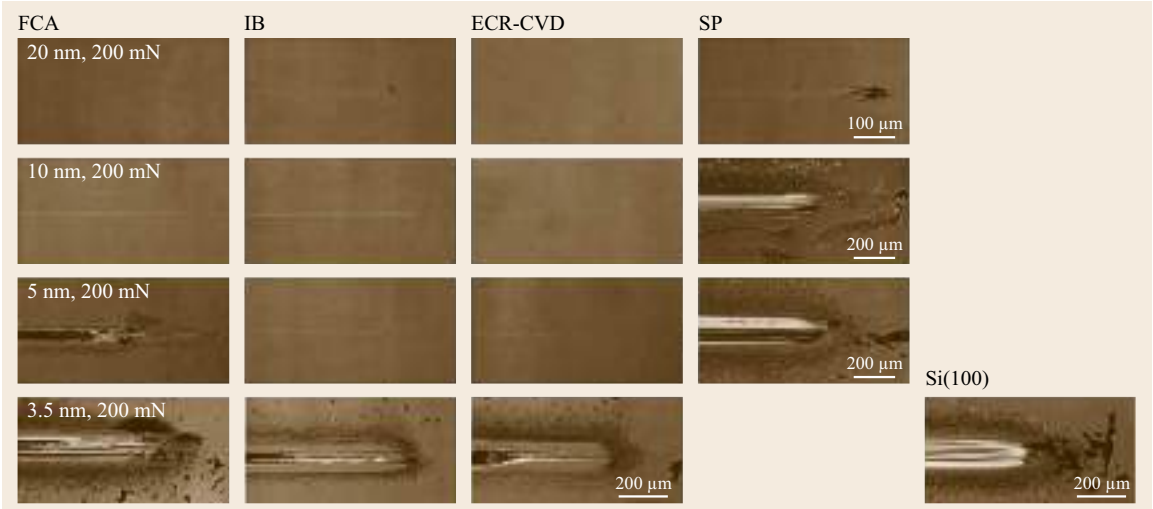


Fig. 33.25 Optical micrographs of wear tracks and debris formed on the various coatings of different thicknesses and silicon substrate when slid against a sapphire ball after sliding distance of 5 m. The end of the wear track is on the right hand side of the image

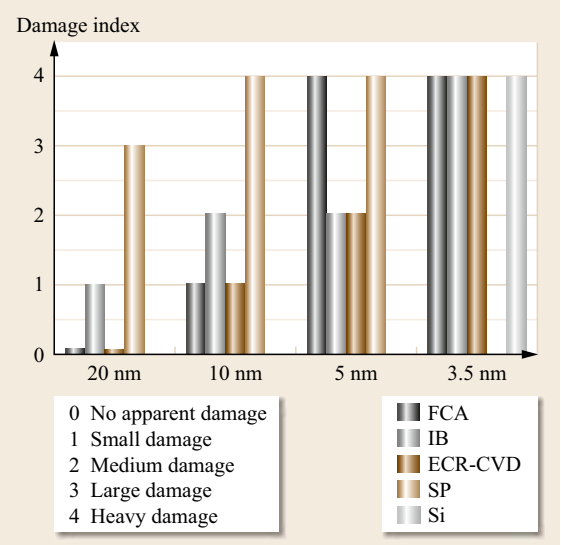


Fig. 33.26 Wear damage index bar chart of the various coatings of different thicknesses and Si(100) substrate based on optical examination of the wear tracks and debris

cient of friction. All 3.5 nm thick coatings failed much quicker as compared to the thicker coatings. It appears that these thin coatings have very low load-carrying capacity and therefore the substrate undergoes deformation almost immediately. This generates stresses at the interface that weaken coating adhesion and lead to delamination of the coating. Another reason may be that the thickness is insufficient to produce a coating comprised of the DLC structure. Instead the bulk may be made up of a matrix characteristic of the interface region where atomic mixing occurs with the substrate and/or any interlayer used. This would also result in poor wear resistance and silicon-like behavior of the coating, especially in the case of FCA coatings, which show the worst performance at 3.5 nm. In contrast to the other coatings, all SP coatings show the worst wear performance at any thickness (Fig. 33.25). This may be due to their poor mechanical properties, such as lower hardness and scratch resistance compared to the other coatings.

Comparison of Figs. 33.20 and 33.26 shows a very good correlation between wear damage and scratch critical loads. Less wear damage corresponds to higher scratch critical load. Based on the data, thicker coatings do show better scratch and wear resistance than thinner coatings. This is probably due to better load-carrying capacity of the thick coatings as compared to the thinner ones. For a given coating thickness, higher hardness and fracture toughness and better adhesion strength are believed to be responsible for the superior wear performance.

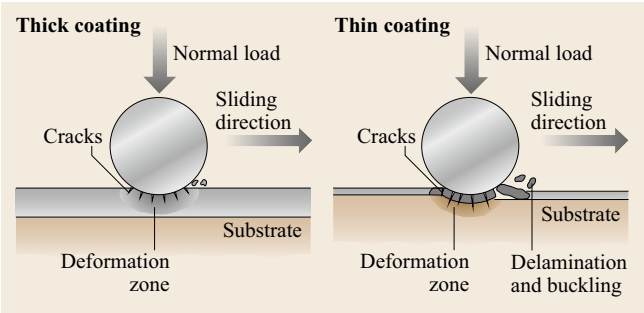


Fig. 33.27 Schematic of wear damage mechanisms of thick and thin DLC coatings (after [33.49])

Effect of Environment

Friction and wear performance of amorphous carbon coatings are found to be strongly dependent on the water vapor content and partial gas pressure in the test environment. Friction data for an amorphous carbon film on a silicon substrate sliding against steel are presented as a function of the partial pressure of water vapor in Fig. 33.28 [33.1, 13, 70, 106, 107]. Friction increases dramatically above a relative humidity of about 40%. At high relative humidity, condensed water vapor forms meniscus bridges at the contacting asperities and the menisci result in an intrinsic attractive force

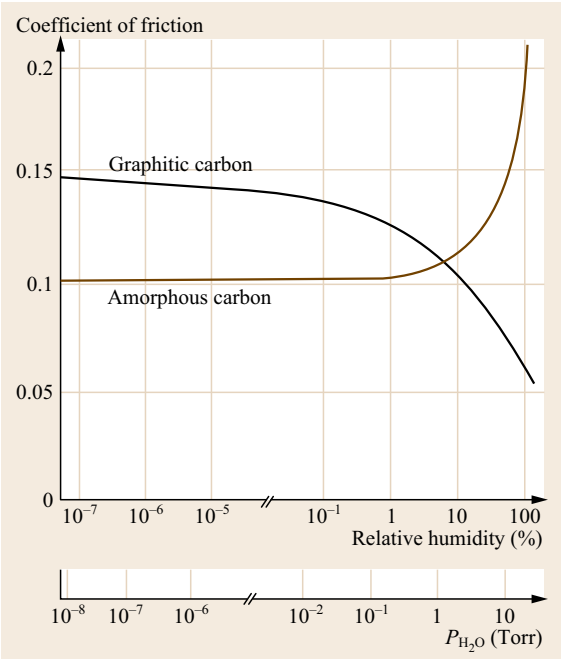


Fig. 33.28 Coefficient of friction as a function of relative humidity and water vapor partial pressure of a RF-plasma-deposited amorphous carbon coating and a bulk graphitic carbon coating sliding against a steel ball

which is responsible for an increase in friction. For completeness, the data for the coefficient of friction of bulk graphitic carbon are also presented in Fig. 33.28. Note that friction decreases with an increase in relative humidity [33.108]. Graphitic carbon has a layered lattice crystal structure. Graphite absorbs polar gases (H_2O, O_2, CO_2, NH_3, etc.) at the edges of the crystallites, which weakens the interlayer bonding forces facilitating interlayer slip and results in lower friction [33.1].

To better study the effect of environment for carbon-coated magnetic disks, a number of tests have been conducted in controlled environments. *Marchon* et al. [33.109] conducted tests in alternating environments of oxygen and nitrogen gases (Fig. 33.29). The coefficient of friction increases as soon as oxygen is added to the test environment whereas in a nitrogen environment the coefficient of friction decreases slightly. Tribochemical oxidation of the DLC coating in the oxidizing environment is responsible for an increase in the coefficient of friction implying wear. *Dugger* et al. [33.110], *Strom* et al. [33.111], *Bhushan* and *Ruan* [33.112], and *Bhushan* et al. [33.72] conducted tests with DLC-coated magnetic disks (with a circa 2 nm thick perfluoropolyether lubricant film) in contact with Al_2O_3-TiC sliders, in different gaseous environments including a high vacuum of 2×10^{-7} Torr (Fig. 33.30). The wear lives are shortest in a high vacuum and the longest in most atmospheres of nitrogen and argon with the following order (from best to worst): argon or nitrogen, Ar + H_2O, ambient air, Ar + O_2, and Ar + H_2O, vacuum. From this sequence of wear performance, we can see that having oxygen and water in an operating environment worsen the wear performance of the coatings, but having nothing in it (i. e., a vacuum) is the worst of all. Indeed, failure mechanisms differ in various environments. In high vacuum, intimate contact between disk and slider surfaces results in significant wear. In ambient air, Ar + O_2, and Ar + H_2O, tribochemical oxidation of the carbon overcoat is responsible for interface failure. For experiments performed in pure argon and nitrogen, mechanical shearing of the asperities cause the formation of debris which is responsible for the formation of scratch marks on the carbon surface as could be observed with an optical microscope [33.72].

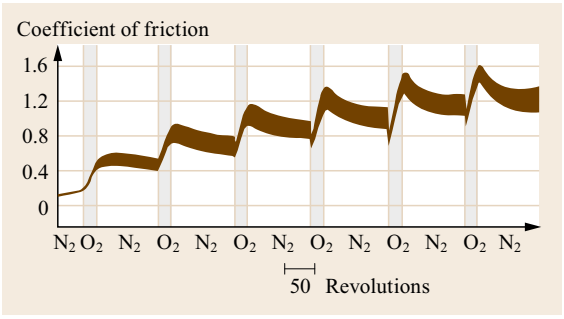


Fig. 33.29 Coefficient of friction as a function of sliding distance for a ceramic slider against a magnetic disk coated with 20 nm thick DC magnetron sputtered DLC coating, measured at a speed of 0.06 m/s and 10 g load. The environment is alternated between oxygen and nitrogen gases (after [33.109])

Functional Tests

Magnetic thin-film heads made with Al_2O_3-TiC substrate are used in magnetic storage applications [33.13]. The multilayered thin-film pole-tip structure present on the head surface wears more rapidly than the Al_2O_3-TiC substrate which is much harder. Pole-tip recession (PTR) is a serious concern in magnetic storage [33.16–20, 113]. Two of the diamond-like carbon coatings superior in mechanical properties – ion beam and cathodic arc carbon – were deposited on the air bearing surfaces of Al_2O_3-TiC head sliders [33.16]. The functional tests were conducted by running a metal-particle MP tape in a computer tape drive. Average PTR as a function of sliding distance data are presented in Fig. 33.31. We note that PTR increases for the uncoated head, whereas for the coated heads there is a slight increase in PTR in early sliding followed by little change. Thus, coatings provide protection.

The micromechanical and accelerated and functional tribological data presented here clearly suggest

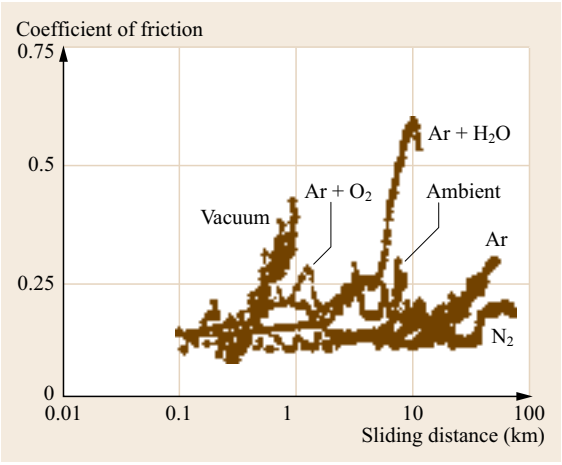


Fig. 33.30 Durability measured by sliding an Al_2O_3-TiC magnetic slider against a magnetic disk coated with a 20 nm thick DC sputtered amorphous carbon coating and a 2 nm thick perfluoropolyether film, measured at a speed of 0.75 m/s and 10 g load. Vacuum refers to 2×10^{-7} Torr (after [33.72])

that there is a good correlation between the scratch resistance and wear resistance measured using accelerated tests and functional tests. Thus, scratch tests can be successfully used to screen coatings for wear applications.

33.3.4 Coating Continuity Analysis

Ultrathin coatings less than 10 nm thick, may not uniformly coat the sample surface. In other words, the coating may be discontinuous and deposited in the form of islands on the microscale. A possible reason for poor wear protection and the nonuniform failure of the coatings may be due to poor coverage of the thin coatings on the substrate. Coating continuity can be studied by using surface analytical techniques such as Auger and/or XPS analyses. Any discontinuity with coating thickness less than sampling depth of the instrument will detect locally the substrate species [33.50, 51, 103].

The results of XPS analysis on various coatings, deposited on Si(100) substrates, over a 1.3 mm^2 region (single point measurement with spot diameter 1300 μm) are shown in Fig. 33.32. The sampling depth is about 2–3 nm. The poor SP coatings show much less carbon content (< 75% atomic concentration) as do the poor 5 nm and 3.5 nm FCA coatings (< 60%) as compared to the IB and ECR-CVD coatings. Silicon is detected in all 5 nm coatings. From the data it is hard to say if the Si is from the substrate or from exposed regions due to coating discontinuity. Based on the sampling depth any Si detected in 3.5 nm coatings would likely be from the substrate. The other interesting observation is that all poor coatings (all SP and FCA 5 and 3.5 nm) show almost twice the oxygen content compared to the other coatings. Any oxygen present may be due to leaks in the deposition chamber and is present in the form of silicon oxides.

AES (Auger electron spectroscopy) measurements, averaged over a scan area of 900 μm^2, were conducted

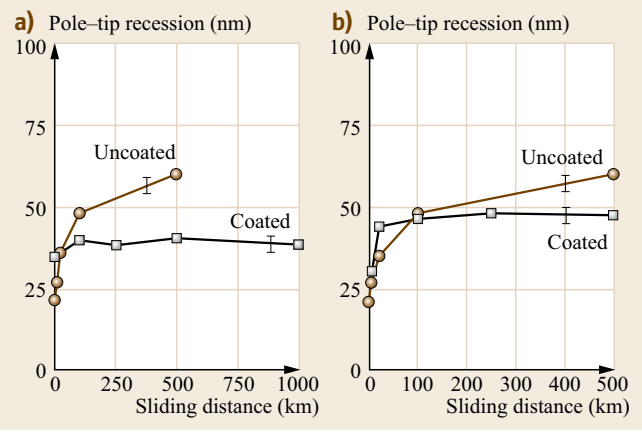


Fig. 33.31a,b Pole-tip recession as a function of sliding distance as measured with an AFM for **(a)** uncoated and 20 nm thick ion-beam carbon-coated, and **(b)** uncoated and 20 nm thick cathodic arc carbon-coated Al_2O_3-TiC heads run against MP tapes (after [33.16])

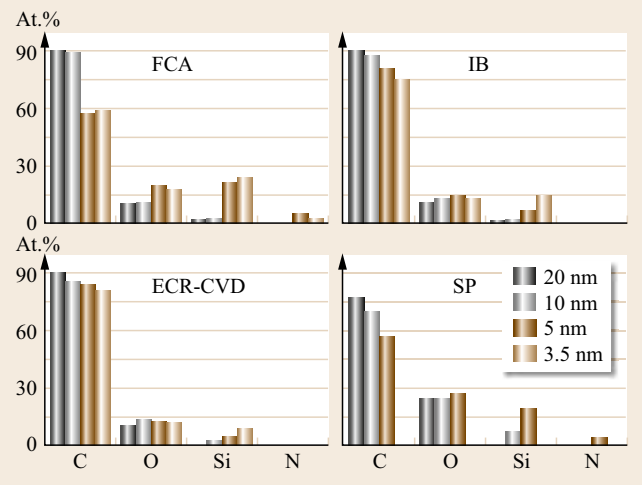


Fig. 33.32 Quantified XPS data for various DLC coatings on Si(100) substrate (after [33.51]). Atomic concentrations are shown

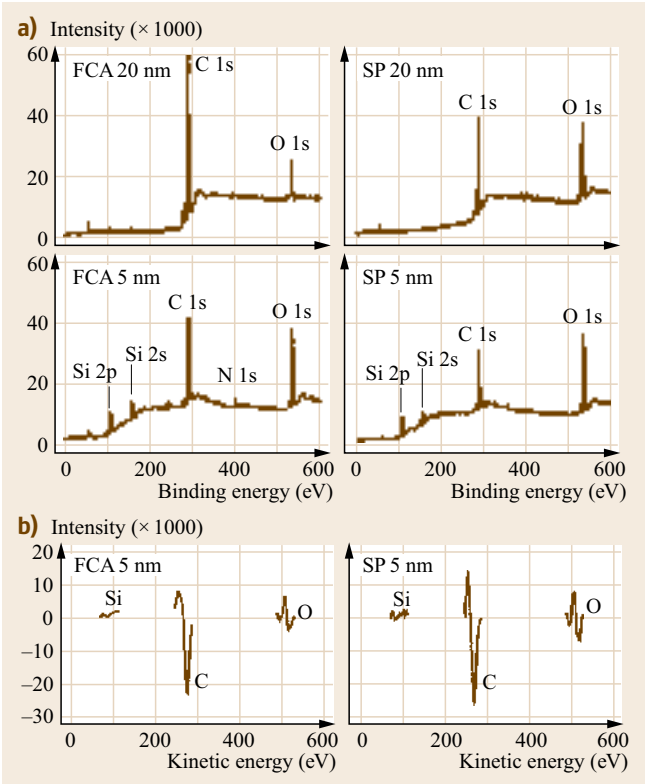


Fig. 33.33 **(a)** XPS spectra for FCA and SP coatings of 5 and 20 nm coating thicknesses on Si(100) substrate, and **(b)** AES spectra for FCA and SP coatings at 5 nm thickness on Si(100) substrate (after [33.51])

on FCA and SP 5 nm coatings at six different regions on each sample. Very little silicon was detected on this scale and the detected peaks were characteristic of oxides. The oxygen levels were comparable to that seen for the good coatings by XPS. These contrasts to the XPS measurements at a larger scale suggest that the coating possesses discontinuities only at isolated areas and that the 5 nm coatings are generally continuous on the microscale. Figure 33.33 shows representative XPS and AES spectra of selected samples.

33.4 Conclusion

Diamond material and its smooth coatings exhibit relatively low wear and friction. Major limitations of true diamond coatings are that they need to be deposited at high temperatures, can only be deposited on selected substrates, and require surface finishing. Mechanical, thermal, and optical properties of hard amorphous carbon (a-C), commonly known as DLC coatings, can be tailored to be similar to diamond. These can be deposited with a large range of thicknesses using a variety of deposition processes on a variety of substrates at or near room temperature. The coatings reproduce substrate topography, avoiding the need for postfinishing. Friction and wear properties of some DLC coatings can be very attractive for tribological applications. The largest industrial application of these coatings is in magnetic storage devices. They are also being used in MEMS/NEMS.

EELS and Raman spectroscopies can be used successfully for chemical characterization of amorphous carbon coatings. The prevailing atomic arrangement in DLC coatings is amorphous or quasiamorphous with small diamond (sp^3), graphite (sp^2), and other unidentifiable micro- or nanocrystallites. Most DLC coatings, except those produced by filtered cathodic arc, contain from a few to about 50 at.% hydrogen. Hydrogen is sometimes deliberately incorporated in the sputtered and ion-plated coatings to tailor their properties.

Amorphous carbon coatings deposited by various techniques exhibit different mechanical and tribological properties. The nanoindenter can be used successfully for measurement of hardness, elastic modulus, fracture toughness, and fatigue life. Microscratch and microwear experiments can be performed using either

a nanoindenter or an AFM. Thin coatings deposited by filtered cathodic arc, ion beam, and ECR-CVD hold promise for tribological applications. Coatings as thin as 5 nm or even thinner provide wear protection. Microscratch, microwear, and accelerated wear testing, if simulated properly, can be used successfully to screen coating candidates for industrial applications. In the examples shown in this chapter, trends observed in the microscratch, microwear, and accelerated macrofriction wear tests are similar to those found in functional tests.

References

33.1 B. Bhushan, B.K. Gupta: *Handbook of Tribology: Materials, Coatings, and Surface Treatments* (McGraw Hill, New York 1991), reprint ed. (Krieger, Malabar 1997)

33.2 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)

33.3 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

33.4 B. Bhushan, B.K. Gupta, G.W. VanCleef, C. Capp, J.V. Coe: Fullerene (C_{60}) films for solid lubrication, Tribol. Trans. **36**, 573–580 (1993)

33.5 B.K. Gupta, B. Bhushan, C. Capp, J.V. Coe: Material characterization and effect of purity and ion implantation on the friction and wear of sublimed fullerene films, J. Mater. Res. **9**, 2823–2838 (1994)

33.6 B.K. Gupta, B. Bhushan: Fullerene particles as an additive to liquid lubricants and greases for low friction and wear, Lubr. Eng. **50**, 524–528 (1994)

33.7 B. Bhushan, V.V. Subramaniam, A. Malshe, B.K. Gupta, J. Ruan: Tribological properties of polished diamond films, J. Appl. Phys. **74**, 4174–4180 (1993)

33.8 B. Bhushan, B.K. Gupta, V.V. Subramaniam: Polishing of diamond films, Diam. Films Technol. **4**, 71–97 (1994)

33.9 P. Sander, U. Kaiser, M. Altebockwinkel, L. Wiedmann, A. Benninghoven, R.E. Sah, P. Koidl: Depth profile analysis of hydrogenated carbon layers on silicon by x-ray photoelectron spectroscopy, auger electron spectroscopy, electron energy-loss spectroscopy, and secondary ion mass spectrometry, J. Vac. Sci. Technol. A **5**, 1470–1473 (1987)

33.10 A. Matthews, S.S. Eskildsen: Engineering applications for diamond-like carbon, Diam. Relat. Mater. **3**, 902–911 (1994)

33.11 A.H. Lettington: Applications of diamond-like carbon thin films, Carbon **36**, 555–560 (1998)

33.12 B.K. Gupta, B. Bhushan: Mechanical and tribological properties of hard carbon coatings for magnetic recording heads, Wear **190**, 110–122 (1995)

33.13 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, Berlin, Heidelberg 1996)

33.14 B. Bhushan: *Mechanics and Reliability of Flexible Magnetic Media*, 2nd edn. (Springer, New York 2000)

33.15 B. Bhushan: Macro- and microtribology of magnetic storage devices. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 1413–1513

33.16 B. Bhushan, S.T. Patton, R. Sundaram, S. Dey: Pole tip recession studies of hard carbon-coated thin-film tape heads, J. Appl. Phys. **79**, 5916–5918 (1996)

33.17 J. Xu, B. Bhushan: Pole tip recession studies of thin-film rigid disk head sliders II: Effects of air bearing surface and pole tip region designs and carbon coating, Wear **219**, 30–41 (1998)

33.18 W.W. Scott, B. Bhushan: Corrosion and wear studies of uncoated and ultra-thin DLC coated magnetic tape-write heads and magnetic tapes, Wear **243**, 31–42 (2000)

33.19 W.W. Scott, B. Bhushan: Loose debris and head stain generation and pole tip recession in modern tape drives, J. Inf. Storage Proc. Syst. **2**, 221–254 (2000)

33.20 W.W. Scott, B. Bhushan, A.V. Lakshmikumaran: Ultrathin diamond-like carbon coatings used for reduction of pole tip recession in magnetic tape heads, J. Appl. Phys. **87**, 6182–6184 (2000)

33.21 B. Bhushan: Nanotribology and nanomechanics of MEMS devices. In: *Proc. 9th Annu. Workshop Micro Electro Mech. Syst.* (IEEE, New York 1996) pp. 91–98

33.22 B. Bhushan (Ed.): *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

33.23 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

33.24 B. Bhushan: Macro- and microtribology of MEMS materials. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 1515–1548

33.25 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Switzerland 2016)

33.26 S. Aisenberg, R. Chabot: Ion beam deposition of thin films of diamond-like carbon, J. Appl. Phys. **49**, 2953–2958 (1971)

33.27 E.G. Spencer, P.H. Schmidt, D.C. Joy, F.J. Sansalone: Ion beam deposited polycrystalline diamond-like films, Appl. Phys. Lett. **29**, 118–120 (1976)

33.28 A. Grill, B.S. Meyerson: Development and status of diamondlike carbon. In: *Synthetic Diamond: Emerging CVD Science and Technology*, ed. by K.E. Spear, J.P. Dismukes (Wiley, New York 1994) pp. 91–141

33.29 Y. Catherine: Preparation techniques for diamond-like carbon. In: *Diamond and Diamond-Like Films and Coatings*, ed. by R.E. Clausing, L.L. Horton, J.C. Angus, P. Koidl (Plenum, New York 1991) pp. 193–227

33.30 J.J. Cuomo, D.L. Pappas, J. Bruley, J.P. Doyle, K.L. Seagner: Vapor deposition processes for amorphous carbon films with sp^3 fractions approaching diamond, J. Appl. Phys. **70**, 1706–1711 (1991)

33.31 J.C. Angus, C.C. Hayman: Low pressure metastable growth of diamond and diamondlike phase, Science **241**, 913–921 (1988)

33.32 J.C. Angus, F. Jensen: Dense diamondlike hydrocarbons as random covalent networks, J. Vac. Sci. Technol. A **6**, 1778–1782 (1988)

33.33 D.C. Green, D.R. McKenzie, P.B. Lukins: The microstructure of carbon thin films, Mater. Sci. Forum **52/53**, 103–124 (1989)

33.34 N.H. Cho, K.M. Krishnan, D.K. Veirs, M.D. Rubin, C.B. Hopper, B. Bhushan, D.B. Bogy: Chemical structure and physical properties of diamond-like amorphous carbon films prepared by magnetron sputtering, J. Mater. Res. **5**, 2543–2554 (1990)

33.35 J.C. Angus: Diamond and diamondlike films, Thin Solid Films **216**, 126–133 (1992)

33.36 B. Bhushan, A.J. Kellock, N.H. Cho, J.W. Ager III: Characterization of chemical bonding and physical characteristics of diamond-like amorphous carbon and diamond films, J. Mater. Res. **7**, 404–410 (1992)

33.37 J. Robertson: Properties of diamond-like carbon, Surf. Coat. Technol. **50**, 185–203 (1992)

33.38 N. Savvides, B. Window: Diamondlike amorphous carbon films prepared by magnetron sputtering of graphite, J. Vac. Sci. Technol. A **3**, 2386–2390 (1985)

33.39 J.C. Angus, P. Koidl, S. Domitz: Carbon thin films. In: *Plasma Deposited Thin Films*, ed. by J. Mort, F. Jensen (CRC, Boca Raton 1986) pp. 89–127

33.40 J. Robertson: Amorphous carbon, Adv. Phys. **35**, 317–374 (1986)

33.41 M. Rubin, C.B. Hooper, N.H. Cho, B. Bhushan: Optical and mechanical properties of DC sputtered carbon films, J. Mater. Res. **5**, 2538–2542 (1990)

33.42 G.J. Vandentop, M. Kawasaki, R.M. Nix, I.G. Brown, M. Salmeron, G.A. Somorjai: Formation of hydrogenated amorphous carbon films of controlled hardness from a methane plasma, Phys. Rev. B **41**, 3200–3210 (1990)

33.43 J.J. Cuomo, D.L. Pappas, R. Lossy, J.P. Doyle, J. Bruley, G.W. Di Bello, W. Krakow: Energetic carbon deposition at oblique angles, J. Vac. Sci. Technol. A **10**, 3414–3418 (1992)

33.44 D.L. Pappas, K.L. Saenger, J. Bruley, W. Krakow, J.J. Cuomo: Pulsed laser deposition of diamond-like carbon films, J. Appl. Phys. **71**, 5675–5684 (1992)

33.45 H.J. Scheibe, B. Schultrich: DLC film deposition by laser-arc and study of properties, Thin Solid Films **246**, 92–102 (1994)

33.46 C. Donnet, A. Grill: Friction control of diamond-like carbon coatings, Surf. Coat. Technol. **94/95**, 456 (1997)

33.47 A. Grill: Tribological properties of diamondlike carbon and related materials, Surf. Coat. Technol. **94/95**, 507 (1997)

33.48 B.K. Gupta, B. Bhushan: Micromechanical properties of amorphous carbon coatings deposited by different deposition techniques, Thin Solid Films **270**, 391–398 (1995)

33.49 X. Li, B. Bhushan: Micro/nanomechanical and tribological characterization of ultra-thin amorphous carbon coatings, J. Mater. Res. **14**, 2328–2337 (1999)

33.50 X. Li, B. Bhushan: Mechanical and tribological studies of ultra-thin hard carbon overcoats for magnetic recording heads, Z. Metallkd. **90**, 820–830 (1999)

33.51 S. Sundararajan, B. Bhushan: Micro/nanotribology of ultra-thin hard amorphous carbon coatings using atomic force/friction force microscopy, Wear **225–229**, 678–689 (1999)

33.52 B. Bhushan: Chemical, mechanical, and tribological characterization of ultra-thin and hard amorphous carbon coatings as thin as 3.5 nm: Recent developments, Diam. Relat. Mater. **8**, 1985–2015 (1999)

33.53 I.I. Aksenov, V.E. Strel'Nitskii: Wear resistance of diamond-like carbon coatings, Surf. Coat. Technol. **47**, 252–256 (1991)

33.54 D.R. McKenzie, D. Muller, B.A. Pailthorpe, Z.H. Wang, E. Kravtchinskaia, D. Segal, P.B. Lukins, P.J. Martin, G. Amaratunga, P.H. Gaskell, A. Saeed: Properties of tetrahedral amorphous carbon prepared by vacuum arc deposition, Diam. Relat. Mater. **1**, 51–59 (1991)

33.55 R. Lossy, D.L. Pappas, R.A. Roy, J.J. Cuomo: Filtered arc deposition of amorphous diamond, Appl. Phys. Lett. **61**, 171–173 (1992)

33.56 I.G. Brown, A. Anders, S. Anders, M.R. Dickinson, I.C. Ivanov, R.A. MacGill, X.Y. Yao, K.M. Yu: Plasma synthesis of metallic and composite thin films with atomically mixed substrate bonding, Nucl. Instrum. Methods B **80/81**, 1281–1287 (1993)

33.57 P.J. Fallon, V.S. Veerasamy, C.A. Davis, J. Robertson, G.A.J. Amaratunga, W.I. Milne, J. Koskinen: Properties of filtered-ion-beam-deposited diamond-like carbon as a function of ion energy, Phys. Rev. B **48**, 4777–4782 (1993)

33.58 S. Anders, A. Anders, I.G. Brown, B. Wei, K. Komvopoulos, J.W. Ager III, K.M. Yu: Effect of vacuum arc deposition parameters on the properties of amorphous carbon thin films, Surf. Coat. Technol. **68/69**, 388–393 (1994)

33.59 S. Anders, A. Anders, I.G. Brown, M.R. Dickinson, R.A. MacGill: Metal plasma immersion ion implantation and deposition using arc plasma sources, J. Vac. Sci. Technol. B **12**, 815–820 (1994)

33.60 S. Anders, A. Anders, I.G. Brown: Transport of vacuum arc plasma through magnetic macroparticle filters, Plasma Sources Sci. **4**, 1–12 (1995)

33.61 D.M. Swec, M.J. Mirtich, B.A. Banks: *Ion Beam and Plasma Methods of Producing Diamondlike Carbon Films* (NASA, Cleveland 1988), Report No. NASA TM102301

33.62 A. Erdemir, M. Switala, R. Wei, P. Wilbur: A tribological investigation of the graphite-to-diamond-like behavior of amorphous carbon films

ion beam deposited on ceramic substrates, Surf. Coat. Technol. **50**, 17–23 (1991)
33.63 A. Erdemir, F.A. Nicols, X.Z. Pan, R. Wei, P.J. Wilbur: Friction and wear performance of ion-beam deposited diamond-like carbon films on steel substrates, Diam. Relat. Mater. **3**, 119–125 (1993)
33.64 R. Wei, P.J. Wilbur, M.J. Liston: Effects of diamond-like hydrocarbon films on rolling contact fatigue of bearing steels, Diam. Relat. Mater. **2**, 898–903 (1993)
33.65 A. Erdemir, C. Donnet: Tribology of diamond, diamond-like carbon, and related films. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 871–908
33.66 J. Asmussen: Electron cyclotron resonance microwave discharges for etching and thin-film deposition, J. Vac. Sci. Technol. A **7**, 883–893 (1989)
33.67 J. Suzuki, S. Okada: Deposition of diamondlike carbon films using electron cyclotron resonance plasma chemical vapor deposition from ethylene gas, Jpn. J. Appl. Phys. **34**, L1218–L1220 (1995)
33.68 B.A. Banks, S.K. Rutledge: Ion beam sputter deposited diamond like films, J. Vac. Sci. Technol. **21**, 807–814 (1982)
33.69 C. Weissmantel, K. Bewilogua, K. Breuer, D. Dietrich, U. Ebersbach, H.J. Erler, B. Rau, G. Reisse: Preparation and properties of hard i-C and i-BN coatings, Thin Solid Films **96**, 31–44 (1982)
33.70 H. Dimigen, H. Hubsch: Applying low-friction wear-resistant thin solid films by physical vapor deposition, Philips Tech. Rev. **41**, 186–197 (1983)
33.71 B. Bhushan, B.K. Gupta, M.H. Azarian: Nanoindentation, microscratch, friction and wear studies for contact recording applications, Wear **181–183**, 743–758 (1995)
33.72 B. Bhushan, L. Yang, C. Gao, S. Suri, R.A. Miller, B. Marchon: Friction and wear studies of magnetic thin-film rigid disks with glass-ceramic, glass and aluminum-magnesium substrates, Wear **190**, 44–59 (1995)
33.73 L. Holland, S.M. Ojha: Deposition of hard and insulating carbonaceous films of an RF target in butane plasma, Thin Solid Films **38**, L17–L19 (1976)
33.74 L.P. Andersson: A review of recent work on hard i-C films, Thin Solid Films **86**, 193–200 (1981)
33.75 A. Bubenzer, B. Dischler, B. Brandt, P. Koidl: RF plasma deposited amorphous hydrogenated hard carbon thin films, preparation, properties and applications, J. Appl. Phys. **54**, 4590–4594 (1983)
33.76 A. Grill, B.S. Meyerson, V.V. Patel: Diamond-like carbon films by RF plasma-assisted chemical vapor deposition from acetylene, IBM J. Res. Dev. **34**, 849–857 (1990)
33.77 A. Grill, B.S. Meyerson, V.V. Patel: Interface modification for improving the adhesion of a-C:H to metals, J. Mater. Res. **3**, 214 (1988)
33.78 A. Grill, V.V. Patel, B.S. Meyerson: Optical and tribological properties of heat-treated diamond-like carbon, J. Mater. Res. **5**, 2531–2537 (1990)
33.79 F. Jansen, M. Machonkin, S. Kaplan, S. Hark: The effect of hydrogenation on the properties of ion beam sputter deposited amorphous carbon, J. Vac. Sci. Technol. A **3**, 605–609 (1985)
33.80 S. Kaplan, F. Jansen, M. Machonkin: Characterization of amorphous carbon-hydrogen films by solid-state nuclear magnetic resonance, Appl. Phys. Lett. **47**, 750–753 (1985)
33.81 H.C. Tsai, D.B. Bogy, M.K. Kundmann, D.K. Veirs, M.R. Hilton, S.T. Mayer: Structure and properties of sputtered carbon overcoats on rigid magnetic media disks, J. Vac. Sci. Technol. A **6**, 2307–2315 (1988)
33.82 B. Marchon, M. Salmeron, W. Siekhaus: Observation of graphitic and amorphous structures on the surface of hard carbon films by scanning tunneling microscopy, Phys. Rev. B **39**, 12907–12910 (1989)
33.83 B. Dischler, A. Bubenzer, P. Koidl: Hard carbon coatings with low optical absorption, Appl. Phys. Lett. **42**, 636–638 (1983)
33.84 D.S. Knight, W.B. White: Characterization of diamond films by Raman spectroscopy, J. Mater. Res. **4**, 385–393 (1989)
33.85 J.W. Ager III, D.K. Veirs, C.M. Rosenblatt: Spatially resolved Raman studies of diamond films grown by chemical vapor deposition, Phys. Rev. B **43**, 6491–6499 (1991)
33.86 W. Scharff, K. Hammer, O. Stenzel, J. Ullman, M. Vogel, T. Frauenheim, B. Eibisch, S. Roth, S. Schulze, I. Muhling: Preparation of amorphous i-C films by ion-assisted methods, Thin Solid Films **171**, 157–169 (1989)
33.87 B. Bhushan, X. Li: Nanomechanical characterization of solid surfaces and thin films, Int. Mater. Rev. **48**, 125–164 (2003)
33.88 X. Li, D. Diao, B. Bhushan: Fracture mechanisms of thin amorphous carbon films in nanoindentation, Acta Mater. **45**, 4453–4461 (1997)
33.89 X. Li, B. Bhushan: Measurement of fracture toughness of ultra-thin amorphous carbon films, Thin Solid Films **315**, 214–221 (1998)
33.90 X. Li, B. Bhushan: Evaluation of fracture toughness of ultra-thin amorphous carbon coatings deposited by different deposition techniques, Thin Solid Films **355/356**, 330–336 (1999)
33.91 X. Li, B. Bhushan: Development of a nanoscale fatigue measurement technique and its application to ultrathin amorphous carbon coatings, Scr. Mater. **47**, 473–479 (2002)
33.92 X. Li, B. Bhushan: Nanofatigue studies of ultrathin hard carbon overcoats used in magnetic storage devices, J. Appl. Phys. **91**, 8334–8336 (2002)
33.93 J. Robertson: Deposition of diamond-like carbon, Philos. Trans. R. Soc. Lond. A **342**, 277–286 (1993)
33.94 S.J. Bull: Tribology of carbon coatings: DLC, diamond and beyond, Diam. Relat. Mater. **4**, 827–836 (1995)
33.95 N. Savvides, T.J. Bell: Microhardness and Young's modulus of diamond and diamondlike carbon films, J. Appl. Phys. **72**, 2791–2796 (1992)
33.96 B. Bhushan, M.F. Doerner: Role of mechanical properties and surface texture in the real area of contact of magnetic rigid disks, ASME J. Tribol. **111**, 452–458 (1989)

33.97 S. Suresh: *Fatigue of Materials* (Cambridge Univ. Press, Cambridge 1991)

33.98 D.B. Marshall, A.G. Evans: Measurement of adherence of residual stresses in thin films by indentation. I. Mechanics of interface delamination, J. Appl. Phys. **15**, 2632–2638 (1984)

33.99 A.G. Evans, J.W. Hutchinson: On the mechanics of delamination and spalling in compressed films, Int. J. Solids Struct. **20**, 455–466 (1984)

33.100 S. Sundararajan, B. Bhushan: Development of a continuous microscratch technique in an atomic force microscope and its application to study scratch resistance of ultrathin hard amorphous carbon coatings, J. Mater. Res. **16**, 437–445 (2001)

33.101 B. Bhushan, B.K. Gupta: Micromechanical characterization of Ni-P coated aluminum-magnesium, glass and glass-ceramic substrates and finished magnetic thin-film rigid disks, Adv. Inf. Storage Syst. **6**, 193–208 (1995)

33.102 X. Li, B. Bhushan: Micromechanical and tribological characterization of hard amorphous carbon coatings as thin as 5 nm for magnetic recording heads, Wear **220**, 51–58 (1998)

33.103 B. Bhushan, V.N. Koinkar: Microscale mechanical and tribological characterization of hard amorphous coatings as thin as 5 nm for magnetic disks, Surf. Coat. Technol. **76/77**, 655–669 (1995)

33.104 V.N. Koinkar, B. Bhushan: Microtribological properties of hard amorphous carbon protective coatings for thin-film magnetic disks and heads, Proc. Inst. Mech. Eng. J. **211**, 365–372 (1997)

33.105 T.W. Wu: Microscratch and load relaxation tests for ultra-thin films, J. Mater. Res. **6**, 407–426 (1991)

33.106 R. Memming, H.J. Tolle, P.E. Wierenga: Properties of polymeric layers of hydrogenated amorphous carbon produced by plasma-activated chemical vapor deposition: tribological and mechanical properties, Thin Solid Films **143**, 31–41 (1986)

33.107 C. Donnet, T. Le Mogne, L. Ponsonnet, M. Belin, A. Grill, V. Patel: The respective role of oxygen and water vapor on the tribology of hydrogenated diamond-like carbon coatings, Tribol. Lett. **4**, 259 (1998)

33.108 F.P. Bowden, J.E. Young: Friction of diamond, graphite and carbon and the influence of surface films, Proc. R. Soc. Lond. A **208**, 444–455 (1951)

33.109 B. Marchon, N. Heiman, M.R. Khan: Evidence for tribochemical wear on amorphous carbon thin films, IEEE Trans. Magn. **26**, 168–170 (1990)

33.110 M.T. Dugger, Y.W. Chung, B. Bhushan, W. Rothschild: Friction, wear, and interfacial chemistry in thin film magnetic rigid disk files, ASME J. Tribol. **112**, 238–245 (1990)

33.111 B.D. Strom, D.B. Bogy, C.S. Bhatia, B. Bhushan: Tribochemical effects of various gases and water vapor on thin film magnetic disks with carbon overcoats, ASME J. Tribol. **113**, 689–693 (1991)

33.112 B. Bhushan, J. Ruan: Tribological performance of thin film amorphous carbon overcoats for magnetic recording rigid disks in various environments, Surf. Coat. Technol. **68/69**, 644–650 (1994)

33.113 B. Bhushan, G.S.A.M. Theunissen, X. Li: Tribological studies of chromium oxide films for magnetic recording applications, Thin Solid Films **311**, 67–80 (1997)

34. Self-Assembled Monolayers for Nanotribology and Surface Protection

Bharat Bhushan

In many applications, hydrophobic and lubricating films are needed to minimize adhesion, stiction, friction, and wear. In various applications, surfaces need to be protected from exposure to the operating environment, and hydrophobic films are of interest. Self-assembled monolayers are molecularly thick, well organized, and chemically bonded to the substrate. Some can be insensitive to the environment and can provide lower adhesion, friction, and wear. Ordered molecular assemblies with high hydrophobicity and desirable nanotribological properties can be engineered using chemical grafting of various polymer molecules with suitable functional head groups, spacer chains, and nonpolar surface terminal groups.

In this chapter, we present an overview of self-assembled monolayers (SAMs) with hydrophobicity and desirable nanotribological properties. SAMs are produced by various organic precursors. We first present a primer to organic chemistry, followed by an overview of selected SAMs with various substrates, spacer chains, and terminal groups in the molecular chains, and an overview of nanotribological properties of SAMs. The contact angle, adhesion, friction, and wear properties of SAMs having various spacer chains with different surface terminal and head groups (hexadecane thiol, biphenyl thiol, perfluoroalkylsilane, alkylsilane, perfluoroalkylphosphonate, and alkylphosphonate) and on various substrates (Au, Si, and Al) are surveyed. Chemical degradation mechanisms and environmental effects are presented. Based on the contact angle and nanotribological properties of various SAM films discovered using an atomic force microscope (AFM), it is found that perfluoroalkylsilane and perfluorophosphonate SAMs exhibit attractive hydrophobic and nanotribological properties.

34.1 **Background** 1179
34.1.1 Need for Hydrophobic Surfaces for Nanotribology 1180
34.1.2 Surface Films for Nanotribology and Surface Protection 1180
34.1.3 Scope of the Chapter 1182
34.2 **A Primer on Organic Chemistry** 1182
34.2.1 Electronegativity/Polarity 1183
34.2.2 Classification and Structure of Organic Compounds 1183
34.2.3 Polar and Nonpolar Groups 1186
34.3 **Self-Assembled Monolayers: Substrates, Spacer Chains, and End Groups in the Molecular Chains** 1186
34.4 **Contact Angle and Nanotribological Properties of SAMs** 1189
34.4.1 Measurement Techniques 1191
34.4.2 Hexadecane Thiol and Biphenyl Thiol SAMs on Au(111) 1192
34.4.3 Perfluoroalkylsilane and Alkylsilane SAMs on Si(100), and Perfluoroalkylphosphonate and Alkylphosphonate SAMS on Al 1198
34.4.4 Chemical Degradation and Environmental Studies 1205
34.5 **Conclusion** 1209
References 1210

34.1 Background

The reliability concerns for engineering devices require the use of hydrophobic and lubricating films to minimize adhesion, stiction, friction, and wear [34.1–9]. In various applications, surfaces need to be protected from exposure to the operating environment. Examples include micro/nanoelectromechanical systems (MEMS/NEMS) and bioMEMS/bioNEMS, requiring relative motion, and magnetic storage devices such as magnetic rigid disk and tape drives. In some biomedical applications, such as biosensors and implantable

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_34

biomedical devices, undesirable protein adsorption, biofouling, and biocompatibility are some of the major issues [34.5, 6, 10]. In micro- and nanofluidic-based sensors, the fluid drag in micro- and nanochannels can be reduced by using hydrophobic coatings.

34.1.1 Need for Hydrophobic Surfaces for Nanotribology

The source of the liquid film at the interface can be either a preexisting film of liquid and/or capillary condensates of water vapor from the environment. If the liquid wets the surface ($0 \leq \theta < 90°$, where θ is the contact angle between the liquid-vapor interface and the liquid-solid interface for a liquid droplet sitting on a solid surface; Fig. 34.1a), the liquid surface is thereby constrained to lie parallel with the surface [34.11–13], and the complete liquid surface therefore is concave in shape; Fig. 34.1b. The direct measurement of contact angle is most widely made from sessile drops. The angle is generally measured by aligning a tangent with the drop profile at the point of contact with the solid surface using a telescope equipped with a goniometer eyepiece. Surface tension results in a pressure difference across any meniscus surface, referred to as capillary pressure or Laplace pressure, and is negative for a concave meniscus [34.14, 15]. The negative Laplace pressure results in an intrinsic attractive (adhesive) force, which depends on the interface roughness (local geometry of interacting asperities and number of asperities), the surface tension, and the contact angle.

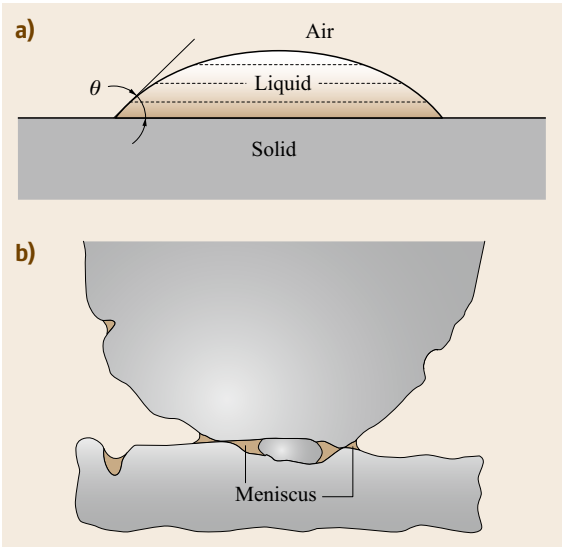


Fig. 34.1 (a) Schematic of a sessile drop on a solid surface and the definition of contact angle, and (b) formation of meniscus bridges as a result of liquid present at an interface

During normal separation, this intrinsic force needs to be overcome [34.14–16]. During sliding, frictional effects need to be overcome, not only because of external load but also because of intrinsic adhesive force. The measured value of high static friction force, contributed largely by liquid-mediated adhesion (meniscus contribution), is generally referred to as *stiction*. It becomes a major concern in micro- and nanodevices operating at ultralow loads as the liquid-mediated adhesive force may be on the same order as the external load. The effect of liquid-mediated adhesion can be minimized by increasing surface roughness and/or the use of a liquid with low surface tension with a film thickness on the order of surface roughness, as well as its chemical bonding to the substrate [34.14–22]. The formation of menisci and/or condensation of water vapor from the environment at the interface can be minimized by the use of hydrophobic (water-fearing) coatings. The surfaces can be made superhydrophobic by the introduction of controlled roughness on the surfaces to take advantage of the so-called Lotus effect [34.10, 23, 24].

34.1.2 Surface Films for Nanotribology and Surface Protection

Surfaces can be treated or coated with a liquid with relatively low surface tension or certain solid films to make them hydrophobic and/or to control adhesion, stiction, friction, and wear.

The classical approach to lubrication uses freely supported multimolecular layers of liquid lubricants [34.2, 4, 14, 15, 21, 25–29]. Boundary lubricant films are formed by either physisorption, chemisorption, or chemical reaction. The physisorbed films can be either monomolecularly or polymolecularly thick. The chemisorbed films are monomolecular, but stoichiometric films formed by chemical reaction can be multilayers. In general, stability and durability of surface films decrease in the following order: chemically reacted films, chemisorbed films, and physisorbed films. A good boundary lubricant should have a high degree of interaction between its molecules and the sliding surface. As a general rule, liquids will have a more desirable performance when they are polar and thus able to grip on solid surfaces (or be adsorbed). Polar lubricants contain reactive functional end groups. Boundary lubrication properties are also dependent upon the molecular conformation and lubricant spreading. It should be noted that the liquid films with the thickness on the order of a few nm may be discontinuous, and may deposit in an island form with nonuniform thickness with a lateral resolution on the nm scale.

Solid films are also commonly used for controlling hydrophobicity and/or adhesion, stiction, friction, and

wear. Hydrophobic films have nonpolar surface terminal groups (to be described later) that repel water. These films have low surface energy (15–30 dyn/cm) and a high contact angle ($\theta \geq 90°$), which minimize wetting [34.26, 30, 31]. Multimolecularly thick (few tenths of nm) films of conventional solid-lubricants have been studied. *Hansma* et al. [34.32] reported the deposition of multimolecularly-thick, highly oriented polytetrafluoroethylene (PTFE) films from the melt, vapor phase or from solution by a mechanical deposition technique involving dragging the polymer at controlled temperature, pressure and speed against a smooth glass substrate. *Scandella* et al. [34.33] reported that the coefficient of nanoscale friction of MoS_2 platelets on mica, obtained by the exfoliation of lithium intercalated MoS_2 in water, was a factor of 1.4 less than that of mica itself. However, MoS_2 is reactive to water, and its friction and wear properties degrade with an increase in humidity [34.14, 15].

Amorphous diamond-like carbon (DLC) coatings can be produced with extremely high hardness and are commercially used as wear-resistant coatings [34.27, 34]. They are widely used in magnetic storage devices [34.2]. Doping of the DLC matrix with elements like hydrogen, nitrogen, oxygen, silicon, and fluorine influences their hydrophobicity and tribological properties [34.27, 35, 36]. Nitrogen and oxygen reduce the contact angle (or increase the surface energy) due to the strong polarity that is formed when these elements are bonded to carbon. On the other hand, silicon and fluorine increase the contact angle ranging from 70 to 100° (or reduce the surface energy to 20–40 dyn/cm), making them hydrophobic [34.37, 38]. Nanocomposite coatings with a diamond-like carbon (a-C:H) network and a glass-like a-Si:O network are generally deposited using a PECVD (plasma-enhanced chemical vapor deposition) technique in which plasma is formed from a siloxane precursor using a hot filament. For a fluorinated DLC, CF_4 as the fluorocarbon source is added to an acetylene plasma. In addition, fluorination of DLC can be achieved by the postdeposition treatment of DLC coatings in a CF_4 plasma. Silicon- and fluorine-containing DLC coatings mainly reduce their polarity due to the loss of sp^2-bonded carbon (due to polarization potential of the involved π electrons) and dangling bonds of the DLC network. As silicon and fluorine are unable to form double bonds, they force carbon into a sp^3 bonding state [34.38]. Friction and wear properties of both silicon-containing and fluorinated DLC coatings have been reported to be superior to those of conventional DLC coatings [34.39, 40]. However, DLC coatings require a line-of-sight deposition process that prevents deposition on complex geometries. Furthermore, it has been reported that some self-assembled monolayers (SAMs) are superior to DLC coatings in hydrophobicity and tribological performance [34.41, 42].

Organized and dense molecular-scale layers of, preferably, long-chain organic molecules are known to be superior lubricants on both macro- and micro/nanoscales as compared with freely supported multimolecular layers [34.4]. Common techniques to produce molecular-scale organized layers are Langmuir–Blodgett (L–B) deposition and chemical grafting of organic molecules to realize SAMs [34.30, 31]. In the L–B technique, organic molecules from suitable amphiphilic molecules are first organized at the air-water interface and then physisorbed on a solid surface to form mono- or multimolecular layers [34.43]. In the case of SAMs, the functional groups of molecules chemisorb on a solid surface, which results in the spontaneous formation of robust, highly ordered, oriented, and dense monolayers [34.31]. In both cases, the organic molecules used have well-distinguished amphiphilic properties (a hydrophilic functional head and a hydrophobic aliphatic tail) so that an adsorption of such molecules on an active inorganic substrate leads to their firm attachment to the surface. Direct organization of SAMs on the solid surfaces allows coating in inaccessible areas. The weak adhesion of classical L–B films to the substrate surface restricts their lifetimes during sliding, whereas certain SAMs can be very durable. As a result, SAMs are of great interest in tribological applications.

Much research into the application of SAMs has been carried out using the so-called soft lithographic technique [34.44, 45]. This is a nonphotolithographic technique. Photolithography is based on a projection-printing system used for projection of an image from a mask to a thin-film photoresist, and its resolution is limited by optical diffraction limits. In soft lithography, an elastomeric stamp or mold is used to generate micropatterns of SAMs by either contact printing (known as microcontact printing or μCP [34.46]), by embossing (nanoimprint lithography) [34.47], or by replica molding [34.48], which circumvent the diffraction limits of photolithography. The stamps are generally cast from photolithographically generated patterned masters, and the stamp material is generally polydimethylsiloxane (PDMS). In μCP, the ink is a SAM precursor to produce nm-thick resists with lines thinner than 100 nm. Soft lithography requires little capital investment. μCP and embossing techniques may be used to produce microdevices that are substantially cheaper and more flexible in choice of material for construction than conventional photolithography (e.g., SAMs and non-SAM entities for μCP and elastomers for embossing).

The largest industrial application for SAMs is in digital micromirror devices (DMDs) used in optical

projection displays [34.49, 50]. The chip set of a DMD consists of half a million to more than two million independently controlled reflective aluminum alloy micromirrors of about 12 μm square. These micromirrors switch forward and backward at a frequency on the order of 5–7 kHz with a rotation of ±12° with respect to the horizontal plane; the movement is limited by a mechanical stop. Mechanical contact leads to stiction and wear in contacting surfaces. A SAM of vapor-deposited perfluorinated *n*-alkanoic acid ($C_nF_{2n-1}O_2H$) (e.g., perfluorodecanoic acid, or PFDA, $CF_3(CF_2)_8COOH$) is used to coat contacting surfaces to make them hydrophobic and thus to minimize meniscus formation. Furthermore, the entire DMD chip set is hermetically sealed in order to prevent particulate contamination and excessive condensation of water at the contacting surfaces. A so-called *getter* strip of PFDA is included inside the hermetically sealed enclosure containing the chip, which acts as a reservoir in order to maintain a PFDA vapor within the package. Degradation mechanisms of SAMs leading to stiction have been studied by *Liu* and *Bhushan* [34.51, 52]. Nanotribological studies of various SAMs on Al substrates have been carried out by *Tambe* and *Bhushan* [34.53], *Bhushan* et al. [34.54], *Hoque* et al. [34.55, 56] and *DeRose* et al. [34.57].

There are various other micro/nanodevices that require SAMs for hydrophobicity in order to minimize meniscus formation. Examples include micromotors, microgears, microvalves, microswitches, mirror-based optical switches, and atomic force microscopy probes [34.5, 6, 10]. Nanotribological studies on Si substrates have been carried out by *Bhushan* et al. [34.58–60], *Kasai* et al. [34.61], *Lee* et al. [34.62], *Tambe* and *Bhushan* [34.53] and *Tao* and *Bhushan* [34.63, 64]. SAMs deposition on Cu surfaces are also being explored for corrosion inhibition for micro/nanoelectronics and/or heat-exchange surfaces exploiting dropwise condensation [34.65–67].

Other industrial applications for SAMs are in the areas of biochemical and optical sensors, devices for use as drug-delivery vehicles, and in the construction of electronic components [34.10, 68–71]. Biochemical sensors require highly sensitive organic layers with tailored biological properties that can be incorporated into electronic, optical, or electrochemical devices. Self-assembled microscopic vesicles are being developed to ferry potentially lifesaving drugs to cancer patients. By assembling organic, metal, and phosphonate molecules (complexes of phosphorous and oxygen atoms) themselves into conductive materials, these can be produced as self-made sandwiches for use as electronic components. Several applications have been proposed based on silicon, glass, or polymer nanochannels, including cell immunoisolation chambers, DNA separation devices, and biocapsules for drug delivery [34.5, 6].

SAMs are also being considered for protection of the surfaces from exposure to the operating environment. They are being developed to reduce corrosion and oxidation of Cu in heat exchangers [34.65–67]. They are being developed to minimize undesirable protein adsorption, biofouling and improve biocompatibility in biosensors and implantable biomedical devices [34.53, 54, 58–63, 72, 73]. These films can also be used to reduce fluid drag in micro/nanochannels.

34.1.3 Scope of the Chapter

An overview of molecularly thick layers of liquid lubricants and conventional solid lubricants can be found in various references, such as [34.2, 4, 10, 14, 15, 18, 19, 21, 27, 29]. In this chapter, we present an overview of SAMs for high hydrophobicity, low adhesion, friction, and wear. SAMs are produced by various organic precursors. We first present a primer on organic chemistry followed by an overview of suitable substrates, spacer chains and end groups in the molecular chains, an overview of contact angle, adhesion, friction, and wear properties of various SAMs, and some concluding remarks.

34.2 A Primer on Organic Chemistry

All organic compounds contain the carbon atom. Carbon, in combination with hydrogen, oxygen, nitrogen, sulfur and phosphor results in a large number of organic compounds. The atomic number of carbon is 6, and its electron structure is $1s^2 2s^2 2p^2$. Two stable isotopes of carbon, ^{12}C and ^{13}C, exist. With four electrons in its outer shell, carbon forms four covalent bonds with each bond resulting from two atoms sharing a pair of electrons. The number of electron pairs that two atoms share determines whether or not the bond is single or multiple. In a single bond, only one pair of electrons is shared by the atoms. Carbon can also form multiple bonds by sharing two or three pairs of electrons between the atoms. For example, the double bond formed by sharing two electron pairs is stronger than a single bond, and it is shorter than a single bond. An organic compound is classified as saturated if it contains only single bonds and as unsaturated if the

molecules possess one or more multiple carbon-carbon bonds.

34.2.1 Electronegativity/Polarity

When two different kinds of atoms share a pair of electrons, a bond is formed in which electrons are shared unequally: one atom assumes a partial positive charge and the other a negative charge with respect to each other. This difference in charge occurs because the two atoms exert unequal attraction in the pair of shared electrons. The attractive force that an atom of an element has for shared electrons in a molecule or polyatomic ion is known as its electronegativity. Elements differ in their electronegativities. A scale of relative electronegatives, in which the most electronegative element, fluorine, is assigned a value of 4.0, was developed by L. Pauling. Relative electronegativities of the elements in the periodic table can be found in most undergraduate chemistry textbooks [34.74]. Relative electronegativity of the nonmetals is high as compared to that of metals. Relative electronegativity of selected elements of interest with high values is presented in Table 34.1.

The polarity of a bond is determined by the difference in electronegativity values of the atoms forming the bond. If the electronegativities are the same the bond is nonpolar, and the electrons are shared equally. In this type of bond, there is no separation of positive and negative charge between atoms. If the atoms have greatly different electronegativities the bond is very polar. A dipole is a molecule that is electrically asymmetrical, causing it to be oppositely charged at two points. As an example, in hydrogen chloride, HCl, both hydrogen and chlorine need one electron to form stable electron configurations. They share a pair of electrons. Chlorine is more electronegative and therefore has a greater attraction for the shared electrons than does hydrogen. As a result, the pair of electrons is displaced towards the chlorine atom, giving it a partial negative charge and leaving the hydrogen atom with a partial positive charge (Fig. 34.2). However, the entire molecule, HCl, is electrically neutral. The hydrogen atom with a partial positive charge (exposed proton on one end) can be easily attracted to the negative charge of other molecules and this is responsible for the polarity of the molecule. A partial charge is usually indicated by δ, and the electronic structure of HCl is given as

$$\overset{\delta+}{\mathrm{H}} : \overset{\delta-}{\underset{\cdot\cdot}{\overset{\cdot\cdot}{\mathrm{Cl}}}} :$$

Similar to the HCl molecule, hydrofluoric acid (HF) is polar, and both behave as a small dipole. On the other hand, methane (CH_4), carbon tetrachloride (CCl_4), and carbon dioxide (CO_2) are nonpolar. In CH_4 and CCl_4, the four C−H and C−Cl polar bonds are identical, and because these bonds emanate from the center to the corners of a tetrahedron in the molecule, the effect of their polarities cancels one another. CO_2 (O=C=O) is nonpolar because carbon-oxygen dipoles cancel each other by acting in opposite directions. Water (H−O−H) is a polar molecule. If the atoms in water were linear as in CO_2, two O−H dipoles would cancel each other, and the molecule would be nonpolar. However, water has a bent structure with an angle of 105° between the two bonds, which is responsible for the water being a polar molecule.

Table 34.1 Relative electronegativity of selected elements

Element	Relative electronegativity
F	4.0
O	3.5
N	3.0
Cl	3.0
C	2.5
S	2.5
P	2.1
H	2.1

34.2.2 Classification and Structure of Organic Compounds

Table 34.2 presents selected organic compounds grouped into classes.

Hydrocarbons

Hydrocarbons are compounds that are composed entirely of carbon and hydrogen atoms bonded to each other by covalent bonds. Saturated hydrocarbons (alkanes) contain single bonds. Unsaturated hydrocarbons that contain carbon-carbon double bonds are called

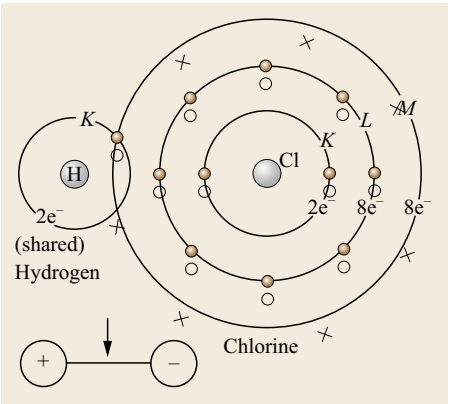


Fig. 34.2 Schematic representation of the formation of a polar HCl molecule

alkenes, and ones with triple bonds are called alkynes. Unsaturated hydrocarbons that contain aromatic rings, e.g., benzene rings, are called aromatic hydrocarbons.

Saturated Hydrocarbons: Alkanes. The alkanes, also known as paraffins, are saturated hydrocarbons: straight- or branched-chain hydrocarbons with only single covalent bonds between the carbon atoms. The general molecular formula for alkanes is C_nH_{2n+2}, where n is the number of carbon atoms in the molecule. Each carbon atom is connected to four other atoms by four single covalent bonds. These bonds are separated by angles of 109.5° (the angle by lines drawn from the center of a regular tetrahedron to its corners). Alkane molecules contain only carbon-carbon and carbon-hydrogen bonds, which are symmetrically directed towards the corners of a tetrahedron. Therefore alkane molecules are essentially nonpolar.

Common alkyl groups have the general formula C_nH_{2n+1} (one hydrogen atom less than the corresponding alkane). The missing H atom may be detached from any carbon in the alkane. The name of the group is formed from the name of the corresponding alkane by replacing the -ane with an -yl ending. Some examples are shown in Table 34.2.

Unsaturated Hydrocarbons. Unsaturated hydrocarbons consist of three families of compounds that contain fewer hydrogen atoms than the alkane with the corresponding number of carbon atoms, and contain multiple bonds between carbon atoms. These include alkenes (with carbon-carbon double bonds), alkynes (with carbon-carbon triple bonds), and aromatic compounds (with benzene rings, which are arranged in a six-membered ring with one hydrogen atom bonded to each carbon atom). Some examples are shown in Table 34.2.

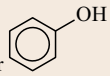
Table 34.2 Names and formulas of selected hydrocarbons

Name	Formula
Saturated hydrocarbons	
Straight-chain alkanes	C_nH_{2n+2}
e.g., methane	CH_4
ethane	C_2H_6 or CH_3CH_3
Alkyl groups	C_nH_{2n+1}
e.g., methyl	$-CH_3$
ethyl	$-CH_2CH_3$
Unsaturated hydrocarbons	
Alkenes	$(CH_2)_n$
e.g., ethene	C_2H_4 or $CH_2{=}CH_2$
propene	C_3H_6 or $CH_3CH{=}CH_2$
Alkynes	
e.g., acetylene	$HC \equiv CH$
Aromatic hydrocarbons	
e.g., benzene	C_6H_5OH or 

Alcohols, Ethers, Phenols, and Thiols

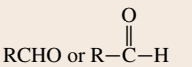
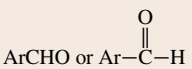
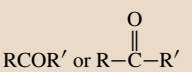
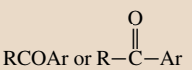
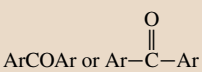
Organic molecules with certain functional groups are synthesized for desirable properties. Alcohols, ethers and phenols are derived from the structure of water by replacing the hydrogen atoms of water with alkyl groups (R) or aromatic (Ar) rings. For example, phenol is a class of compounds that has a hydroxy group

Table 34.3 Names and formulas of selected alcohols, ethers, phenols and thiols

Name	Formula[a]
Alcohols	R−OH
e.g., methanol	CH_3OH
ethanol	CH_3CH_2OH
Ethers	R−O−R′
e.g., dimethyl ether	$CH_3{-}O{-}CH_3$
diethyl ether	$CH_3CH_2{-}O{-}CH_2CH_3$
Phenols	 C_6H_5OH or
Thiols	−SH
e.g., methanethiol	CH_3SH

[a] The letters R- and R′- represent an alkyl group. The R- groups in ethers can be the same or different and can be alkyl or aromatic (Ar) groups

Table 34.4 Names and formulas of selected aldehydes and ketones

Name	Formula[a]
Aldehydes	 RCHO or R−C(=O)−H
	 ArCHO or Ar−C(=O)−H
e.g., methanal or formaldehyde	HCHO
ethanal or acetaldehyde	CH_3CHO
Ketones	 RCOR′ or R−C(=O)−R′
	 RCOAr or R−C(=O)−Ar
	 ArCOAr or Ar−C(=O)−Ar
e.g., butanone or methyl ethyl ketone	$CH_3COCH_2CH_3$

[a] The letters R and R′ represent an alkyl group and Ar represents an aromatic group

attached to an aromatic ring (benzene ring). Organic compounds that contain the –SH group are analogs of alcohols, and are known as thiols. Some examples are shown in Table 34.3.

Aldehydes and Ketones

Both aldehydes and ketones contain the carbonyl group C=O, a carbon-oxygen double bond. Aldehydes have at least one hydrogen atom bonded to the carbonyl group, whereas ketones have only alkyl or aromatic group bonded to the carbonyl group. The general formula for the saturated homologous series of aldehydes and ketones is $C_nH_{2n}O$. Some examples are shown in Table 34.4.

Carboxyl Acids and Esters

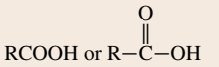
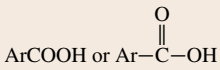
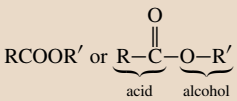
The functional group of the carboxylic acids is known as a carboxyl group represented as −COOH. Carboxylic acids can be either aliphatic (RCOOH) or aromatic (ArCOOH). The carboxylic acids with even numbers of carbon atoms, n, ranging from 4 to about 20 are called fatty acids (e.g., $n = 10$, 12, 14, 16 and 18 are called capric acid, lauric acid, myristic acid, palmitic acid, and stearic acid respectively).

Esters are alcohol derivatives of carboxylic acids. Their general formula is RCOOR′, where R may be a hydrogen, alkyl group, or aromatic group, and R′ may be an alkyl group or aromatic group, but not a hydrogen. Esters are found in fats and oils. Some examples are shown in Table 34.5.

Amides and Amines

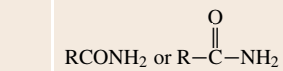
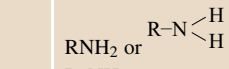
Amides and amines are organic compounds containing nitrogen. Amides are nitrogen derivatives of carboxylic acids. The carbon atom of a carbonyl group is bonded

Table 34.5 Names and formulas of selected carboxylic acids and esters

Name	Formula
Carboxylic acid[a]	 RCOOH or R–C(=O)–OH
	 ArCOOH or Ar–C(=O)–OH
e.g., methanoic acid (formic acid)	HCOOH
ethanoic acid (acetic acid)	CH_3COOH
octadecanoic acid (stearic acid)	$CH_3(CH_2)_{16}COOH$
Esters[b]	 RCOOR′ or R–C(=O)–O–R′ (acid, alcohol)
e.g., methyl propanoate	$CH_3CH_2COOCH_3$

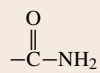
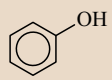
[a] The letter R represents an alkyl group and Ar represents an aromatic group
[b] The letter R represents hydrogen, alkyl group or aromatic group and R′ represents alkyl group or aromatic group

Table 34.6 Names and formulas of selected organic nitrogen compounds (amides and amines)

Name	Formula[a]
Amides	 $RCONH_2$ or R–C(=O)–NH_2
e.g., methanamide (formamide)	$HCONH_2$
ethanamide (acetamide)	CH_3CONH_2
Amines	 RNH_2 or R_2NH or R_3N
e.g., methylamine	CH_3NH_2
ethylamine	$CH_3CH_2NH_2$

[a] The letter R represents an alkyl group or aromatic group

Table 34.7 Some examples of polar (hydrophilic) and nonpolar (hydrophobic) groups

Name	Formula[a]
Polar	
Alcohol (hydroxyl)	−OH
Carboxyl acid	−COOH
Aldehyde	−COH
Ketone	 R–C(=O)–R
Ester	−COO−
Carbonyl	>C=O
Ether	R−O−R
Amine	$-NH_2$
Amide	 –C(=O)–NH_2
Phenol	 OH
Thiol	−SH
Trichlorosilane	$SiCl_3$
Nonpolar	
Methyl	$-CH_3$
Trifluoromethyl	$-CF_3$
Aryl (benzene ring)	

[a] The letter R represents an alkyl group

directly to a nitrogen atom of an $-NH_2$, $-NHR$, or $-NR_2$ group. The characteristic structure of amide is $RCONH_2$.

An amine is a substituted ammonia molecule which, has a general structure of RNH_2, R_2NH or R_3N, where R is an alkyl or aromatic group. Some examples are shown in Table 34.6.

34.2.3 Polar and Nonpolar Groups

Table 34.7 summarizes polar and nonpolar groups commonly used in the construction of hydrophobic and hydrophilic molecules. Table 34.8 lists the relative polarity of selected polar groups [34.75]. Thiol, silane, carboxylic acid, and alcohol (hydroxyl) groups are the most commonly used polar anchor groups for their attachment to surfaces. Silane anchor groups are commonly used for Si or SiO_2 surfaces as $-Si-O-$ bonds are strong. Methyl and trifluoromethyl are commonly used as end groups for hydrophobic film surfaces.

Table 34.8 Organic groups listed in the increasing order of polarity

Alkanes
Alkenes
Aromatic hydrocarbons
Ethers
Trichlorosilanes
Aldehydes, ketones, esters, carbonyls
Thiols
Amines
Alcohols, phenols
Amides
Carboxylic acids

34.3 Self-Assembled Monolayers: Substrates, Spacer Chains, and End Groups in the Molecular Chains

SAMs are formed as a result of spontaneous self-organization of functionalized organic molecules onto the surfaces of appropriate substrates into stable, well-defined structures; Fig. 34.3. The final structure is close to or at thermodynamic equilibrium, and as a result it tends to form spontaneously and rejects defects. SAMs consist of three building groups: a head group that binds strongly to a substrate, a surface terminal (tail or end) group that constitutes the outer surface of the film, and a spacer chain (backbone chain) that connects the head and surface terminal groups. The SAMs are named based on the surface terminal group followed by the spacer chain and the head group (or type of compound formed at the surface). In order to control hydrophobicity, adhesion, friction, and wear, it should be strongly adherent to the substrate, and the surface terminal group of the organic molecular chain should be nonpolar. For a strong attachment of the organic molecules to the substrate, the head group of the molecular chain should contain a polar terminal group resulting in the exothermic process (energies on the order of tens of kcal/mol), i. e., it results in an apparent pinning of the head group to a specific site on the surface through a chemical bond. Furthermore, the molecular structure and any cross-linking would have a significant effect on their friction and wear performance. The substrate surface should have a high surface energy (hydrophilic), so that there will be a strong tendency for molecules to adsorb on the surface. The surface should be highly functional with polar groups and dangling bonds (generally unpaired electrons) so that they can react with organic molecules and provide a strong bond. Because of the exothermic head group-substrate interactions, molecules try to occupy every available binding site on the surface, and during this process they generally push together molecules that have already adsorbed. The process results in the formation of ordered molecular assemblies. The interactions between molecular chains are van der Waals or electrostatic types with energies on the order of a few (< 10) kcal/mol, exothermic. The molecular chains in SAMs are not perpendicular to the surface; the tilt angle depends on the anchor group as well

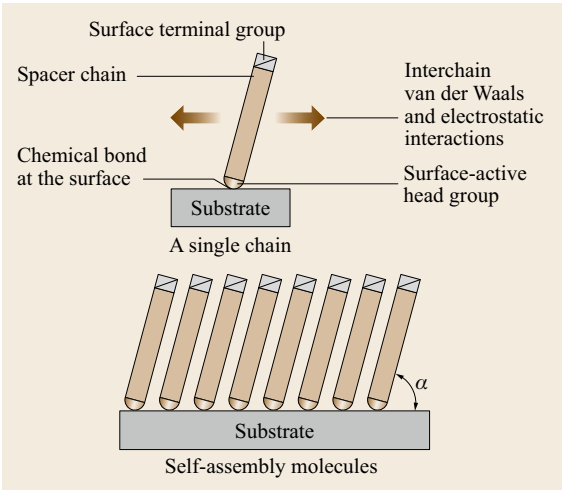


Fig. 34.3 Schematic of a self-assembled monolayer on a surface and the associated forces

as on the substrate and the spacer group. For example, the tilt angle for alkanethiolate on Au is typically about a 30–35° angle with respect to substrate normal.

Table 34.9 lists selected systems that have been used for the formation of SAMs [34.45]. The spacer chain of the SAM is mostly an alkyl chain ($(-CH_2)_n$) or made of a derivatized alkyl group. By attaching different terminal groups at the surface, the film surface can be made to attract or repel water. The commonly used surface terminal group of a hydrophobic film with low surface energy, in the case of a single alkyl chain, is a nonpolar methyl ($-CH_3$) or trifluoromethyl ($-CF_3$) group. For a hydrophilic film, the commonly used surface terminal groups are alcohol ($-OH$) or carboxylic acid ($-COOH$) groups. The surface active head groups most commonly used are thiol ($-SH$), silane (e.g., trichlorosilane or $-SiCl_3$), and carboxyl ($-COOH$) groups. The substrates most commonly used are gold, silver, platinum, copper, hydroxylated (activated) surfaces of SiO_2 on Si, Al_2O_3 on Al, and glass.

As an example of substrate, epitaxial Au film on glass, mica or single-crystal silicon, produced by e-beam evaporation, is commonly used because it can be deposited on smooth surfaces as a film that is atomically flat and defect free. Bulk single-crystal Si, sputtered Al film and bulk Al sheets, and copper disks with natural oxide layers, have been selected because of their use in the construction of MEMS/NEMS (e.g., digital projection displays) and heat exchange applications respectively. The substrate surface should be clean before deposition. For silicon substrates, a concentrated HF solution (typically 49% HF) is commonly used to remove the oxide layer followed by a rinse with deionized (DI) water [34.60, 62]. Hydrogen passivates the surface by saturating the dangling bonds, and this results in a hydrogen-terminated silicon surface with hydrophobic properties. For deposition of multimolecularly thick polymer films with nonpolar ends, hydrophobic substrates may lead to a coated surface with a high contact angle, and are preferred. For SAM deposition, the substrate should be hydrophilic in order to form strong interfacial bonds with their head groups. Hydroxylation of oxide surfaces is carried out to make them hydrophilic. Silicon and other metals get oxidized and get hydroxylated to some degree when exposed to the environment. Bulk silicon, polysilicon film or SiO_2 film surfaces are commonly treated to produce a hydroxylated silica surface by immersion in piranha solution (a mixture of typically 3 : 1 v/v 98% H_2SO_4 : 30% H_2O_2) at temperatures of about 90 °C for about 30 min followed by a rinse in deionized water [34.29, 60, 62]. Piranha solution also removes any organic and metallic contaminants, whereas HF would not necessarily remove organics. Oxygen plasma is another technique used for hydroxylation for SiO_2 as well as polymer surfaces [34.59, 61, 64, 73]. For complex silicon geometries or fine structures, such as AFM tips, oxygen plasma may be preferable. Figure 34.4 shows the schematics of surfaces after various surface treatments. Surfaces after piranha or oxygen plasma treatment remain hydrophilic for a few hours to about a day and become hydrophobic when they come in contact with carbon. They can retain hydrophilicity longer in dry nitrogen. To retain hydrophobicity, polymers are generally stored in DI water. Treated surfaces with HF remain hydrophobic for about 2–3 h and can retain hydrophobicity longer in dry nitrogen.

Table 34.9 Selected substrates and precursors that have commonly been used for formation of SAMs

Substrate	Precursor	Binding with substrate[a]
Au	R SH (thiol)	RS-Au
Au	Ar SH (thiol)	ArS-Au
Au	RSSR′ (disulfide)	RS-Au
Au	RSR′ (sulfide)	
Si/SiO_2, glass	R $SiCl_3$ (trichlorosilane)	Si–O–Si (siloxane)
Si/Si–H	R COOH (carboxyl)	R–Si
Metal oxides (e.g., Al_2O_3, SnO_2, TiO_2)	RCOOH (carboxyl)	RCOO–…MO_n

[a] R represents alkane (C_nH_{2n+2}) and Ar represents aromatic hydrocarbon. It consists of various surface active head groups with mostly methyl terminal groups

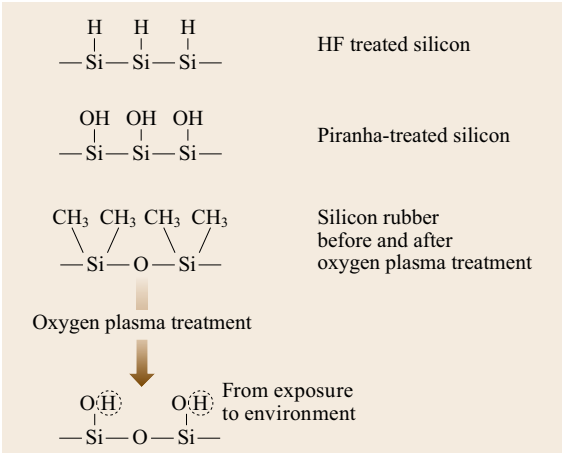


Fig. 34.4 Schematic showing HF-treated silica and the hydroxylation process occurring on a silica and elastomeric surface using piranha solution and oxygen plasma, respectively

For organic molecules to pack together and provide a better ordering, a substrate for given molecules should be selected such that the cross-sectional diameter of the spacer chains of the molecule is equal to or smaller than the distance between the anchor groups attached to the substrate. For the case of alkanethiol film, the advantage of Au substrate over SiO_2 substrate is that it results in better ordering because the cross-sectional diameter of the alkane molecule is slightly smaller than the distance between sulfur atoms attached to the Au substrate (≈ 0.53 nm). The thickness of the film can be controlled by varying the length of the hydrocarbon chain, and the surface properties of the film can be modified by the terminal group.

SAMs are usually produced by immersing a substrate in a solution containing the precursor (ligand) that is reactive to the substrate surface or by exposing the substrate to the vapor of the reactive chemical precursors [34.30]. A schematic of the vapor deposition system is shown in Fig. 34.5 [34.59, 60, 73]. Samples are placed in the quartz reaction tube. The silane bubbler is used for introducing gas-phase silane into the quartz reaction tube placed in an oven at a controlled temperature. An inert gas flow (N_2) is used as a carrier gas. A by-products condenser is used for trapping the by-products and/or nonreacted silanes.

Research on some SAMs has been widely reported. SAMs of long-chain fatty acids $C_nH_{2n+1}COOH$ or $(CH_3)(CH_2)_nCOOH$ ($n = 10$, 12, 14 or 16) on glass or alumina substrates have been widely studied since the 1950s [34.25, 26, 30, 31, 76]. Probably the most studied SAMs to date are *n*-alkanethiolate (*n*-alkyl and *n*-alkane are used interchangeably) monolayers $CH_3(CH_2)_nS$-prepared from the adsorption of alkanethiol $-(CH_2)_nSH$ solution onto an Au film [34.45, 77–81] and *n*-alkylsiloxane monolayers. Siloxane (Si–O–Si) refers to the bond and silane (Si_nX_{2n+2}, which includes covalently bonded compounds containing the elements Si and other atoms or groups such as H and Cl to form SiH_4 and $SiCl_4$ respectively) refers to the head group of the precursor. These terms are used interchangeably. Monolayers are produced by adsorption of *n*-alkyltrichlorosilane $-(CH_2)_nSiCl_3$ onto a hydroxylated Si/SiO_2 substrate with siloxane (Si–O–Si) binding; see Fig. 34.6 [34.53, 58–63, 82]. *Tambe* and *Bhushan* [34.53], *Bhushan* et al. [34.54], *Hoque* et al. [34.55, 56, 65–67, 83]. *DeRose* et al. [34.57] have produced perfluoroalkylsilane and perfluoroalkylphosphonate on Al and Cu surfaces.

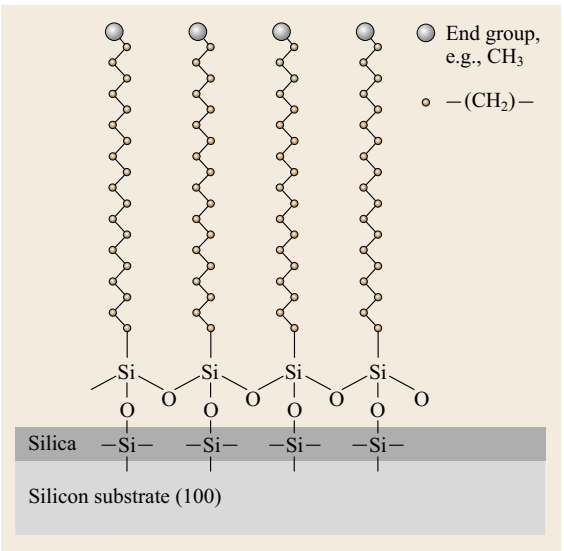


Fig. 34.6 Schematics of a methyl-terminated, *n*-alkylsiloxane monolayer on Si/SiO_2

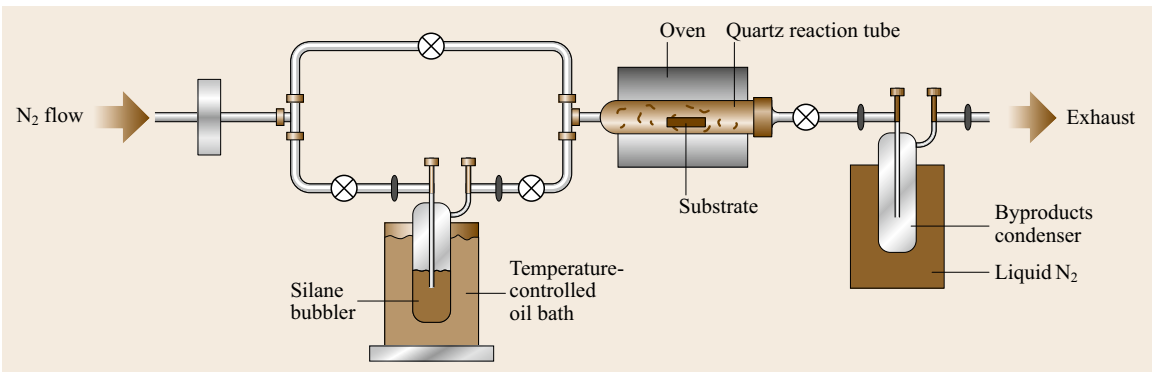


Fig. 34.5 Schematic showing vapor-phase deposition system for silane SAMs (after [34.59])

34.4 Contact Angle and Nanotribological Properties of SAMs

The basis for the molecular design and tailoring of SAMs should start from a complete knowledge of the interrelationships between the molecular structure and contact angle and nanotribological properties of SAMs, as well as a deep understanding of the adhesion, friction, and wear mechanisms of SAMs at the molecular level. Friction and wear studies of SAMs have been carried out on macro- and nanoscales. Macroscale tests are conducted using a so-called pin-on-disk tribotester apparatus in which a ball specimen slides against a lubricated flat specimen [34.14, 15]. Nanoscale tests are conducted using an atomic force/friction force microscope (AFM/FFM) [34.1, 4, 14, 15, 84]. In the AFM/FFM experiments, a sharp tip of a radius ranging from about 5 to 50 nm slides against a SAM specimen. A Si_3N_4 tip is commonly used for friction studies, and a Si or natural diamond tip is commonly used for scratch, wear, and indentation studies.

In early studies, the effect of chain length of the carbon atoms of fatty acid monolayers on the coefficient of friction and wear on the macroscale was studied by *Bowden* and *Tabor* [34.25] and *Zisman* [34.26]. *Zisman* [34.26] reported that for the monolayers deposited on a glass surface sliding against a stainless steel surface, there is a steady decrease in friction with increasing chain length. At a significantly long chain length, the coefficient of friction reaches a lower limit; Fig. 34.7a. He further reported that monolayers having a chain length below 12 carbon atoms behave as liquids (poor durability), and those with a chain length of 12–15 carbon atoms behave like a plastic solid (medium durability), whereas those with a chain length above 15 carbon atoms behave like a crystalline solid (high durability). Investigations by *Ruhe* et al. [34.86] indicated that the lifetime of the alkylsilane monolayer coating on a silicon surface increases greatly with an increase in the chain length of the alkyl substituent. *DePalma* and *Tillman* [34.87] showed that a monolayer of *n*-octadecyltrichlorosilane (n-$C_{18}H_{37}SiCl_3$, OTS) is an effective lubricant on silicon.

With the development of AFM techniques, researchers have successfully characterized the nanotribological properties of self-assembled monolayers [34.1, 4, 6, 84]. Studies by *Bhushan* et al. [34.58] showed that C_{18} alkylsiloxane films exhibit the lowest coefficient of friction and can withstand a much higher normal load during sliding as compared to L–B films, soft Au films, and hard SiO_2 coatings. *McDermott* et al. [34.85] studied the effect of the length of alkyl chains on the frictional properties of methyl-terminated *n*-alkylthiolate $CH_3(CH_2)_nS-$ films chemisorbed on Au(111) using an AFM. They reported that the longer-chain monolayers exhibit a markedly lower friction and a reduced propensity to wear than shorter-chain monolayers; Fig. 34.7b. These results are in good agreement with the macroscale results by *Zisman* [34.26]. They also conducted infrared reflection spectroscopy to measure the bandwidth of the methylene stretching mode ($\nu_a(CH_2)$), which exhibits a qualitative correlation with the packing density of the chains. It was

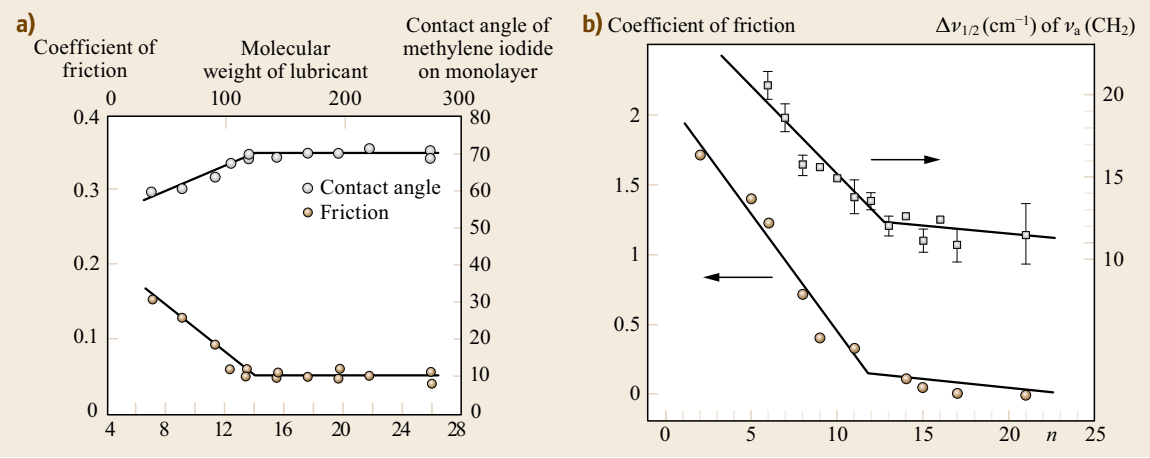


Fig. 34.7 **(a)** Effect of chain length (or molecular weight) on coefficient of macroscale friction of stainless steel sliding on glass lubricated with a monolayer of fatty acid, and contact angle of methyl iodide on condensed monolayers of fatty acid on glass (after [34.26]). **(b)** Effect of chain length of methyl-terminated *n*-alkanethiolate over Au film $AuS(CH_2)_nCH_3$ on the coefficient of microscale friction and peak bandwidth at half maximum ($\Delta\nu_{1/2}$) for the bandwidth of the methylene stretching mode ($\nu_a(CH_2)$) (after [34.85])

found that the chain structures of monolayers prepared with longer chain lengths are more ordered and more densely packed in comparison to those of monolayers prepared with shorter chain lengths. They further reported that the ability of the longer-chain monolayers to retain molecular-scale order during shear leads to a lower observed friction. Monolayers having a chain length of more than 12 carbon atoms, preferably 18 or more, are desirable for tribological applications. (Incidentally, monolayers with 18 carbon atoms, octadecanethiol films, have been widely studied.)

Xiao et al. [34.88] and *Lio* et al. [34.89] also studied the effect of the length of the alkyl chains on the frictional properties of *n*-alkanethiolate films on gold and *n*-alkylsilane films on mica. Friction was found to be particularly high with short chains of less than eight carbon atoms. Thiols and silanes exhibit similar friction forces for the same n when $n > 11$, while for $n < 11$, silanes exhibit higher friction, larger than that for thiols by a factor of about 3 for $n = 6$. The increase in friction was attributed to the large number of dissipative modes in the less ordered chains that occurs when going from a thiol to a silane anchor or when decreasing n. Longer chains ($n > 11$), stabilized by van der Waals attraction, form more compact and rigid layers and act as better lubricants. *Schonherr* and *Vancso* [34.90] also correlated the magnitude of friction with the order of the alkane chains. The disorder of the short-chain hydrocarbon disulfide SAMs was found to result in a significant increase in the magnitude of friction.

Tsukruk and *Bliznyuk* [34.91] studied the adhesion and friction between a Si sample and a Si_3N_4 tip, in which both surfaces were modified by $-CH_3$-, $-NH_2$- and $-SO_3H$-terminated silane-based SAMs. Various polymer molecules for the backbone were used. They reported that a very broad maximum adhesive force in a pH range from 4 to 8 with a minimum adhesion at pH > 9 and pH < 3 were observed for all of the studied mating surfaces. This observation can be understood by considering a balance of electrostatic and van der Waals interaction between composite surfaces with multiple isoelectric points. The friction coefficient of NH_2/NH_2 and SO_3H/SO_3H mating SAMs was very high in aqueous solutions. Cappings of NH_2-modified surfaces (3-aminopropyltriethoxysilane) with rigid and soft polymer layers resulted in a significant reduction in adhesion to a level lower than that of untreated surface [34.92]. *Fujihira* et al. [34.93] studied the influence of surface terminal groups of SAMs and functional tips on adhesive force. It was found that the adhesive forces measured in air increase in the order of CH_3/CH_3, CH_3/COOH, and COOH/COOH.

Bhushan and *Liu* [34.79], *Liu* et al. [34.80], and *Liu* and *Bhushan* [34.81, 94] studied adhesion, friction and wear properties of alkylthiol and biphenylthiol SAMs on Au(111) films. They explained the friction mechanisms using a molecular spring model in which local stiffness and intermolecular forces govern the friction properties. They studied the influence of relative humidity, temperature, and velocity on adhesion and friction. They also investigated the wear mechanisms of SAMs by a continuous microscratch AFM technique.

Fluorinated carbon (fluorocarbon) molecules are known to have low surface energy and are commonly used for lubrication [34.14, 15]. *Bhushan* and *Cichomski* [34.73] deposited fluorosilane SAMs on polydimethylsiloxane (PDMS). To make the hydrophobic PDMS surface chemically active, it was oxygenated using an oxygen plasma, which introduces silanol groups (SiOH). They reported that SAM-coated PDMS was more hydrophobic with lower adhesion, friction, and wear. *Bhushan* et al. [34.59, 60], *Kasai* et al. [34.61], *Lee* et al. [34.62], *Tambe* and *Bhushan* [34.53] and *Tao* and *Bhushan* [34.64] studied the adhesion, friction, and wear of methyl- and/or perfluoro-terminated alkylsilanes on silicon. They reported that perfluoroalkylsilane SAMs exhibited lower surface energy, higher contact angle, lower adhesive force, and lower wear as compared to that of alkylsilanes. *Kasai* et al. [34.61] also reported the influence of relative humidity, temperature, and velocity on adhesion and friction. *Tao* and *Bhushan* [34.63] studied degradation mechanisms of alkylsilanes and perfluoroalkylsilane SAMs on Si. They reported that oxygen in the air causes thermal oxidation of SAMs.

Tambe and *Bhushan* [34.53], *Bhushan* et al. [34.54], *Hoque* et al. [34.55, 56] and *DeRose* et al. [34.57] studied the nanotribological properties of methyl- and perfluoro-terminated alkylphosphonate, perfluorodecyldimethylchlorosilane, and perfluorodecanoic acid on aluminum, which is of industrial interest. *Hoque* et al. [34.83] and *DeRose* et al. [34.57] studied the nanotribological properties of alkylsilanes and perfluroalkylsilanes on aluminum. *Hoque* et al. [34.65–67] studied the nanotribological properties of alkylphosphonate and perfluoroalkylsilane SAMs on copper. The authors found that these SAMs on aluminum and copper perform well irrespective of the substrate used. They confirmed the presence of respective films using x-ray photoelectron spectroscopy (XPS).

Hoque et al. [34.65–67] studied the chemical stability of various SAMs deposited on Cu substrates via exposure to various corrosive conditions. *DeRose* et al. [34.57] studied the chemical stability of various SAMs deposited on Al substrates via exposure to corrosive conditions (aqueous nitric acid solutions of a low pH of 1.8 at temperatures ranging from 60 to 80 °C for times ranging from 30 to 70 min). The exposed samples

were characterized by XPS and contact angle measurements. They reported that perfluorodecanoic acid/Al is less stable than perfluorodecylphosphonate/Al and octadecylphosphonate/Al, but more stable than perfluorodecyldimethylchlorosilane/Al, which has implications in DMD applications, discussed earlier. In general, chemical stability data of various SAMs deposited on Cu and Al surfaces to corrosive environments have been reported by these authors. Based on these studies, it was concluded that chemisorption occurs at the interface and is responsible for strong interfacial bonds.

To date, the contact angle and nanotribological properties of alkanethiol, biphenylthiol, alkylsilane, perfluoroalkylsilane, alkylphosphonate and perfluoroalkylphosphane SAMs have been widely studied. In this chapter, we review, in some detail, the nanotribological properties of various SAMs having alkyl and biphenyl spacer chains with different surface terminal groups ($-CH_3$, $-CF_3$) and head groups ($-S-H$, $-Si-O-$, $-OH$, and $P-O-$), which have been investigated by AFM at various operating conditions; Fig. 34.8a,b [34.53, 54, 59, 61, 63, 79–81, 94]. Hexadecane thiol (HDT), 1,1′-biphenyl-4-thiol (BPT), and cross-linked BPT (BPTC) were deposited on Au(111) films on Si(111) substrates by immersing the substrate in a solution containing the precursor (ligand) that is reactive to the substrate surface. Crosslinked BPTC was produced by irradiation of BPT monolayers with low-energy electrons. Perfluoroalkylsilane and alkylsilane SAMs were deposited on Si(100) by exposing the substrate to the vapor of the reactive chemical precursors. Perfluoroalkylphosphonate and alkylphosphonate SAMs were deposited on sputtered Al film on Si substrate as well as bulk Al substrates. Representative data follow.

34.4.1 Measurement Techniques

Experimental techniques used for measurement of the static contact angle, surface energy, adhesion, friction, and wear are described next.

Static Contact Angle and Surface Energy Measurements Using DI Water

The static contact angle, a measure of water repellency, was measured using a Rame–Hart model 100 contact angle goniometer (Mountain Lakes, NJ, USA) [34.95, 96]. Typically, 10 µl droplets of DI water were used for making contact angle measurements. At least two measurements of the contact angle were made and were

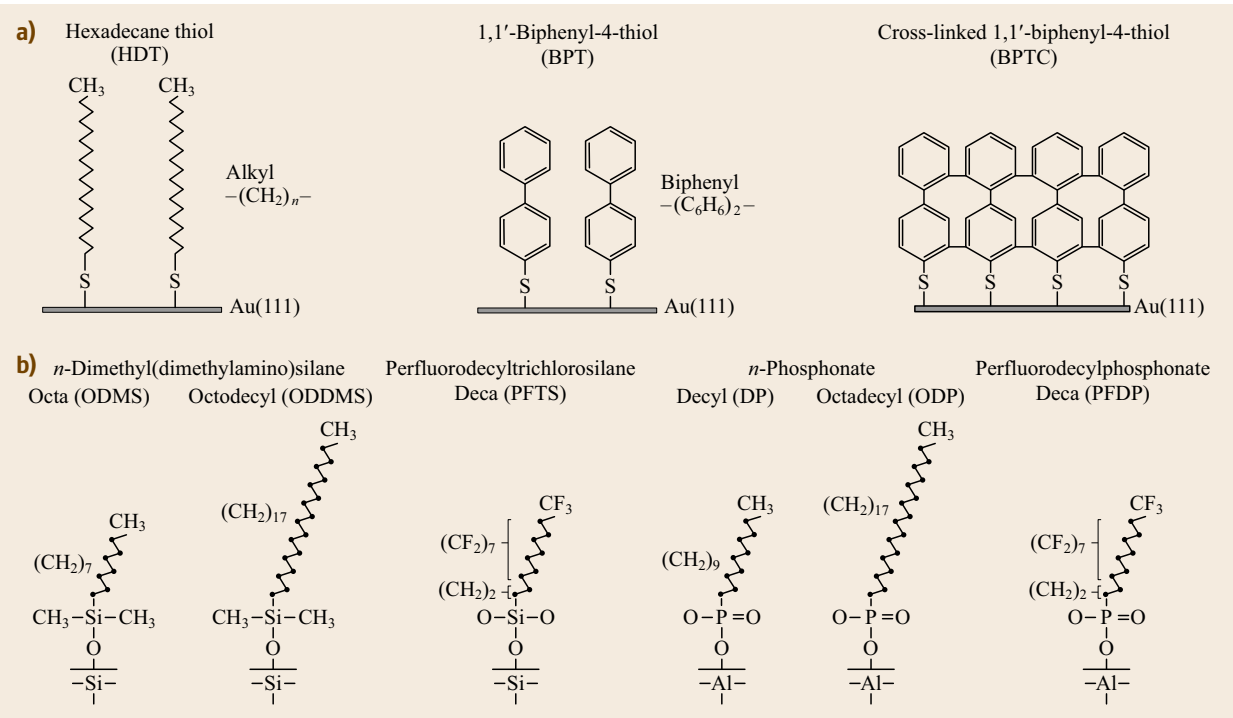


Fig. 34.8a,b Schematics of the structures of **(a)** hexadecane and biphenyl thiol SAMs on Au(111) substrates, and **(b)** perfluoroalkylsilane and alkylsilane SAMs on Si with native oxide substrates, and perfluoroalkylphosphonate and alkylphosphonate SAMs on Al with native oxide

found to be reproducible within $\pm 2^\circ$. The critical surface tension, a measure of interfacial surface energy, was obtained from the so-called Zisman plot. Contact angles of SAMs with liquids with a range of surface tensions were measured. The cosines of the contact angles were plotted as a function of surface tension of the various *n*-alkane liquids (hexadecane, dodecane, undecane and decane) used. The plot is a linear line for low-polarizable SAMs. The horizontal intercept of the line passing through cos(contact angle)= 1 provides the critical surface tension, which is a measure of the surface energy of the SAM [34.59].

AFM Adhesion and Friction Measurements

The adhesion and friction tests were conducted using a commercial AFM system. Square pyramidal Si_3N_4 tips with a 30–50 nm tip radius on gold back-coated triangular Si_3N_4 cantilevers with a typical spring constant of 0.58 N/m were used. Adhesion can be calculated using either force calibration plots or from the negative intercepts on the friction force versus normal plot loads. Both methods generally yield similar results. The force calibration plot technique was used in this study. The coefficient of friction was obtained from the slope of the friction force versus normal load plots. Normal loads typically ranged from 5 to 100 nN. Friction force measurements were generally performed at a scan rate of 1 Hz along the fast scan axis and over a scan size of 2 μm by 2 μm. The fast scan axis was perpendicular to the longitudinal direction of the cantilever. The friction force was calibrated by the method described in [34.4, 6, 84].

AFM Wear Measurements

Wear tests were conducted using a diamond tip with a nominal radius of 50 nm and nominal cantilever stiffness of 10 N/m. Wear tests were performed on a 1 μm × 1 μm scan area at the desired normal load and at a scan rate of 1 Hz. After each wear test, a 3 μm × 3 μm area was imaged, and the average wear depth was calculated.

Effect of Relative Humidity, Temperature, and Sliding Velocity

The influence of relative humidity on adhesive force, friction force, and wear was studied in an environmentally controlled chamber. Relative humidity was controlled by introducing a mixture of dry and moist air inside the chamber. The temperature was maintained at $22 \pm 1\,^\circ$C. The sample was kept in the environmental chamber at desired humidity for at least 2 h prior to the tests so that the system could reach equilibrium conditions.

In order to study the effect of temperature on adhesion and friction force, the samples were placed on a thermal stage during the measurements. A glass plate was placed under the thermal stage to prevent the heat from being transported away. The temperature range studied was from 20 to 110 °C. The relative humidity was maintained at $50 \pm 5\%$ during the temperature effect measurements.

The effect of sliding velocity on friction force was monitored in ambient conditions using a high-velocity piezoelectric stage designed for achieving high relative sliding velocities on a commercial AFM setup [34.97]. The traveling distance of the sample, i. e., the scan size, was set at 25 μm, while the scan frequency was varied between 0.1 Hz (5 μm/s) and 100 Hz (5000 μm/s).

Chemical Degradation and Environmental Studies

The chemical degradation experiments were carried out in a high-vacuum tribotest apparatus [34.98, 99]. The system was equipped with a mass spectrometer so that gaseous emissions from the interface could be monitored in situ during the sliding in high vacuum and other controlled environments. The normal loads and friction forces at the contacting interface were measured using resistive-type strain-gage transducers. For the sliding tests, the coated flat sample was glued on a flat surface at the end of a rotating shaft. The sample was slid against a Si(100) wafer mounted on the flat surface of a slider integrated with a flexible cantilever used in magnetic rigid disk drives. The sliding speed used was 0.3 m/s, and the applied pressure was 50 kPa. The environmental effects were investigated in high vacuum (2×10^{-7} Torr), argon, dry air (less than 2% relative humidity (RH)), ambient air (30% RH), and high-humidity air (70% RH).

34.4.2 Hexadecane Thiol and Biphenyl Thiol SAMs on Au(111)

Hexadecane thiol on Au(111) film was selected as it is a widely studied film. Biphenyl thiol was selected to study the effect of rigidity on nanotribological performance. Biphenyl thiol film was cross-linked to further increase its stiffness.

Surface Roughness, Adhesion, and Friction

Surface height and friction force images of SAMs were recorded simultaneously on an area of 1 μm × 1 μm by an AFM, and adhesive forces were measured by using the force calibration mode in an AFM [34.79].

For further analysis, presented later in this chapter, the measured roughness, thickness, tilt angles, and spacer chain lengths of Si(111), Au(111) and various SAMs are listed in Table 34.10 [34.79]. The roughness of BPT is very close to that of Au(111), but the

Table 34.10 The R_a roughness, thickness, tilt angles and spacer chain lengths of SAMs

Samples	R_a roughness[a] (nm)	Thickness[b] (nm)	Tilt angle[b] (deg)	Spacer length[c] (nm)
Si(111)	0.07			
Au(111)	0.37			
HDT	0.92	1.89	30	1.91
BPT	0.36	1.25	15	0.89
BPTC	0.14	1.14	25	0.89

[a] Measured by an AFM with 1 μm × 1 μm scan size, using Si_3N_4 tip under 3.3 nN normal load
[b] The thickness and tilt angles of BPT and BPTC are reported by *Geyer* et al. [34.78]. The thickness and tilt angles of HDT are reported by *Ulman* [34.31]
[c] The spacer chain lengths of alkylthiols were calculated by the method reported by *Miura* et al. [34.100]. The spacer chain lengths of biphenyl thiols were calculated by the data reported by *Ratajczak-Sitarz* et al. [34.101]

roughness of BPTC is lower than that of Au(111) and BPT; this is caused by electron irradiation. Table 34.10 indicates that the roughness value of HDT is much higher than its substrate roughness of Au(111). This is caused by local aggregation of organic compounds on the substrates during SAM deposition. Table 34.10 also indicates that the thicknesses of biphenyl thiol SAMs are generally smaller than the alkylthiol, because of shorter spacer chains in biphenyl thiol.

The average values and standard deviation of the adhesive force and coefficient of friction are presented in Fig. 34.9 [34.79]. Based on the data, the adhesive force and coefficient of friction of SAMs are less than

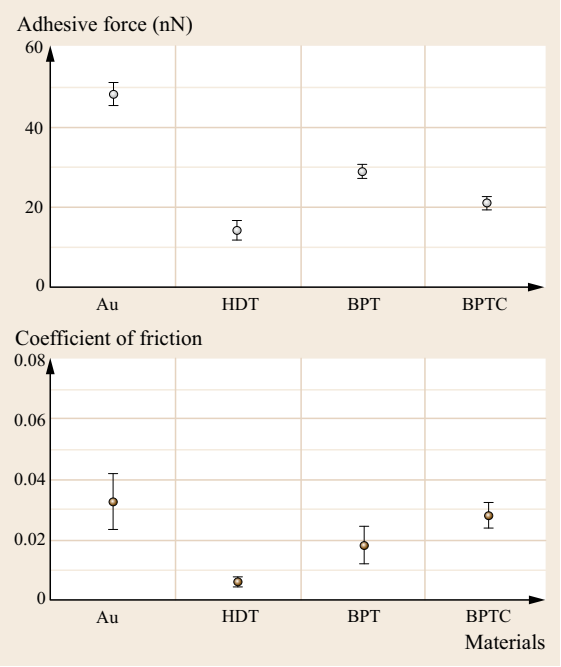


Fig. 34.9 Adhesive forces and coefficients of friction of Au(111) and various SAMs

those of their corresponding substrates. Among various films, HDT exhibits the lowest values. The ranking of adhesive forces F_a is in the following order: $F_{a\text{-Au}} > F_{a\text{-BPT}} > F_{a\text{-BPTC}} > F_{a\text{-HDT}}$, and the ranking of the coefficients of friction μ is in the following order: $\mu_{Au} > \mu_{BPTC} > \mu_{BPT} > \mu_{HDT}$. The ranking of various SAMs for adhesive force and coefficient of friction are similar. It suggests that alkylthiol and biphenyl thiol SAMs can be used as effective molecular lubricants for micro/nanodevices.

In micro/nanoscale contact, liquid capillary condensation is one of the sources for the adhesion and friction. In the case of a sphere in contact with a flat surface, the attractive Laplace force caused by water capillary is [34.14, 15],

$$F_L = 2\pi R\gamma_{la}(\cos\theta_1 + \cos\theta_2)\,, \tag{34.1}$$

where R is the radius of the sphere, γ_{la} is the surface tension of the liquid against air, and θ_1 and θ_2 are the contact angles between liquid and flat and spherical surfaces respectively. In an AFM adhesive study, the tip-flat sample contact is just like a sphere in contact with a flat surface, and the liquid is water. Since a single tip was used in the adhesion measurements, $\cos\theta_2$ can be treated as a constant. Therefore,

$$\begin{aligned} F_L &= 2\pi R\gamma_{la}(1+\cos\theta_1) - 2\pi R\gamma_{la}(1-\cos\theta_2) \\ &= 2\pi R\gamma_{la}(1+\cos\theta_1) - C\,, \end{aligned} \tag{34.2}$$

where C is a constant.

Based on the following Young–Dupre equation, the work of adhesion W_a (the work required to pull apart the unit area of the solid-liquid interface) can be written as [34.12]

$$W_a = \gamma_{la}(1+\cos\theta_1)\,. \tag{34.3}$$

It indicates that W_a is determined by the contact angle of SAMs, i. e., is influenced by the surface chemistry properties (polarization and hydrophobicity) of SAMs. By substituting (34.3) into (34.2), F_L can be expressed as

$$F_L = 2\pi R W_a - C \,. \tag{34.4}$$

When the influence of other factors, such as van der Waals force, are very small on the adhesive force, then adhesive force $F_a \approx F_L$. Thus the adhesive force F_a should be proportional to the work of adhesion W_a.

Contact angle is a measure of the wettability of a solid by a liquid and determines the W_a value [34.95, 96]. The static contact angles of distilled water on Au(111) and SAMs were measured and are summarized in Fig. 34.10a [34.79]. For water, $\gamma_{la} = 72.6\,\mathrm{mJ/m^2}$ at 22 °C, and by using (34.3), the W_a data are obtained and presented in Fig. 34.10b. The W_a can be ranked in the following order: $W_{a\text{-}Au}(97.1) > W_{a\text{-}BPT}(86.8) > W_{a\text{-}BPTC}(82.1) > W_{a\text{-}HDT}(61.4)$. Except $W_{a\text{–}Au}$, this order exactly matches the order of adhesion force in Fig. 34.9. The relationship between F_a and W_a is summarized in Fig. 34.11 [34.79]. It indicates that the adhesive force F_a (nN) increases with the work of adhesion $W_a(\mathrm{mJ/m^2})$ by the linear relationship

$$F_a = 0.57 W_a - 22 \,. \tag{34.5}$$

These experimental results agree well with the modeling prediction presented earlier in (34.4). It proves that on the nanoscale at ambient conditions the adhesive force of SAMs is mainly influenced by the water capillary force. Though neither HDT nor BPT has polar surface groups, the surface terminal of HDT has a symmetrical structure, which causes a smaller electrostatic attractive force and yields a smaller adhesive force than BPT. It is believed that the easy attachment of Au onto the tip should be one of the reasons behind the large adhesive force, which does not fit the linear relationship described by (34.5).

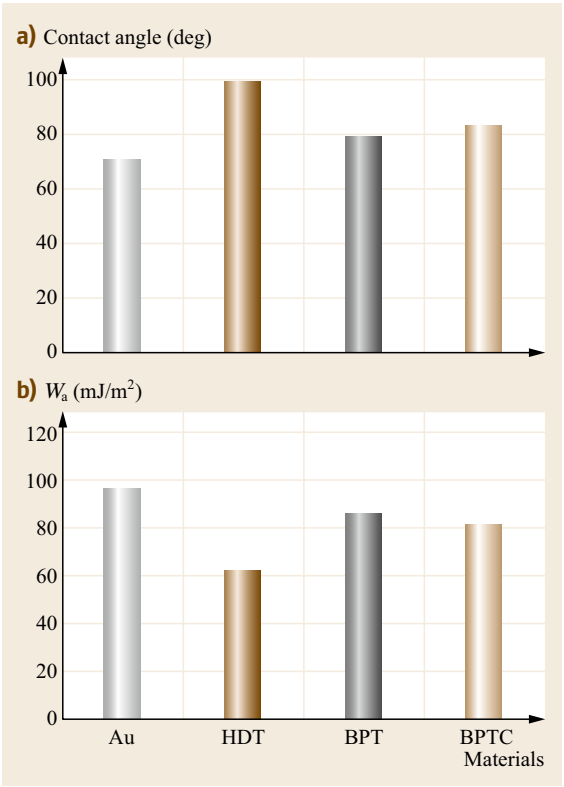


Fig. 34.10 (a) The static contact angle, and (b) work of adhesion of Au(111) and various SAMs. All of the points in this figure represent the mean value of six measurements. The uncertainty associated with the average contact angle is within ±2°

Stiffness, Molecular Spring Model, and Micropatterned SAMs

Next the friction mechanisms of SAMs were examined. The Monte Carlo simulation of the mechanical relaxation of a $CH_3(CH_2)_{15}SH$ self-assembled monolayer performed by *Siepman* and *McDonald* [34.102] indicated that SAMs compress and respond nearly elastically to microindentation by an AFM tip when the load is below a critical normal load. Compression can lead to major changes in the mean molecular tilt (i. e., orientation), but the original structure is recovered as the normal load is removed. *Garcia-Parajo* et al. [34.103] also reported the compression and relaxation of octadecyltrichlorosilane (OTS) film during loading and unloading.

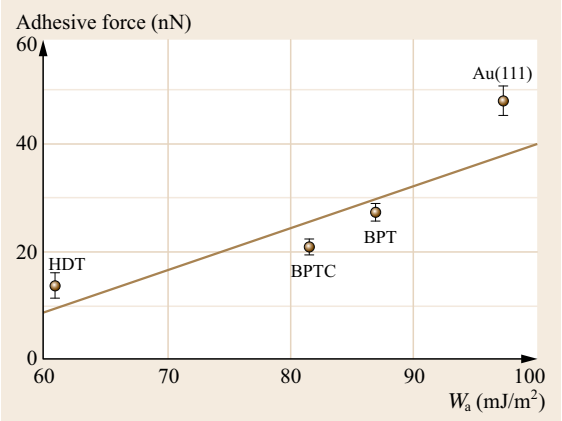


Fig. 34.11 Relationship between the adhesive forces and work of adhesion for different specimens

To study the difference in the stiffness of various films, the stiffness properties were measured by an AFM in force modulation mode [34.4, 6, 81, 84]. They reported that BPT was stiffer than HDT. Since BPT has a rigid benzene structure, it is more difficult to compress than HDT. Figure 34.12 shows the variation of the displacement with normal load during indentation mode. It clearly indicates that SAMs can be compressed. At a given normal load, long carbon chain structure SAMs such as HDT are easily compressed as compared to rigid benzene ring structure SAMs such as BPT, which implies that BPT is more rigid than HDT.

In order to explain the frictional difference of SAMs, based on the friction and stiffness measurements by AFM and the Monte Carlo simulation, a molecular spring model is presented in Fig. 34.13. It is believed that the self-assembled molecules on a substrate are just like assembled molecular springs anchored to the substrates [34.79]. An AFM tip sliding on the surface of SAMs is like a tip sliding on the top of *molecular springs or brush*. The molecular spring assembly has compliant features and can experience compression and orientation under normal load. The orientation of the molecular springs or brush reduces the shearing force at the interface, which in turn reduces the friction force. The possibility of orientation is determined by the spring constant of a single molecule (local stiffness), as well as the interaction between the neighboring molecules, which can be reflected by packing density or packing energy. It should be noted that the orientation can lead to conformational defects along the molecular chains, which leads to energy dissipation. In the study of BPT by AFM, it was found that after the first several scans, the friction force is significantly reduced, but the surface height does not show any apparent change. This suggests that the molecular orientation can be facilitated by initial sliding and is reversible [34.94].

Based on the stiffness measurements obtained using Fig. 34.12 and the view of molecular structures in Fig. 34.13, biphenyl is a more rigid structure due to the contribution of two rigid benzene rings. Therefore the spring constant of BPT is larger than that of HDT. The hydrogen (H^+) in a biphenyl chain has an electrostatic attractive force with the π electrons in the neighboring benzene ring. Thus the intermolecular force between biphenyl chains is stronger than that for alkyl chains. The larger spring constant and stronger intermolecular force of BPT require a larger external force to allow it to orient, thus causing a higher coefficient of friction. The cross-linking of BPT leads to a larger packing energy for BPTC. Therefore it requires a larger external force to allow BPTC orientation, i. e., the coefficient of friction of BPTC is higher than that of BPT.

An elegant way to demonstrate the influence of molecular stiffness on friction is to investigate SAMs with different structures on the same wafer. For this purpose, a micropatterned SAM was prepared. First the biphenyldimethylchlorosilane (BDCS) was deposited on silicon by a typical self-assembly method [34.81]. Then the film was partially cross-linked using a mask technique by low-energy electron irradiation. Finally the micropatterned BDCS films were realized, which had the as-deposited and cross-linked coating regions on the same wafer. The local stiffness properties of these micropatterned samples were investigated by the force modulation AFM technique [34.104]. The variation in the deflection amplitude provides a measure

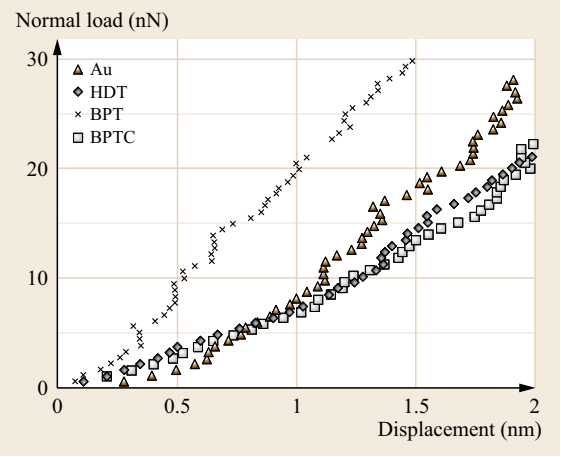


Fig. 34.12 Normal load versus displacement curves of Au(111) and various SAMs

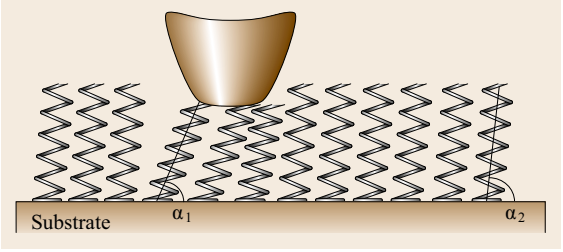


Fig. 34.13 Molecular spring model of SAMs. In this figure, $\alpha_1 < \alpha_2$, which is caused by the orientation under the normal load applied by an AFM tip. The orientation of the molecular springs reduces the shearing force at the interface, which in turn reduces the friction force. The molecular spring constant, as well as the intermolecular forces can determine the magnitude of the coefficients of friction of SAMs. In this figure, the size of the tip and molecular springs are not exactly to scale (after [34.79])

of the relative local stiffness of the surface. Surface height, stiffness, and friction images of the micropatterned biphenyldimethylchlorosilane (BDCS) specimen were obtained and are presented in Fig. 34.14. The circular areas correspond to the as-deposited film and the remaining area to the cross-linked film. Figure 34.14a indicates that cross-linking caused by the low-energy electron irradiation leads to about a 0.5 nm decrease of the surface height of BDCS films. The corresponding stiffness images indicate that the cross-linked area has a higher stiffness than the as-deposited area. Figure 34.14b indicates that the as-deposited area (area with higher surface height) has a lower friction force. Obviously, these data on the micropatterned sample proves that the local stiffness of SAMs has an influence on their friction performance. Higher stiffness leads to larger friction force. These results provide strong proof of the suggested molecular spring model.

In summary, it is found that SAMs exhibit compliance and can experience compression and orientation under normal load. The orientation of SAMs reduces the shear stress at the interface; therefore SAMs can serve as good lubricants. The molecular spring constant (local stiffness), as well as the intermolecular forces, can influence the magnitude of the coefficients of friction of SAMs.

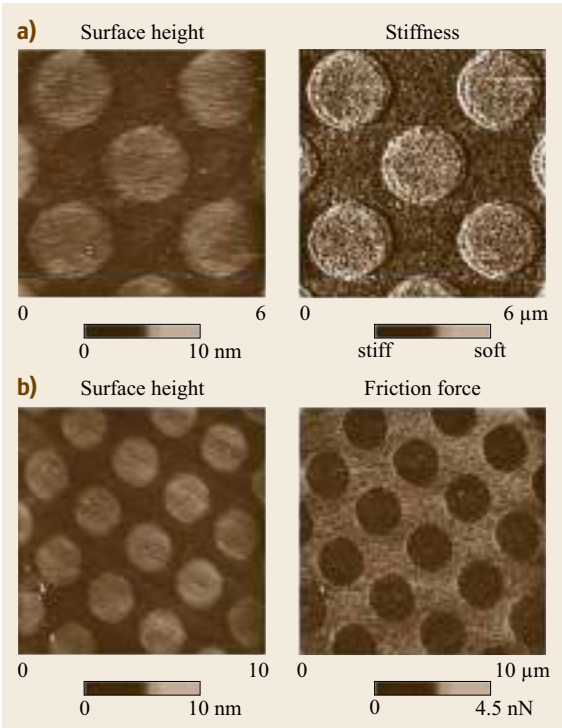


Fig. 34.14 (a) AFM grayscale surface height and stiffness images, and (b) AFM grayscale surface height and friction force images of micropatterned BDCS (after [34.81])

Wear and Scratch Resistance

Wear resistance was studied on an area of $1\,\mu m \times 1\,\mu m$. The variation of wear depth with normal loads is presented in Fig. 34.15 [34.79]. HDT exhibits the best wear resistance. For all of the tested SAMs, in the wear depth as a function of normal load curves, there appears a critical normal load, which is marked by arrows in Fig. 34.15. When the normal load is smaller than the critical normal load, the monolayer only shows a slight height change in the scan areas. When the normal load is higher than the critical value, the height change of SAM increases dramatically. Relocation and accumulation of BPT molecules has been observed during the initial several scans, which leads to the formation of a larger terrace. Wear studies of a single BPT terrace indicate that the wear life of BPT increases exponentially with terrace size [34.80, 81].

The scratch resistance of Au(111) and SAMs was studied by a continuous AFM microscratch technique. Figure 34.16a shows coefficient of friction profiles as a function of increasing normal load and corresponding tapping mode AFM surface height images of the scratches captured on Au(111) and SAMs [34.81]. Figure 34.16a indicates that there is an abrupt increase in the coefficient of friction for all of the tested samples. The normal load associated with this event is termed the critical load (indicated by the arrows labeled *A*). At the initial stages of the scratch, all the samples exhibit a low coefficient of friction indicating that the friction force is dominated by the shear component. This is in agreement with analysis of the AFM images, which shows negligible damage on the surfaces prior to the critical load. At the critical load, a clear groove is formed, which is accompanied by the formation of

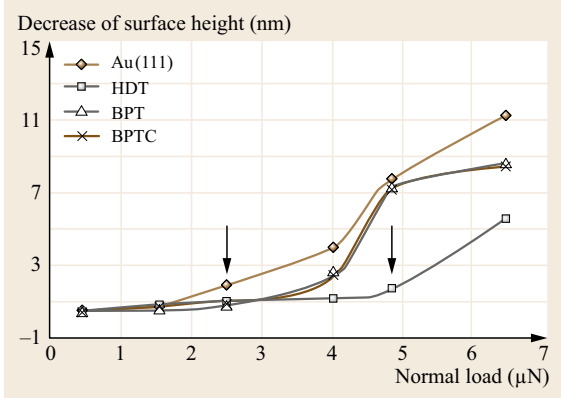


Fig. 34.15 Wear depth as a function of normal load after one scan cycle

material pileup at the sides of the scratch. This suggests that the initial damage that occurs at the critical load is due to plowing associated with plastic deformation, and this causes the sharp rise in the coefficient of friction. Beyond the critical load, debris can be seen in addition to material pileup at the sides of the scratch. Figure 34.16b summarizes the critical loads for the various samples obtained in this study. It clearly indicates that all SAMs can increase the critical load of the corresponding substrate.

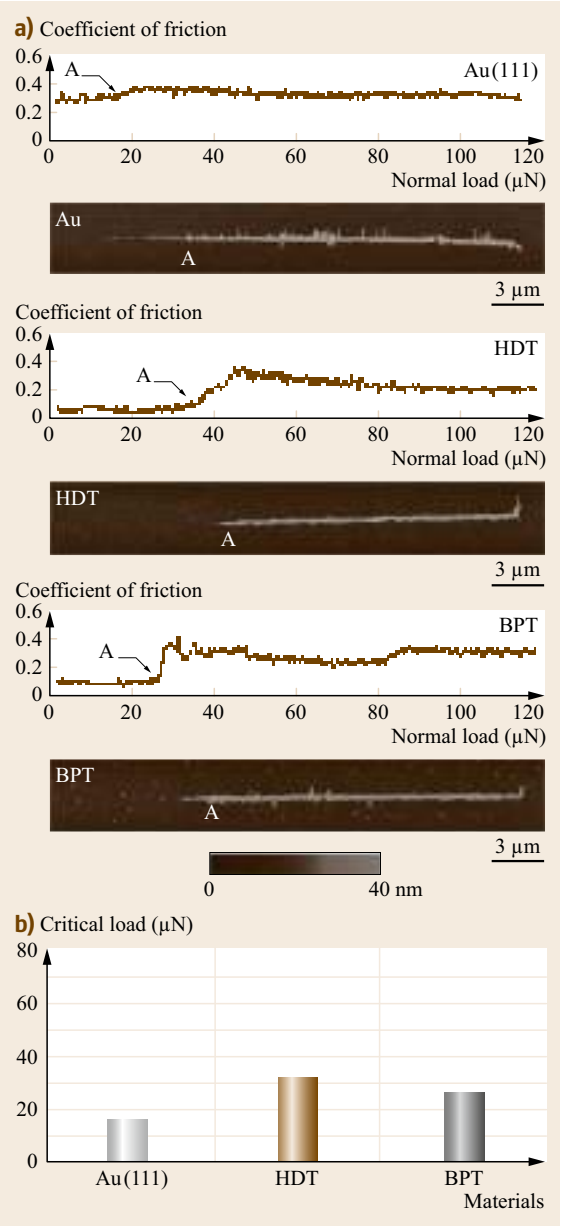


Fig. 34.16 (a) Coefficient of friction profiles during scratch as a function of normal load and corresponding AFM surface height images, and **(b)** critical loads estimated from the coefficient of friction profile and AFM images for Au(111), HDT/Au(111), and BPT/Au(111) (after [34.81])

The mechanisms responsible for a sudden drop in decrease in surface height with an increase in load during wear and scratch tests need to be understood. *Barrena* et al. [34.105] observed that the height of self-assembled alkylsilanes decreases in discrete amounts with normal load. This step-like behavior is due to the discrete molecular tilts, which are dictated by the geometrical requirements of the close packing of molecules. Only certain angles are allowed due to the zigzag arrangement of the carbon atoms. The relative height of the monolayer under pressure can be calculated by

$$\frac{h}{L} = \left[1 + \left(\frac{na}{d}\right)^2\right]^{-1/2}, \qquad (34.6)$$

where L is the total length of the molecule, h is the height of the SAMs in the tilt configuration (monolayer thickness), a is the distance between alternate carbon atoms in the molecule, d is the separation of the molecules, and n is the step number. The values of a (0.25 nm) and d (0.47 nm) are used in the calculation for HDT. The calculated and measured relative heights of HDT are listed in Table 34.11. When the normal loads are smaller than the critical values in Fig. 34.15, the measured relative height values of HDT are very close to the calculated values. It means that HDT underwent step tilting below critical normal loads.

The residual SAM thickness after wear under critical normal load was measured by profiling the worn film using an AFM. The results are listed in Table 34.12. For an alkanethiol monolayer, the relationship between

Table 34.11 Calculated $[1 + (na/d)^2]^{-1/2}$ and measured (h/L) relative heights of HDT self-assembled monolayers (after [34.79])

Steps (n)	Calculated[a] $[1 + (na/d)^2]^{-1/2}$	Measured (h/L)
1	0.883	
2	0.685	0.674[b]
3	0.531	0.532[c]
4	0.425	0.416[d]
5	0.352	0.354[e]
6	0.299	

[a] Calculations are based on the assumption that the molecules tilt in discrete steps (n), upon compression with a diamond AFM tip [34.105]
[b–e] These measured values correspond to the normal loads of 0.50, 1.57, 2.53, and 4.03 μN respectively

Table 34.12 Calculated L_0 and measured residual film thickness for SAMs under critical load

	L_0^a (nm)	Residual thickness[b] (nm)
HDT	0.24	0.25
BPT	0.39	0.42
BPTC	0.33	0.38

[a] Calculated by the equation of $h = b\cos(\alpha)n + L_0$ [34.100]
[b] Measured by AFM using diamond tip under critical normal load. All of the data are the mean value of three tests

the monolayer thickness h and intercept length L_0 can be expressed as (Fig. 34.17)

$$h = b\cos(\alpha)n + L_0\,, \tag{34.7}$$

where b is the length of the projection of the C–C bond onto the main chain axis ($b = 0.127$ nm for alkanethiol), n is the chain length defined by $CH_3(CH_2)_nSH$, and α is the tilt angle [34.100]. For BPT and BPTC, based on the same principle, and using the bond lengths reported in reference [34.101], the L_0 values are also calculated; Table 34.12. It indicates that the measured residual thickness values of SAMs under critical load are very close to the calculated intercept length L_0 values. It means that under the critical normal load, the Si_3N_4 tip approaches the interface, and SAMs wear away severely from the substrate. This is due to the interface chemical adsorption bond strength (S-Au) being generally smaller than the other chemical bond strengths in SAM spacer chains (Table 34.14, to be introduced later).

According to the wear and scratch results reported here and the above discussion, the transition of the wear mechanisms of SAMs with increasing normal load is illustrated in Fig. 34.18. Below the critical normal load SAMs undergo step orientation, at the critical load SAMs wear away from the substrate due to the weak interface bond strengths, while above the critical normal load severe wear take place on the substrate. In order to improve wear resistance, the interface bond must be enhanced; a rigid spacer chain and a hard substrate are also preferred.

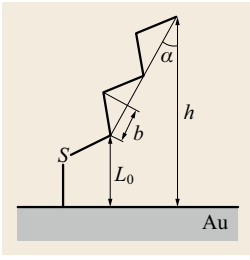


Fig. 34.17 Illustration of the relationship between the components of the equation $h = b\cos(\alpha)n + L_0$ (after [34.79])

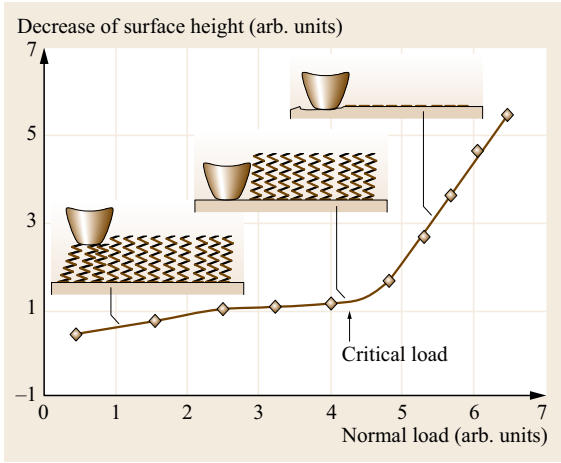


Fig. 34.18 Illustration of the wear mechanisms of SAMs with increasing normal load (after [34.81])

34.4.3 Perfluoroalkylsilane and Alkylsilane SAMs on Si(100), and Perfluoroalkylphosphonate and Alkylphosphonate SAMs on Al

Perfluorodecyltricholorosilane (PFTS),

$$CF_3{-}(CF_2)_7{-}(CH_2)_2{-}SiCl_3\,,$$

n-octyldimethyl (dimethylamino)silane (ODMS),

$$CH_3{-}(CH_2)_n{-}Si(CH_3)_2{-}N(CH_3)_2 \quad (n = 7)\,,$$

and *n*-octadecyldimethyl(dimethylamino)silane ($n = 17$) (ODDMS) vapor-deposited on Si(100) substrate and perfluorodecylphosphonate (PFDP)

$$CF_3{-}(CF_2)_7{-}(CH_2)_2{-}\underset{\underset{O}{\|}}{\overset{\overset{O}{|}}{P}}{-}OH\,,$$

decylphosphonate (DP)

$$CH_3{-}(CH_2)_n{-}\underset{\underset{O}{\|}}{\overset{\overset{O}{|}}{P}}{-}OH \quad (n = 9)\,,$$

and octadecylphosphonate (ODP) ($n = 17$) liquid-deposited on sputtered Al film on Si substrate were selected. Perfluoro SAMs were selected because fluorinated films are known to have low surface energy. Two chain lengths of alkylsilanes (with 8 and 18 carbon atoms) were selected to compare their nanotribological

performance with that of the former as well as to study the effect of chain length. Al substrate was selected because of the application of Al micromirrors in digital projection displays. Perfluoroalkylphosphane (with 10 carbon atoms) and alkylphosphonate SAMs (with 10 and 18 carbon atoms) on Al were selected.

Static Contact Angle and Surface Free Energy Measurements

Static contact angles and surface energy were measured to get a measure of the hydrophobicity. Figure 34.19 shows a Zisman plot for the SAMs deposited on Si and their surface energy values using various alkane liquids [34.59]. Critical surface tension, a measure of surface energy values for the two SAMs, are presented in the figure. Zisman analysis for the Si substrate was not available because the alkane liquids used for the measurement instantly spread on these surfaces. For PFTS, significant reduction in the critical surface tension or surface energy was observed (12.9 mN/m for PFTS/Si), as compared to ODMS (24.7 mN/m for ODMS/Si) and ODDMS (23.9 mN/m for ODDMS/Si). The surface energy for ODMS and ODDMS was comparable. It suggests that the surface was comparably covered by the SAMs without bare substrate appearing.

For comparison, static contact angles were measured for various SAMs on Si and Al substrates. The measured values are compared among the samples in Fig. 34.20a [34.54, 59]. A summary of root-mean-square (RMS) roughnesses measured by using an AFM, and static contact angles and film thickness measured using an ellipsometer, are summarized in Table 34.13 [34.54, 61]. Significant improvement

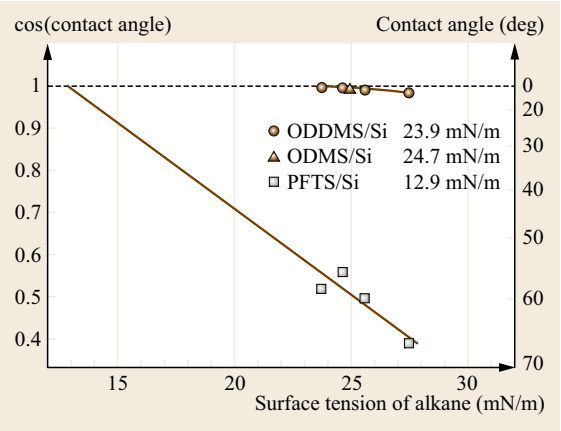


Fig. 34.19 Zisman plot for PFTS/Si, ODMS/Si, and ODDMS/Si used for calculating critical surface tension, a measure of surface energy, which is given by the *x*-intercept (cos(contact angle) = 1) of the fitted line to the data

in the water repellency was observed for perfluorinated SAMs as compared to bare Si and Al substrates. Static contact angles of alkylsilanes and alkylphosphonates were also higher than corresponding substrates, but lower than corresponding perfluorinated films. The contact angle generally increased with a decrease in surface energy [34.106], which is consistent with the data obtained. The contact angles can be influenced by the packing density as well as the sample roughness [34.107]. The higher contact angles for the SAMs deposited on Al substrates than those on Si substrate are probably due to this effect. The $-CH_3$ groups in ODMS, ODDMS, DP, and ODP are nonpolar and are known to contribute to the water repellency. Perfluorinated SAMs exhibited the highest contact angle among the SAMs tested in this study.

AFM Adhesion and Friction Measurements Under Ambient Conditions

Figure 34.20a shows the adhesive force, friction force, and the coefficient of friction measured under ambient conditions using an AFM, and Fig. 34.20b shows the friction force versus normal load plots for various SAMs deposited onto the Si and Al substrates [34.54, 59]. Figure 34.21 shows surface height and friction force maps for Si and PFTS, ODMS, and ODDMS on Si [34.59].

The bare substrates showed higher adhesive force than the SAMs coatings. ODMS and ODDMS shows an adhesive force comparable to DP and ODP despite their lower water contact angles. These SAMs have the same tail groups, and during AFM measurements the AFM tip interacts only with the tail groups, whereas the contact angles can also be influenced by the head groups in these SAMs. This is probably the reason why the adhesive forces for these SAMs are comparable. PFTS and PFDP, which have the highest contact angles, showed the lowest adhesion.

Friction force images of SAMs on Si exhibit more uniform contrast than those of bare Si. The coefficient of friction was higher for the bare substrates as compared to the corresponding SAMs deposited on them. The SAMs deposited on the Si substrate showed lower coefficients of friction than those deposited on the Al substrates. The primary reason for this is believed to be the higher roughness of the Al substrates. The SAMs with fluorocarbon backbone chains were found to have higher coefficients of friction than those with hydrocarbon backbone chains. This might be attributed to the higher stiffness of the fluorocarbon backbone [34.54, 61]. For the fluorocarbon backbone chains, it is harder to rotate the backbone structure due to the larger size of the F atoms in comparison to the H atoms [34.108]. The C–C bonds of hydrocarbon chains can, on the other

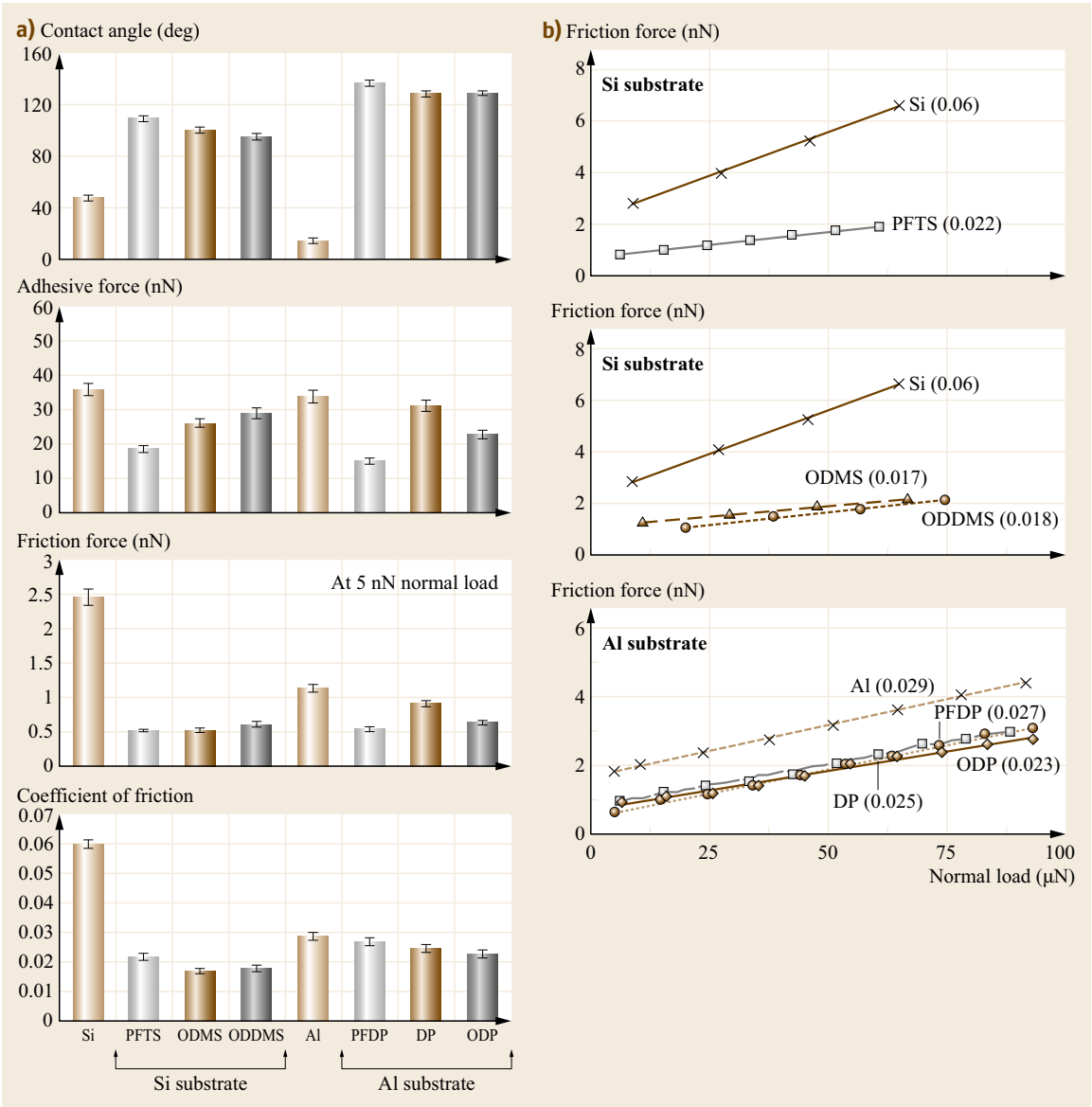


Fig. 34.20 (a) The static contact angle, adhesive force, friction force and coefficient of friction measured using an AFM for various SAMs on Si and Al substrates, and (b) friction force versus normal load plots for various SAMs on Si and Al substrates (after [34.54, 59])

hand, rotate more freely. We have presented earlier a molecular spring or brush model to explain why less-compliant SAMs show larger friction. The SAMs with a higher spring constant or stiffer backbone structure may need more energy to be elastically deformed during sliding; therefore friction is higher for these SAMs. In terms of the chain length effect, it has been reported that the coefficient of friction for SAM surfaces decreases with the carbon backbone chain length (n) up to 12 carbon atoms ($n \approx 12$) [34.85]. This chain length effect for the coefficient of friction was not obvious in these data.

Effect of Relative Humidity, Temperature, and Sliding Velocity on AFM Adhesion and Friction

The effect of relative humidity for various SAMs on adhesion and friction was studied. Adhesive force, friction force at 5 nN of normal load, coefficient of friction, and microwear data are presented in Fig. 34.22 [34.54,

Table 34.13 A summary of RMS roughness, contact angle and film thickness of various SAMs

SAM/substrate	Acronym	RMS roughness (nm)	Contact angle (deg)	Film thickness (nm)
Silicon(111)[a]	Si	0.07	48	–
Perfluorodecyltricholorosilane/Si[a]	PFTS/Si	0.09	112	≈ 1.8
n-Octyldimethyl(dimethylamino)silane/Si[a]	ODMS/Si	0.10	99	≈ 1.9
n-Octadecyldimethyl(dimethylamino)silane/Si[b]	ODDMS/Si	0.10	92	≈ 2.1
Aluminum[b]	Al	32	< 15	–
Perfluorodecylphosphonate	PFDP	34	137	≈ 1.9
Decylphosphonate/Al[b]	DP/Al	31	129	≈ 1.9
Octadecylphosphonate/Al[b]	ODP/Al	36	130	≈ 2.1

[a] *Kasai* et al. [34.61]
[b] *Bhushan* et al. [34.54]

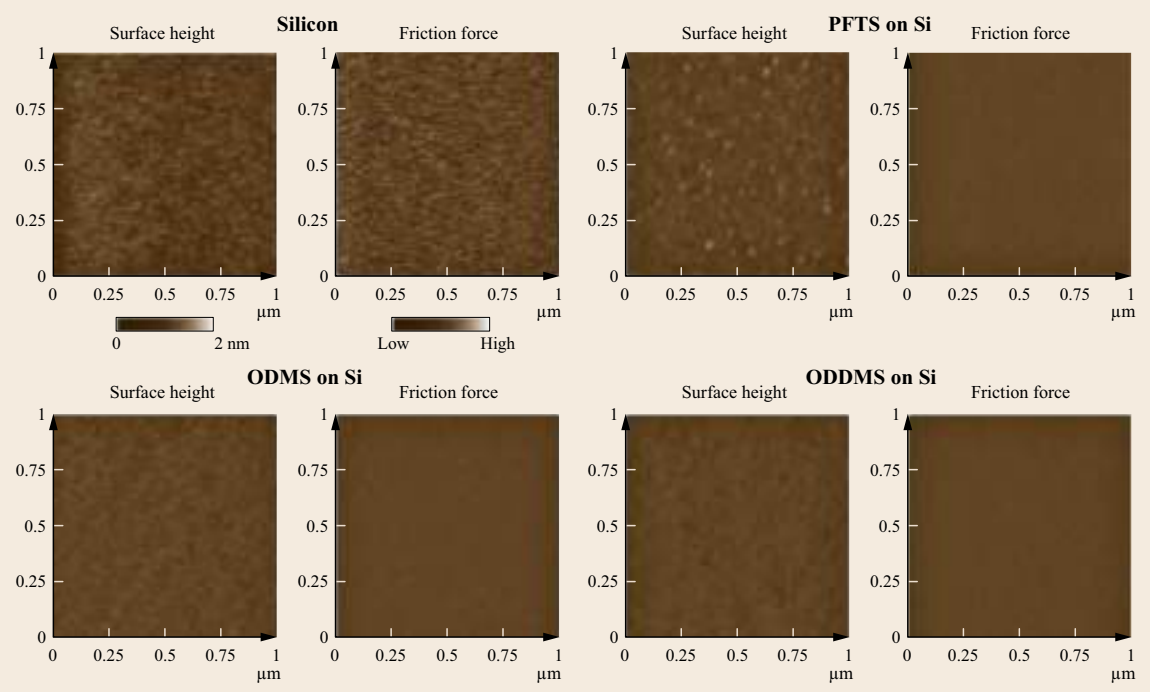


Fig. 34.21 Surface height and friction force maps for Si, PFTS/Si, ODMS/Si, and ODDMS/Si (after [34.59])

61]. The result of adhesive force for silicon showed an increase with relative humidity; Fig. 34.22a. This is expected since the surface of silicon is hydrophilic, as shown in Fig. 34.20a. More condensation of water at the tip–sample interface at higher humidity increases the adhesive force due to the capillary effect. On the other hand, the adhesive force for the SAMs showed a very weak dependency on the change in humidity. This occurs since the surface of the SAMs is hydrophobic. The adhesive force of ODMS/Si and ODDMS/Si showed a slight increase from 75 to 90% RH. Such an increase was absent for PFTS/Si, possibly because of the hydrophobicity of PFTS/Si. The Al substrate is hydrophilic and hence shows relative humidity dependence, Fig. 34.22b. The PFDP, DP, and ODP SAMs deposited on Al substrates showed almost no change in adhesive force with humidity. The highly hydrophobic nature of these monolayers means that the contribution of water menisci to the overall adhesive force is negligible at all humidities.

The friction force of silicon showed an increase with relative humidity up to about 75% RH and a slight decrease beyond this point; Fig. 34.22a. The initial increase can result from the increase in adhesive force. The decrease in friction force at higher humidity could be attributed to the lubricating effect of the water layer.

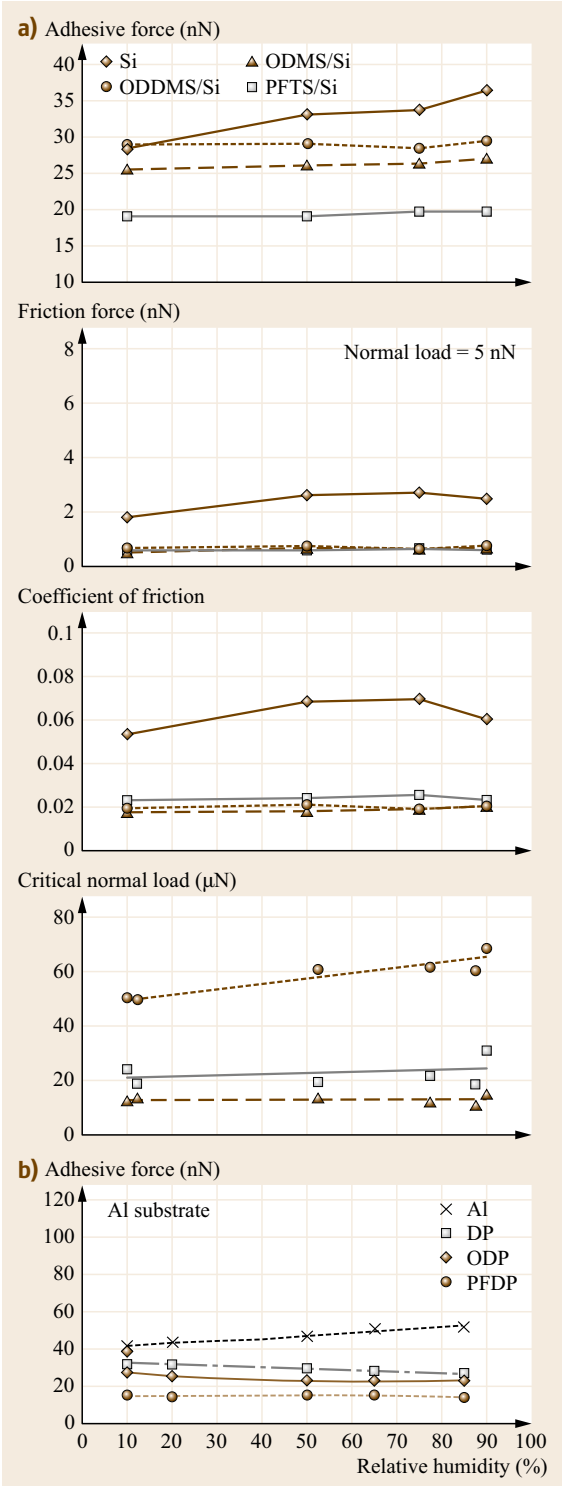


Fig. 34.22a,b Relative humidity effect on **(a)** adhesive force, friction force, coefficient of friction and microwear for various SAMs on Si substrates (after [34.61]), and **(b)** adhesive force for various SAMs on Al substrates (after [34.54]) ◀

This effect is more pronounced in the coefficient of friction. Since the adhesive force increased and coefficient of friction decreased in this range, those effects cancel each other out and the resulting friction force showed only slight changes. On the other hand, the friction force and coefficient of friction of SAMs showed very small changes with relative humidity, like those found for adhesive force. This suggests that the adsorbed water layer on the surface maintained a similar thickness throughout the relative humidity range tested. The differences among the SAM types were small within the measurement error, however a closer look at the coefficient of friction for ODMS/Si showed a slight increase from 75 to 90% RH as compared to PFTS/Si, possibly due to the same reason as for the adhesive force increment. The inherent hydrophobicity of SAMs means that they did not show much relative humidity dependence.

Figure 34.23a shows the effect of temperature on adhesive force, friction force at 5 nN of normal load, and coefficient of friction for various SAMs on Si substrate [34.61]. Figure 34.23b shows the effect of temperature on the adhesive force for SAMs on Al [34.54]. The adhesive force for silicon showed an increase with the temperature, from room temperature (RT) to about 55 °C, followed by a decrease from 55 to 75 °C, and eventually leveled off from 75 to 110 °C. The adhesive force increased for Al substrates up to 50 °C and then decreased to a stable value for higher temperatures. The initial increase of adhesive force for Si and Al substrates at lower temperatures is not well understood. The observed decrease could be attributed to the desorption of water molecules on the surface. After almost full depletion of the water layer, the adhesive force remains constant. The SAMs with hydrocarbon backbones on both Si and Al substrates showed similar behavior as those of the Si and Al substrates, but the initial increase in the adhesive force with temperature was smaller. The SAMs with fluorocarbon backbone chains showed almost no temperature dependence. For the SAMs with hydrocarbon backbone chains, the initial increase in adhesive force is believed to be caused by the melting of the SAM film. The melting point for a linear carbon chain molecule such as $CH_3(CH_2)_{14}CH_2OH$ is 50 °C [34.109]. With an increase in temperature, the SAM film softens, thereby increasing the real area of contact and consequently the adhesive force. Once the temperature is higher than the melting point, the lubrication regime is changed from boundary lubrication in a solid SAM to liquid lubrication in the melted SAM [34.81].

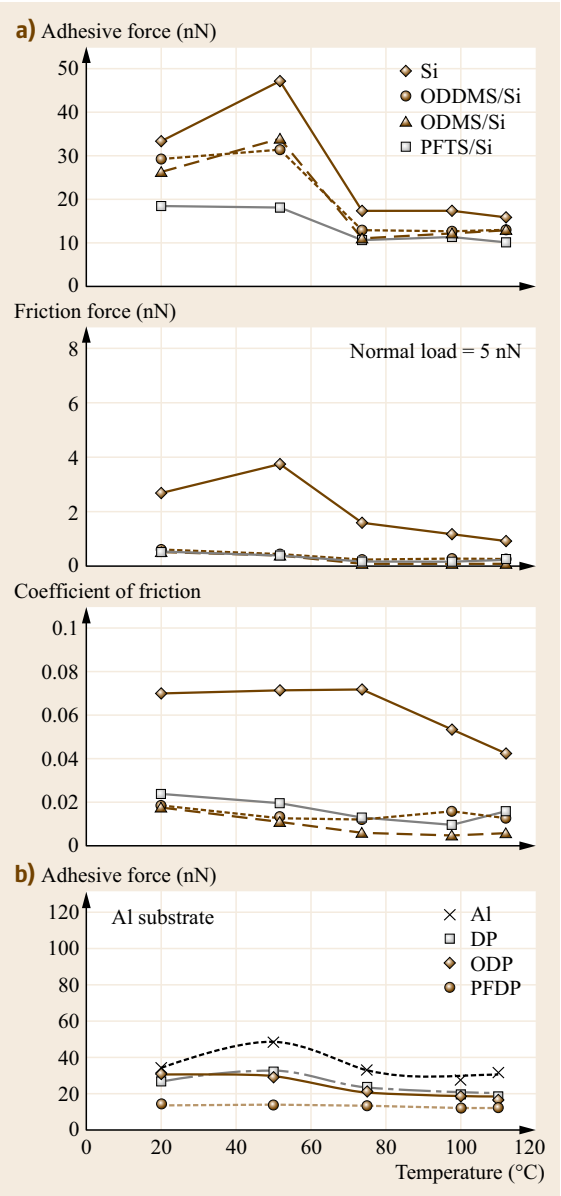


Fig. 34.23a,b Temperature effect **(a)** on adhesive force, friction force, and coefficient of friction for various SAMs on Si substrates (after [34.61]), and **(b)** on adhesive force for various SAMs on Al substrates (after [34.54])

The friction force for silicon showed an increase with temperature followed by a steady decrease. The friction force is highly affected by the change in adhesion. The decrease in friction can result from the depletion of the water layer. The coefficient of friction for silicon remained constant followed by a decrease starting at about 80 °C. For SAMs, the coefficient of friction exhibited a monotonic decrease with temperature. The decrease in friction and coefficient of friction for SAMs possibly results from the decrease in stiffness. As introduced before, the spring model suggests a smaller friction for more compliant SAMs [34.81]. The difference among the SAM types was not significant. PFTS could maintain its stiffness more than ODMS and ODDMS when temperature was increased [34.110], however it was not pronounced in the results.

Figure 34.24a shows the effect of sliding velocity on adhesive force, friction force and coefficient of friction for various SAMs on a Si substrate [34.61]. The adhesive force for silicon remained rather constant at a lower sliding velocity, and then increased rapidly. A similar trend was found for the SAMs. The increase in adhesive force for silicon is believed to be the result of a tribochemical reaction at the tip–sample interface [34.29] and increase of contact area by mechanical plowing. For the SAMs, the higher adhesive force at the higher velocity can result from viscous drag of the SAM molecules [34.111]. SAMs can be detached from the surface and attached to an AFM tip. In addition, another reason may be an increase of contact area, which may be caused by more penetration of the AFM tip into the SAMs. The rate of increase was larger for ODMS than PFTS, presumably because of the higher stiffness and more densely packed structure of PFTS.

Coefficient of friction showed an increase with sliding velocity and reached a plateau for Si, ODMS/Si, and ODDMS/Si. As the sliding speed is increased, the reorientation of the SAMs needs additional work, which might lead to increases in friction. For PFTS, coefficient of friction decreased at larger sliding velocity, forming a peak. The peak structure may result from the viscoelastic property of SAMs [34.112].

Figure 34.24b shows the effect of sliding velocity on the friction force for various SAMs on Al substrates [34.54]. Friction force increases slowly at lower sliding velocities for the bare Al substrate and SAMs, followed by a rapid increase at higher sliding velocities, except for PFDP. The increase in friction force at high velocities ($> 1\,\mathrm{mm/s}$) is the result of asperity impacts and a corresponding high frictional energy dissipation at the sliding interface for Al [34.113]. *Tambe* and *Bhushan* [34.111] extended the *molecular spring* model presented by *Bhushan* and *Liu* [34.79] to explain this velocity-dependent increase in friction force for compliant SAM molecules. Based on this model for the DP and ODP SAMs, the increase in friction force is believed to result from the reorientation of the SAM molecules under the tip load and during tip motion. The reorientation of the SAMs can act as an additional hindrance to tip motion when the AFM tip reverses during scanning and thus results in higher friction. The

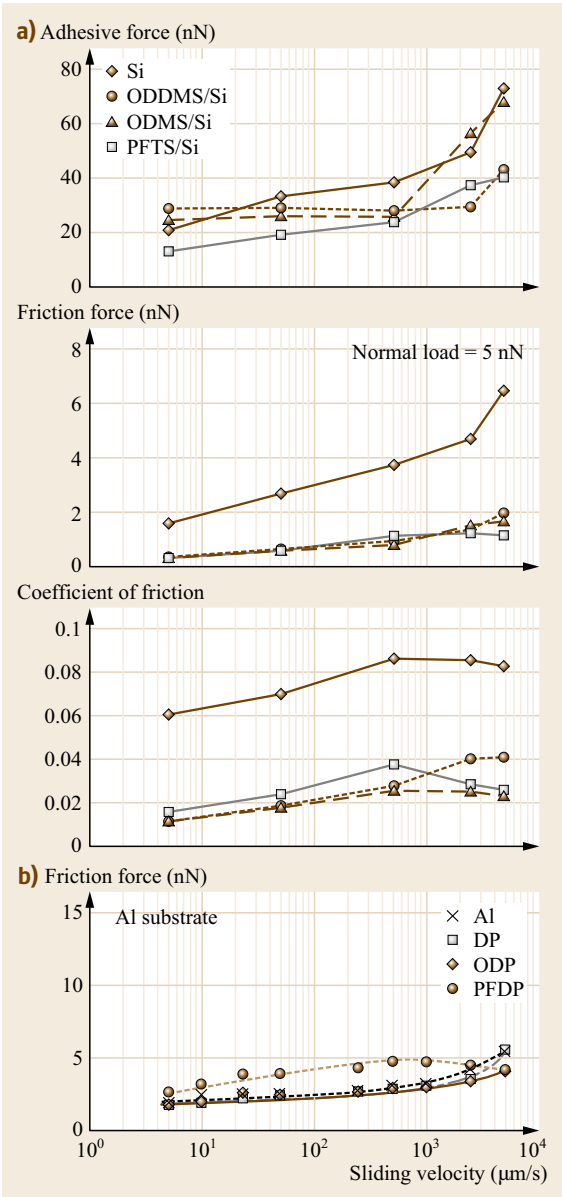


Fig. 34.24a,b Sliding velocity effect **(a)** on adhesive force, friction force, and coefficient of friction for various SAMs on Si substrates (after [34.61]), and **(b)** on friction force for various SAMs on Al substrates (after [34.54])

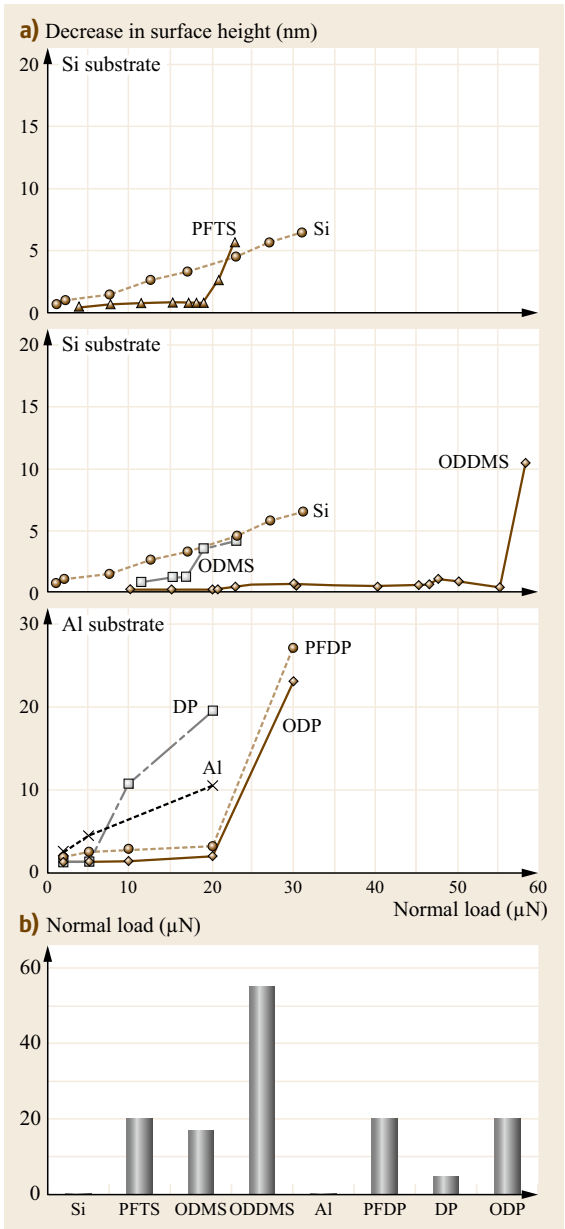


Fig. 34.25 **(a)** Decrease in surface height as a function of normal load after one scan cycle for various SAMs on Si and Al substrates, and **(b)** comparison of critical loads for failure during wear tests for various SAMs (after [34.54, 61])

molecules can consequently get entangled and/or get detached from the substrate and attach to the AFM tip.

AFM Wear Measurements

Figure 34.25a shows the relationship between the decrease in surface height as a function of the normal load during wear tests [34.54, 61]. As shown in the figure, the SAMs exhibit a critical normal load, beyond which the surface height drastically decreases. Figure 34.25a also shows the wear behavior of the Al and Si substrates. Unlike the SAMs, the substrates show a monotonic decrease in surface height with the increasing normal load with wear initiating from the

very beginning, i. e., even for low normal loads. Si (Young's modulus of elasticity, $E = 130$ GPa [34.114, 115], hardness, $H = 11$ GPa [34.27]) is relatively hard in comparison to Al ($E = 77$ GPa, $H = 0.41$ GPa), and hence the decrease in surface height for Al is much larger than that for Si for similar normal loads.

The critical loads corresponding to the sudden failure of SAMs are shown in Fig. 34.25b. Amongst all the SAMs, ODDMS shows the best performance in the wear tests, and this is believed to be because of the longer chain length effect. Fluorinated SAMs – PFTS and PFDP – show a higher critical load as compared to ODMS and DP with similar chain lengths. ODP shows a higher critical load as compared to DP because of its longer chain length. The mechanism of failure of compliant SAMs during wear tests has been presented earlier in Fig. 34.18. It is believed that the SAMs fail mostly due to shearing of the molecule at the head group, that is, by means of shearing of the molecules off the substrate. Table 34.14 gives the bond strengths for various intermolecular bonds. The weakest bonds are at the interface, and hence failure is expected to be initiated at the interface.

To study the effect of relative humidity, wear tests were performed at various humidities. The bottom of Fig. 34.22a shows critical normal load as a function of relative humidity. The critical normal load showed weak dependency on relative humidity for ODMS/Si and PFTS/Si and was larger for ODMS/Si than PFTS/Si throughout the humidity range. This suggests that water molecules could penetrate into the ODDMS, which might work as lubricants [34.81, 118]. This effect was absent for PFTS/Si and ODMS/Si.

34.4.4 Chemical Degradation and Environmental Studies

Chemical degradation and environmental studies were carried out for HDT/Au, PFTS/Si, ODMS/Si, and ODDMS/Si films.

Chemical Degradation Studies

The coefficient of friction and detected gaseous products for HDT/Au are shown in Fig. 34.26a [34.63]. A normal pressure of 50 kPa was applied to HDT films. The coefficient of friction increased after a sliding distance of about 10 m. During sliding, $(CH_2)_{15}S$, C_2H_3, CH_3, CH_2, and H_2 were detected by mass spectrometer. The partial pressure of HS fragments is of interest as it corresponds to the interface bonds, and it is reported here. The increase of partial pressure for $(CH_2)_{15}S$ was much more than that of other species, due to the breaking of the S-Au bond. Partial pressures of C_2H_3, CH_3, CH_2, and H_2 were also found to increase during the sliding. There was no noticeable change in the partial pressure of HS.

HDT film is deposited on an Au(111) layer. The bond strength of S–Au is 184 kJ/mol (Table 34.14), which is lower than that of C–C bonds (425 kJ/mol), C−H bonds (422 kJ/mol), and C–S bonds (286 kJ/mol) in the alkyl chain. Since the S–Au bond is the weakest bond in the alkanethiol chain, the whole chain should be sheared away from the substrate. Because the upper

Table 34.14 Typical bond strengths[a] in SAMs

SAMs	Bond	HDT (kJ/mol)	BPT (kJ/mol)	Bond	PFTS (kJ/mol)	ODMS or ODDMS (kJ/mol)	PFDP (kJ/mol)	DP and ODP (kJ/mol)
Interfacial bonds	S–Au	184[b]	184[b]	Si–O	242[c]	242[c]		
	S–C	286[a]			800[d]	800[d]		
	C_6H_5−S		362[a]	Si–C	414[a]	414[a]		
				Al–O			511[a]	511[a]
				P–C			513[a]	513[a]
				P–O			599[a]	599[a]
	C–C			C–C				
Bonds in backbone	CH_2−CH_2	326[e]		CH_2−CH_2	326[e]	326[e]	326[e]	326[e]
	CH_3−CH_2	≈ 305[a]		CF_2−CF_2	≈ 326[f]		≈ 326[f]	
	C_6H_5		strong	CF_2−CH_2	≈ 326[f]		≈ 326[f]	
				CF_3−CF_2	≈ 326[f]		≈ 326[f]	
				CH_3−CH_2		≈ 305[a]		≈ 305[a]

[a] *Lide* [34.109]
[b] Chemical adsorption bond from *Lio* et al. [34.89]
[c] Chemical adsorption bond from *Hoshino* [34.116]
[d] In diatomic molecules
[e] *Cottrell* [34.117]
[f] Due to the C–C bond, should be close to that of CH_2−CH_2

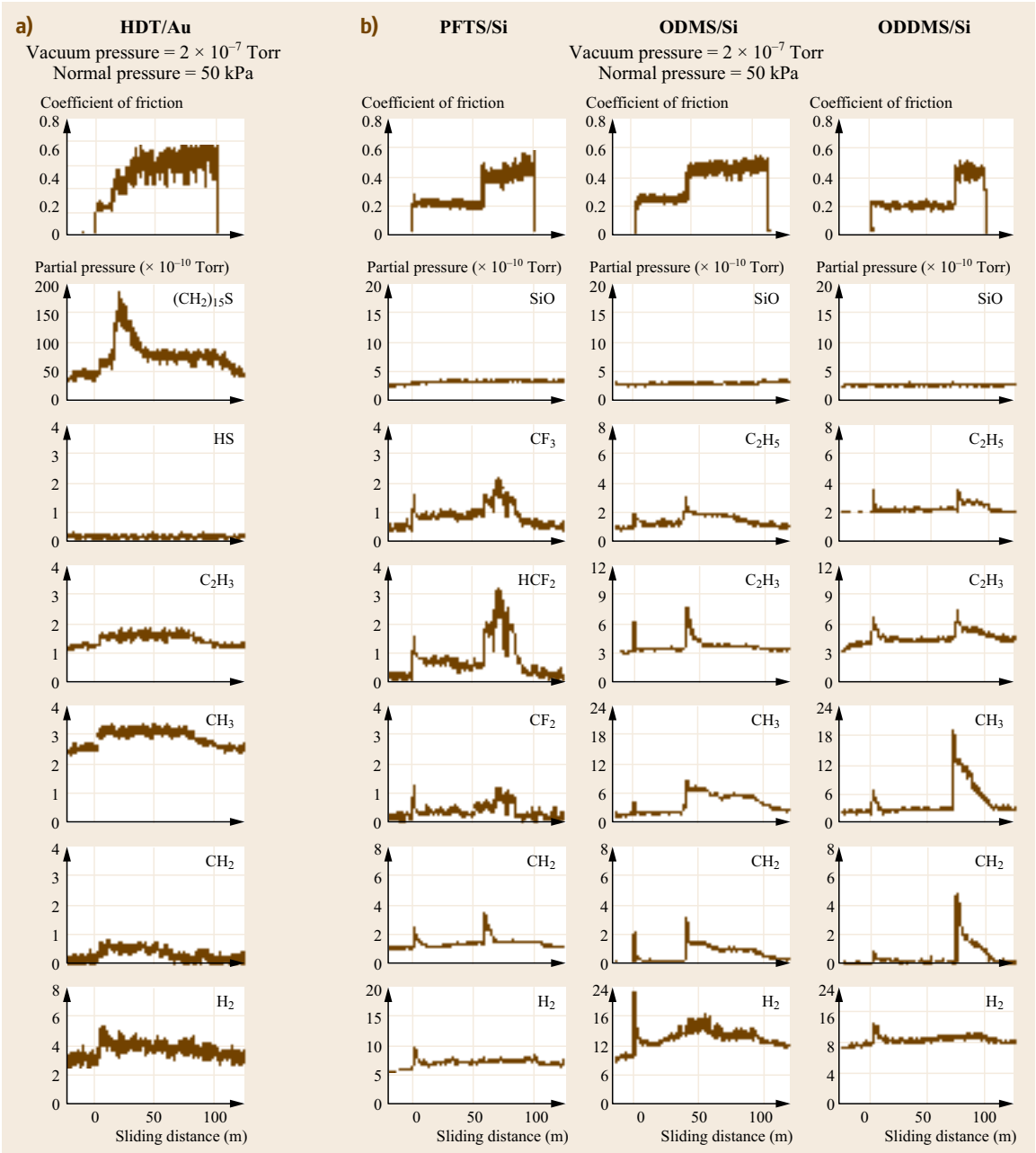


Fig. 34.26a,b Coefficients of friction and mass spectral data on **(a)** HDT/Au (1.9 nm), and **(b)** PFTS/Si (1.8 nm), ODMS/Si (≈ 1.9 nm), and ODDMS/Si (≈ 2.1 nm) in high vacuum (after [34.63])

atomic mass unit (amu) limit of the mass spectrometer used is 250, we monitored $(CH_2)_{15}S$ (amu = 242), which is the chain with CH_3 sheared. The rate of generation of $(CH_2)_{15}S$ is much larger than that of other species. This suggests that the mechanical shear of the whole alkanethiol chain is the dominant factor causing the failure of the HDT film. The cleavage of the S–Au bonds has been reported in the literature. Based on the bond strength as well as the above studies, mechanical shearing of the C–C bonds and C–H bonds does not likely happen during sliding. The reaction induced by low-energy electrons, generated by triboelectrical emission during the sliding, could be responsible for the degradation of the alkanethiol chain. Thermal desorp-

tion of HDT from Au is another possible degradation mechanism of HDT.

The coefficient of friction and generated gaseous products for PFTS/Si, ODMS/Si, and ODDMS/Si are shown in Fig. 34.26b [34.63]. The coefficients of friction for PFTS/Si, ODMS/Si, and ODDMS/Si increase sharply after a certain sliding distance, which indicates the degradation of the film. At the same time, gaseous products of CF_3, HCF_2, CF_2, CH_2, and H_2 were detected for PFTS/Si, and C_2H_5, C_2H_3, CH_3, CH_2, and H_2 were detected for ODMS/Si, and ODDMS/Si.

PFTS/Si showed lower friction than ODMS/Si in the tests. ODDMS/Si showed lower friction than both PFTS/Si and ODMS/Si. This is because of the chain length effect. It has been reported that the coefficient of friction for SAM surfaces decreases with the carbon backbone chain length (n) when the carbon atoms are less than 12. For chains with more than 12 carbons, the change of carbon atoms will not influence the coefficient of friction to a noticeable extent.

PFTS/Si showed a higher durability than ODMS/Si. For the case of a perfluorinated carbon backbone, it is harder to rotate the backbone structure (due to the different sizes of F atoms versus H atoms), which implies that this structure is more rigid than a hydrocarbon backbone [34.108]. *Chambers* [34.119] reported that the C–C bond strength increases when hydrogen is replaced with fluorine. This suggests that the rigid perfluorinated carbon backbone may be responsible for the increased durability. The length of the alkyl chain also influences the desorption energy of alkanes. Based on studies of the adsorption of alkanes on Cu(100), Au(111), Pt(110), and others, the physisorption energy increases with the alkyl chain length [34.120–122]. Therefore, ODDMS is more durable than ODMS.

During sliding on PFTS films, gaseous products of CF_3, HCF_2, CF_2, CH_2, and H_2 were detected. From the structure of perfluoroalkylsilane, the only source of H on the molecular chain that would cause a partial pressure increase of H_2 is from the $(CH_2)_2$, which is located at the bottom of the chain. Since the partial pressure of H_2 increases immediately after sliding and remains high until the end of sliding, it is probably generated by the low-energy electrons coming from a triboelectrical emission. The partial pressure of CH_2 had a sharp peak at the beginning of sliding and at the moment when friction changed. Meanwhile, the partial pressures of CH_3, HSF_2, and CF_2 increased significantly when the friction increased. For ODMS and ODDMS, C_2H_5, C_2H_3, CH_3, CH_2, and H_2 were detected during sliding. The partial pressure of the carbon-related products increases considerably when the friction is increased. The SiO, which is related to the interface bonds, has no noticeable change during sliding.

Perfluoroalkylsilanes and alkylsilanes are attached to the naturally oxidized silicon by Si–O bonds. The Si–O bond strength varies over a large range (Table 34.14), depending on the forming condition. In the alkylsilane chain, C–Si bond strength (414 kJ/mol) is slightly lower than C–C bond strength. Based on Table 34.14, interfacial bonds (Si–O) are weaker than the C–C bonds in the backbone. It is believed that the cleavage of films occurs at the interface. We have previously reported evidence of the cleavage of the interfacial bonds using an AFM. To explain the hydrogen, C_1 and C_2 hydrocarbon (in the tests for PFTS/Si, ODMS/Si, and ODDMS/Si) or fluorocarbon (in the tests for PFTS/Si) products, *Kluth* et al. [34.123] suggested that the alkylsilane (perfluoroalkylsilane as well) chains break and create radicals. The radical could remain on the surface and decompose to generate a shorter radical and an alkene. The radical could repeatedly decompose to ever-shorter radicals and alkenes as long as it remains on the surface.

A summary of the coefficients of friction and durability of all films in a vacuum is presented in Fig. 34.27 [34.63].

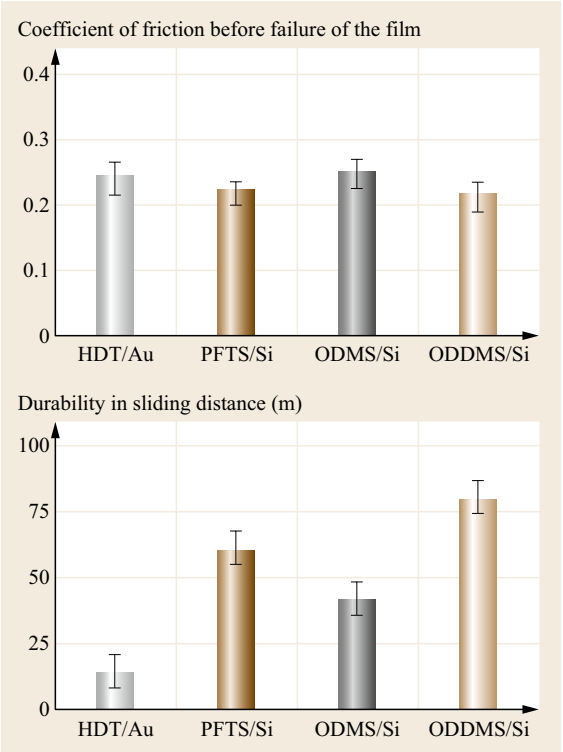


Fig. 34.27 Coefficient of friction and durability comparison of HDT/Au, PFTS/Si, ODMS/Si, and ODDMS/Si in high vacuum. Error bars represent $\pm 3\sigma$ based on five measurements (after [34.63])

Environmental Studies

To study the effect of environment, the friction tests on HDT/Au, PFTS/Si, ODMS/Si, and ODDMS/Si were conducted in high vacuum, argon, dry air (less than 2% RH), air with 30% RH, and air with 70% RH; see Fig. 34.28 [34.63]. By comparing the coefficient of friction in different environments, it was found that the friction in argon is the lowest for the SAMs tested. In high vacuum, the intimate contact leads to high friction. In dry air, the friction is higher than in the argon. This shows that oxygen has an apparent effect on the performance of SAMs. *Kim* et al. [34.124] studied the thermal stability of alkylsiloxane SAMs in air. They found that the alkylsiloxane decomposes at about 200 °C, which is much lower than the decomposition temperature of 470 °C in vacuum reported by *Kluth* et al. [34.123]. This difference could be attributed to the oxygen in air. The water contained in the air is found to have a significant influence on the friction of SAMs. A humidity effects study of alkylsilane on mica substrate by *Tian* et al. [34.118] indicated that the water molecules can

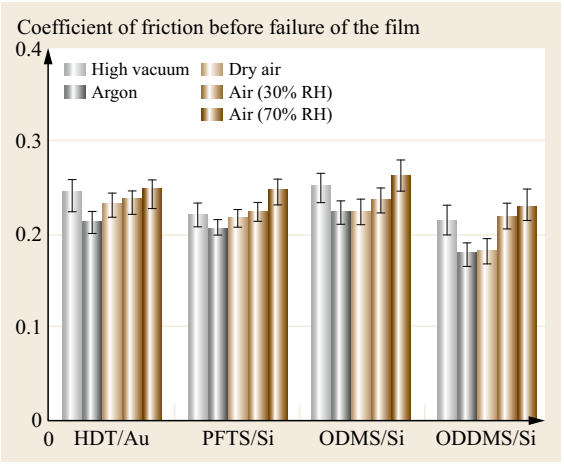


Fig. 34.29 Comparison of coefficient of friction data for HDT/Au, PFTS/Si, ODMS/Si, and ODDMS/Si in high vacuum, argon, and air with different humidity levels. Error bars represent $\pm 3\sigma$ based on five measurements (after [34.63])

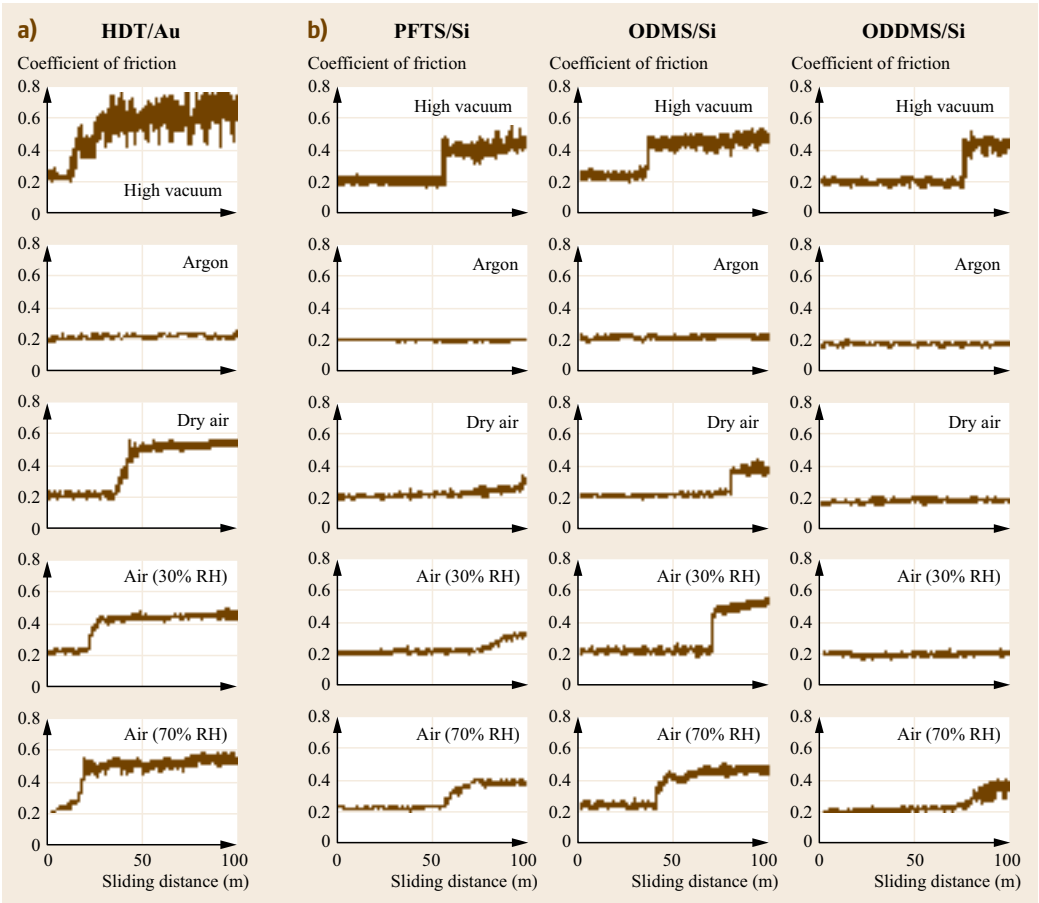


Fig. 34.28 Coefficient of friction data in high vacuum, argon, and air with different humidity levels of **(a)** HDT/Au (1.9 nm) and **(b)** PFTS/Si (1.8 nm), ODMS/Si ($\approx$ 1.9 nm), and ODDMS/Si ($\approx$ 2.1 nm) (after [34.63])

penetrate the alkylsilane film, which alters their molecular chain ordering and can also detach the alkylsilane molecules from the substrate.

A summary of the coefficients of friction before the failure of the lubricant films in various environments is presented in Fig. 34.29 [34.63]. The data in Fig. 34.29 are average values based on five measurements. To summarize the highlights, friction of tested lubricant films is high in high vacuum because of the intimate contact between the lubricants and the counterpart surface. Friction of tested lubricant films is lower in argon than in dry air. Friction of SAMs is significantly influenced by water molecules.

34.5 Conclusion

The adhesive force can be significant in an interface with ultrasmooth surfaces and it can be on the same order as the external load if the latter is small, such as in micro- and nanodevices. High adhesion also leads to stiction, friction, and wear issues in sliding surfaces. In various applications, surfaces need to be protected from exposure to the operating environment, and hydrophobic films are of interest. Films are also needed to provide low adhesion, stiction, friction, and wear. These films should be molecularly thick, well organized, chemically bonded to the substrate, and insensitive to environment. Ordered molecular assemblies with high hydrophobicity can be engineered using chemical grafting of various polymer molecules with suitable functional head groups and nonpolar surface terminal groups.

The contact angle, adhesion, friction, and wear properties of SAMs having alkyl, biphenyl, and perfluoroalkyl spacer chains with different surface terminal groups ($-CH_3$ and $-CF_3$) and head groups ($-SH$, $-Si-O-$, $-OH$, and $P-O-$) were presented in this chapter. It was found that the adhesive force varies linearly with the work of adhesion value of SAMs, which indicates that capillary condensation of water plays an important role to the adhesion of SAMs on the nanoscale at ambient conditions. SAMs with high-compliance long carbon spacer chains exhibited the lowest adhesive force and friction force. The friction data were explained using a molecular spring model, in which the local stiffness and intermolecular force govern frictional performance. The results of the stiffness and friction characterization of the micropatterned sample with different structures supports this model. Perfluoroalkylsilane and perfluoroalkylphosphonate SAMs exhibited lower surface energy, higher contact angle, and lower adhesive force as compared to that of alkylsilane and alkylphosphonate SAMs, respectively. The substrate had little effect. The coefficients of friction of various SAMs were comparable.

The influence of relative humidity on adhesion and friction of SAMs was dominated by the thickness of the adsorbed water layer. At higher humidity, water increases friction through increased adhesion by the meniscus effect in the contact zone. With an increase in temperature, in the case of Si(111), the desorption of the adsorbed water layer and reduction of the surface tension of water reduced the adhesive force and friction force. A decrease in adhesion and friction with temperature was found for all films. An increase in adhesive force and friction with sliding velocity was found for all films.

PFTS/Si showed a better wear resistance than ODMS/Si. ODDMS/Si showed a better wear resistance than ODMS/Si due to the chain length effect. Wear behavior of the SAMs was mostly determined by the molecule-substrate bond strengths. Similar trends were observed for films on Al substrates.

The nanotribological characterization studies of SAMs deposited on Si and Al substrates are summarized in Table 34.15 [34.53]. SAMs deposited on the Si and Al substrates showed low friction and low adhesion, both of which are desirable for MEMS/NEMS applications.

Table 34.15 Summary of nanotribological characterization studies for SAMs on Si and Al substrates

	SAMs property	Friction force	Adhesive force	Wear
Substrate	Hard	High	Low	Low
	Soft	Low	Low	High
Chemical structure	Linear chain molecule	High	Low	High
	Ring molecule	High	High	Low
Backbone	Fluorocarbon backbone	Low	Low	Low
	Hydrocarbon backbone	Low	High	High
Chain length	Long backbone chain		High	High
	Short backbone chain		Low	Low

Based on studies in high vacuum (2×10^{-7} Torr), the coefficients of friction of the SAMs are in the following order (from low to high): ODDMS/Si, PFTS/Si, HDT/Au, ODMS/Si. HDT on Au showed lower durability than perfluoroalkylsilane and alkylsilane on Si because of the weak interfacial bond. PFTS/Si had better durability than ODMS/Si. This indicates that fluorination of alkylsilane can improve durability. ODDMS/Si was more durable than ODMS/Si and PFTS/Si because of the chain length effect. The friction of SAMs in high vacuum was higher than in argon because of intimate contact. Based on studies in argon and air with various relative humidities, oxygen can increase friction and decrease durability of SAMs. The water molecule can detach SAM molecules from the substrate and result in high friction and low durability.

In summary, based on the contact angle, adhesion, friction, and wear measurements of SAM films by AFM, these films exhibit attractive hydrophobic and tribological properties. Fluorinated SAMs appear to exhibit superior performance. SAM films should find many applications including in micro- and nanodevices requiring good nanotribological properties and surface protection.

References

34.1 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374**, 607–616 (1995)

34.2 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)

34.3 B. Bhushan (Ed.): *Tribology Issues and Opportunities in MEMS* (Kluwer Academic, Dordrecht 1998)

34.4 B. Bhushan (Ed.): *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

34.5 B. Bhushan: Nanotribology and nanomechanics of MEMS/NEMS and BioMEMS/BioNEMS materials and devices, Microelectron. Eng. **84**, 387–412 (2007)

34.6 B. Bhushan: Nanotribology and nanomechanics in nano/biotechnology, Phil. Trans. R. Soc. A **366**, 1499–1537 (2008)

34.7 K.F. Man, B.H. Stark, R. Ramesham: *A Resource Handbook for MEMS Reliability* (JPL, Jet Propulsion Laboratory, California Institute of Technology, Pasadena 1998)

34.8 K.F. Man: MEMS reliability for space applications by elimination of potential failure modes through testing and analysis, http://www-rel.jpl.nasa.gov/0rg/5053/atpo/products/Prod-map.html (2002)

34.9 D.M. Tanner, N.F. Smith, L.W. Irwin, W.P. Eaton, K.S. Helgesen, J.J. Clement, W.M. Miller, J.A. Walraven, K.A. Peterson, P. Tangyunyong, M.T. Dugger, S.L. Miller: *MEMS reliability: Infrastructure, rest structure, experiments, and failure modes, SAND2000-0091* (Sandia National Laboratories, Albuquerque 2000)

34.10 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Switzerland 2016)

34.11 A.W. Adamson: *Physical Chemistry of Surfaces*, 5th edn. (Wiley, New York 1990)

34.12 J.N. Israelachvili: *Intermolecular and Surface Forces*, 2nd edn. (Academic, London 1992)

34.13 M.E. Schrader, G.I. Loeb (Eds.): *Modern Approaches to Wettability* (Plenum, New York 1992)

34.14 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)

34.15 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

34.16 S. Cai, B. Bhushan: Meniscus and viscous forces during separation of hydrophilic and hydrophobic surfaces with liquid mediated contacts, Mater. Sci. Eng. R **61**, 78–106 (2008)

34.17 B. Bhushan: Contact mechanics of rough surfaces in tribology: Multiple asperity contact, Tribol. Lett. **4**, 1–35 (1998)

34.18 B. Bhushan (Ed.): *Modern Tribology Handbook, Vol. 1 – Principles of Tribology* (CRC, Boca Raton 2001)

34.19 B. Bhushan (Ed.): *Modern Tribology Handbook, Vol. 2 – Materials, Coatings, and Industrial Applications* (CRC, Boca Raton 2001)

34.20 B. Bhushan: Adhesion and stiction: Mechanisms, measurement techniques, and methods for reduction, J. Vac. Sci. Technol. B **21**, 2262–2296 (2003)

34.21 B. Bhushan, Z. Zhao: Macro- and microscale tribological studies of molecularly-thick boundary layers of perfluoropolyether lubricants for magnetic thin-film rigid disks, J. Inf. Storage Proc. Syst. **1**, 1–21 (1999)

34.22 B. Bhushan, W. Peng: Contact mechanics of multilayered rough surfaces, Appl. Mech. Rev. **55**, 435–480 (2002)

34.23 B. Bhushan, Y.C. Jung: Wetting, adhesion, and friction of superhydrophobic and hydrophilic leaves and fabricated micro/nanopatterned surfaces, J. Phys. Condens. Matter **20**(22), 225010 (2008)

34.24 M. Nosonovsky, B. Bhushan: Roughness-induced superhydrophobicity: A way to design non-adhesive surfaces, J. Phys. Condens. Matter **20**, 225009 (2008)

34.25 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids, Part I* (Clarendon, Oxford 1950)

34.26 W.A. Zisman: Friction, durability and wettability properties of monomolecular films on solids. In: *Friction and Wear*, ed. by R. Davies (Elsevier, Am-

sterdam 1959) pp. 110–148
34.27 B. Bhushan: Chemical, mechanical and tribological characterization of ultra-thin and hard amorphous carbon coatings as thin as 3.5 nm: Recent developments, Diam. Relat. Mat. **8**, 1985–2015 (1999)
34.28 V.N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14**, 2378–2391 (1996)
34.29 H. Liu, B. Bhushan: Nanotribological characterization of molecularly-thick lubricant films for applications to MEMS/NEMS by AFM, Ultramicroscopy **97**, 321–340 (2003)
34.30 A. Ulman: *An Introduction to Ultrathin Organic Films: From Langmuir-Blodgett to Self-Assembly* (Academic, San Diego 1991)
34.31 A. Ulman: Formation and structure of self-assembled monolayers, Chem. Rev. **96**, 1533–1554 (1996)
34.32 H. Hansma, F. Motamedi, P. Smith, P. Hansma, J.C. Wittman: Molecular resolution of thin, highly oriented poly (tetrafluoroethylene) films with the atomic force microscope, Polymer **33**, 647–649 (1992)
34.33 L. Scandella, A. Schumacher, N. Kruse, R. Prins, E. Meyer, R. Luethi, L. Howald, M. Scherge, J.A. Schaefer: Surface modification and mechanical properties of bulk silicon. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 529–537
34.34 A. Erdemir, C. Donnet: Tribology of diamond, diamond-like carbon, and related films. In: *Materials, Coatings, and Industrial Applications*, Modern Tribology Handbook, Vol. 2, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 871–908
34.35 V.F. Dorfman: Diamond-like nanocomposites (DLN), Thin Solid Films **212**, 267–273 (1992)
34.36 M. Grischke, K. Bewilogua, K. Trojan, H. Dimigan: Application-oriented modification of deposition process for diamond-like-carbon based coatings, Surf. Coat. Technol. **74/75**, 739–745 (1995)
34.37 R.S. Butter, D.R. Waterman, A.H. Lettington, R.T. Ramos, E.J. Fordham: Production and wetting properties of fluorinated diamond-like carbon coatings, Thin Solid Films **311**, 107–113 (1997)
34.38 M. Grischke, A. Hieke, F. Morgenweck, H. Dimigan: Variation of the wettability of DLC coatings by network modification using silicon and oxygen, Diam. Rel. Mater. **7**, 454–458 (1998)
34.39 C. Donnet, J. Fontaine, A. Grill, V. Patel, C. Jahnes, M. Belin: Wear-resistant fluorinated diamondlike carbon films, Surf. Coat. Technol. **94/95**, 531–536 (1997)
34.40 D.J. Kester, C.L. Brodbeck, I.L. Singer, A. Kyriakopoulos: Sliding wear behavior of diamond-like nanocomposite coatings, Surf. Coat. Technol. **113**, 268–273 (1999)
34.41 H. Liu, B. Bhushan: Adhesion and friction studies of microelectromechanical systems/nanoelectromechanical systems materials using a novel microtriboapparatus, J. Vac. Sci. Technol. A **21**, 1528–1538 (2003)
34.42 B. Bhushan, H. Liu, S.M. Hsu: Adhesion and friction studies of silicon and hydrophobic and low friction films and investigation of scale effects, ASME J. Tribol. **126**, 583–590 (2004)
34.43 J.A. Zasadzinski, R. Viswanathan, L. Madsen, J. Garnaes, D.K. Schwartz: Langmuir-blodgett films, Science **263**, 1726–1733 (1994)
34.44 J. Tian, Y. Xia, G.M. Whitesides: Microcontact printing of SAMs. In: *Thin Films – Self-Assembled Monolayers of Thiols*, Vol. 24, ed. by A. Ulman (Academic, San Diego 1998) pp. 227–254
34.45 Y. Xia, G.M. Whitesides: Soft lithography, Angew. Chem. Int. Ed. **37**, 550–575 (1998)
34.46 A. Kumar, G.M. Whitesides: Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol ink followed by chemical etching, Appl. Phys. Lett. **63**, 2002–2004 (1993)
34.47 S.Y. Chou, P.R. Krauss, P.J. Renstrom: Imprint lithography with 25-nanometer resolution, Science **272**, 85–87 (1996)
34.48 Y. Xia, E. Kim, X.M. Zhao, J.A. Rogers, M. Prentiss, G.M. Whitesides: Complex optical surfaces formed by replica molding against elastomeric masters, Science **273**, 347–349 (1996)
34.49 L.J. Hornbeck: The DMD projection display chip: A MEMS-based technology, MRS Bulletin **26**, 325–328 (2001)
34.50 M.R. Douglass: Lifetime estimates and unique failure mechanisms of the digital micromirror device (DMD). In: *Proc. 36th Annu. Int. Reliab. Phys. Symp., Reno* (1988) pp. 9–16
34.51 H. Liu, B. Bhushan: Nanotribological characterization of digital micromirror devices using an atomic force microscope, Ultramicroscopy **100**, 391–412 (2004)
34.52 H. Liu, B. Bhushan: Investigation of nanotribological and nanomechanical properties of the digital micromirror device by atomic force microscope, J. Vac. Sci. Technol. A **22**, 1388–1396 (2004)
34.53 N.S. Tambe, B. Bhushan: Nanotribological characterization of self assembled monolayers deposited on silicon and aluminum substrates, Nanotechnology **16**, 1549–1558 (2005)
34.54 B. Bhushan, M. Cichomski, E. Hoque, J.A. DeRose, P. Hoffmann, H.J. Mathieu: Nanotribological characterization of perfluoroalkylphosphonate self-assembled monolayers deposited on aluminum-coated silicon substrates, Microsyst. Technol. **12**, 588–596 (2006)
34.55 E. Hoque, J.A. DeRose, P. Hoffmann, H.J. Mathieu, B. Bhushan, M. Cichomski: Phosphonate self-assembled monolayers on aluminum surfaces, J. Chem. Phys. **124**, 174710 (2006)
34.56 E. Hoque, J.A. DeRose, G. Kulik, P. Hoffmann, H.J. Mathieu, B. Bhushan: Alkylphosphonate modified aluminum oxide surfaces, J. Phys. Chem. B **110**, 10855–10861 (2006)
34.57 J.A. DeRose, E. Hoque, B. Bhushan, H.J. Mathieu: Characterization of perfluorodecanote self-assembled monolayers on aluminum and com-

parison of stability with phosphonate and siloxy self-assembled monolayers, Surf. Sci. **602**, 1360–1367 (2008)
34.58 B. Bhushan, A.V. Kulkarni, V.N. Koinkar, M. Boehm, L. Odoni, C. Martelet, M. Belin: Microtribological characterization of self-assembled and langmuir-blodgett monolayers by atomic and friction force microscopy, Langmuir **11**, 3189–3198 (1995)
34.59 B. Bhushan, T. Kasai, G. Kulik, L. Barbieri, P. Hoffmann: AFM study of perfluorosilane and alkylsilane self-assembled monolayers for anti-stiction in MEMS/NEMS, Ultramicroscopy **105**, 176–188 (2005)
34.60 B. Bhushan, D. Hansford, K.K. Lee: Surface modification of silicon surfaces with vapor phase deposited ultrathin fluorosilane films for biomedical devices, J. Vac. Sci. Technol. A **24**, 1197–1202 (2006)
34.61 T. Kasai, B. Bhushan, G. Kulik, L. Barbieri, P. Hoffmann: Nanotribological study of perfluorosilane SAMs for anti-stiction and low wear, J. Vac. Sci. Technol. B **23**, 995–1003 (2005)
34.62 K.K. Lee, B. Bhushan, D. Hansford: Nanotribological characterization of perfluoropolymer thin films for BioMEMS applications, J. Vac. Sci. Technol. A **23**, 804–810 (2005)
34.63 Z. Tao, B. Bhushan: Degradation mechanisms and environmental effects on perfluoropolyether self assembled monolayers and diamondlike carbon films, Langmuir **21**, 2391–2399 (2005)
34.64 Z. Tao, B. Bhushan: Surface modification of AFM silicon probes for adhesion and wear reduction, Tribol. Lett. **21**, 1–16 (2006)
34.65 E. Hoque, J.A. DeRose, P. Hoffmann, B. Bhushan, H.J. Mathieu: Chemical stability of nonwetting, low adhesion self-assembled monolayer films formed by perfluoroalkylsilazation of copper, J. Chem. Phys. **126**, 114706 (2007)
34.66 E. Hoque, J.A. DeRose, B. Bhushan, H.J. Mathieu: Self-assembled monolayers on aluminum and copper oxide surfaces: Surface and characteristics, nanotribological properties, and chemical stability. In: *Applied Scanning Probe Methods–Characterization*, Vol. IX, ed. by B. Bhushan, H. Fuchs, M. Tomitori (Springer, Berlin, Heidelberg 2008) pp. 235–281
34.67 E. Hoque, J.A. DeRose, B. Bhushan, K.W. Hipps: Low adhesion, non-wetting phosphonate self-assembled monolayer films formed on copper oxide surfaces, Ultramicroscopy **109**, 1015–1022 (2009)
34.68 A. Manz, H. Becker (Eds.): *Microsystem Technology in Chemistry and Life Sciences* (Springer, Berlin, Heidelberg 1998)
34.69 J. Cheng, L.J. Krica (Eds.): *Biochip Technology* (Harwood Academic, New York 2001)
34.70 M.J. Heller, A. Guttman (Eds.): *Integrated Microfabricated Biodevices* (Marcel Dekker, New York 2001)
34.71 A. van der Berg (Ed.): *Lab-on-a-Chip: Chemistry in Miniaturized Synthesis and Analysis Systems* (Elsevier, Amsterdam 2003)
34.72 D.R. Tokachichu, B. Bhushan: Bioadhesion of polymers for BioMEMS, IEEE Trans. Nanotech. **5**, 228–231 (2006)
34.73 B. Bhushan, M. Cichomski: Nanotribological characterization of vapor phase deposited fluorosilane self-assembled monolayers deposited on polydimethylsiloxane surfaces for biomedical micro-/nanodevices, J. Vac. Sci. Technol. A **25**, 1285–1293 (2007)
34.74 M. Hein, L.R. Best, S. Pattison, S. Arena: *Introduction to General, Organic, and Biochemistry*, 6th edn. (Brooks/Cole, Pacific Grove 1997)
34.75 J.R. Mohrig, C.N. Hammond, T.C. Morrill, D.C. Neckers: *Experimental Organic Chemistry* (W.H. Freeman, New York 1998)
34.76 A. Ulman (Ed.): *Characterization of Organic Thin Films* (Butterworth-Heineman, Boston 1995)
34.77 C. Jung, O. Dannenberger, Y. Xu, M. Buck, M. Grunze: Self-assembled monolayers from organosulfur compounds: A comparison between sulfides, disulfides, and thiols, Langmuir **14**, 1103–1107 (1998)
34.78 W. Geyer, V. Stadler, W. Eck, M. Zharnikov, A. Golzhauser, M. Grunze: Electron-induced crosslinking of aromatic self-assembled monolayers: Negative resists for nanolithography, Appl. Phys. Lett. **75**, 2401–2403 (1999)
34.79 B. Bhushan, H. Liu: Nanotribological properties and mechanisms of alkylthiol and biphenyl thiol self-assembled monolayers studied by atomic force microscopy, Phys. Rev. B **63**(24), 245412 (2001)
34.80 H. Liu, B. Bhushan, W. Eck, V. Stadler: Investigation of the adhesion, friction, and wear properties of biphenyl thiol self-assembled monolayers by atomic force microscopy, J. Vac. Sci. Technol. A **19**, 1234–1240 (2001)
34.81 H. Liu, B. Bhushan: Investigation of nanotribological properties of alkylthiol and biphenyl thiol self-assembled monolayers, Ultramicroscopy **91**, 185–202 (2002)
34.82 S.R. Wasserman, Y.T. Tao, G.M. Whitesides: Structure and reactivity of alkylsiloxane monolayers formed by reaction of alkylchlorosilanes on silicon substrates, Langmuir **5**, 1074–1089 (1989)
34.83 E. Hoque, J.A. DeRose, P. Hoffmann, B. Bhushan, H.J. Mathieu: Alkylperfluorosilane self-assembled monolayers on aluminum: A comparison with alkylphosphonate self-assembled monolayers, J. Phys. Chem. C **111**, 3956–3962 (2007)
34.84 B. Bhushan: *Nanotribology and Nanomechanics I and II*, 3rd edn. (Springer, Berlin, Heidelberg 2011)
34.85 M.T. McDermott, J.B.D. Green, M.D. Porter: Scanning force microscopic exploration of the lubrication capabilities of *n*-alkanethiolate monolayers chemisorbed at gold: Structural basis of microscopic friction and wear, Langmuir **13**, 2504–2510 (1997)
34.86 J. Ruhe, V.J. Novotny, K.K. Kanazawa, T. Clarke, G.B. Street: Structure and tribological properties of ultrathin alkylsilane films chemisorbed to solid surfaces, Langmuir **9**, 2383–2388 (1993)

34.87 V. DePalma, N. Tillman: Friction and wear of self-assembled tricholosilane monolayer films on silicon, Langmuir **5**, 868–872 (1989)

34.88 X. Xiao, J. Hu, D.H. Charych, M. Salmeron: Chain length dependence of the frictional properties of alkylsilane molecules self-assembled on mica studied by atomic force microscopy, Langmuir **12**, 235–237 (1996)

34.89 A. Lio, D.H. Charych, M. Salmeron: Comparative atomic force microscopy study of the chain length dependence of frictional properties of alkanethiol on gold and alkylsilanes on mica, J. Phys. Chem. B **101**, 3800–3805 (1997)

34.90 H. Schonherr, G.J. Vancso: Tribological properties of self-assembled monolayers of fluorocarbon and hydrocarbon thiols and disulfides on Au(111) studied by scanning force microscopy, Mater. Sci. Eng. C **8–9**, 243–249 (1999)

34.91 V.V. Tsukruk, T. Nguyen, M. Lemieux, J. Hazel, W.H. Weber, V.V. Shevchenko, N. Klimenko, E. Sheludko: Tribological properties of modified MEMS surfaces. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 607–614

34.92 V.V. Tsukruk, V.N. Bliznyuk: Adhesive and friction forces between chemically modified silicon and silicon nitride surfaces, Langmuir **14**, 446–455 (1998)

34.93 M. Fujihira, Y. Tani, M. Furugori, U. Akiba, Y. Okabe: Chemical force microscopy of self-assembled monolayers on sputtered gold films patterned by phase separation, Ultramicroscopy **86**, 63–73 (2001)

34.94 H. Liu, B. Bhushan: Orientation and relocation of biphenyl thiol self-assembled monolayers, Ultramicroscopy **91**, 177–183 (2002)

34.95 R.J. Good, C.J.V. Oss: *Modern Approaches to Wettability – Theory and Applications* (Plenum, New York 1992)

34.96 M.H.V.C. Adao, B.J.V. Saramago, A.C. Fernandes: Estimation of the surface properties of styrene-acrylonitrile random copolymers from contact angle measurements, J. Coll. Inter. Sci. **217**, 94–106 (1999)

34.97 N.S. Tambe, B. Bhushan: A new atomic force microscopy based technique for studying nanoscale friction at high sliding velocities, J. Phys. D **38**, 764–773 (2005)

34.98 B. Bhushan, J. Ruan: Tribological performance of thin film amorphous carbon overcoats for magnetic recording disks in various environments, Surf. Coat. Technol. **68/69**, 644–650 (1994)

34.99 B. Bhushan, L. Yang, C. Gao, S. Suri, R.A. Miller, B. Marchon: Friction and wear studies of magnetic thin film rigid disks with glass-ceramic, glass and aluminum-magnesium substrates, Wear **190**, 44–59 (1995)

34.100 Y.F. Miura, M. Takenga, T. Koini, M. Graupe, N. Garg, R.L. Graham, T.R. Lee: Wettability of self-assembled monolayers generated from CF_3-terminated alkanethiols on gold, Langmuir **14**, 5821–5825 (1998)

34.101 M. Ratajczak-Sitarz, A. Katrusiak, Z. Kaluski, J. Garbarczyk: 4,4′-Biphenyldithiol, Acta Crystallogr. C **43**, 2389–2391 (1987)

34.102 J.I. Siepman, I.R. McDonald: Monte carlo simulation of the mechanical relaxation of a self-assembled monolayer, Phys. Rev. Lett. **70**, 453–456 (1993)

34.103 M. Garcia-Parajo, C. Longo, J. Servat, P. Gorostiza, F. Sanz: Nanotribological properties of octadecyltrichlorosilane self-assembled ultrathin films studied by atomic force microscopy: Contact and tapping modes, Langmuir **13**, 2333–2339 (1997)

34.104 D. DeVecchio, B. Bhushan: Localized surface elasticity measurements using an atomic force microscope, Rev. Sci. Instrum. **68**, 4498–4505 (1997)

34.105 E. Barrena, S. Kopta, D.F. Ogletree, D.H. Charych, M. Salmeron: Relationship between friction and molecular structure: Alkysilane lubricant films under pressure, Phys. Rev. Lett. **82**, 2880–2883 (1999)

34.106 N. Eustathopoulos, M. Nicholas, B. Drevet: *Wettability at High Temperature* (Pergamon, Amsterdam 1999)

34.107 S. Ren, S. Yang, Y. Zhao, T. Yu, X. Xiao: Preparation and characterization of ultrahydrophobic surface based on a stearic acid self-assembled monolayer over polyethyleneimine thin films, Surf. Sci. **546**, 64–74 (2003)

34.108 E.S. Clark: The molecular conformations of polytetrafluoroethylene: Forms II and IV, Polymer **40**, 4659–4665 (1999)

34.109 D.R. Lide: *CRC Handbook of Chemistry and Physics*, 85th edn. (CRC, Boca Raton 2004)

34.110 W.D. Callister: *Materials Science and Engineering*, 4th edn. (Wiley, New York 1997)

34.111 N.S. Tambe, B. Bhushan: Friction model for velocity dependence of nanoscale friction, Nanotechnology **16**, 2309–2324 (2005)

34.112 S.C. Clear, P.F. Nealey: The effect of chain density on the frictional behavior of surfaces modified with alkylsilanes and immersed in n-alcohols, J. Chem. Phys. **114**, 2802–2811 (2001)

34.113 N.S. Tambe, B. Bhushan: Durability studies of micro/nanoelectromechanical systems materials, coatings and lubricants at high sliding velocities (up to 10 mm/s) using a modified atomic force microscope, J. Vac. Sci. Technol. A **23**, 830–835 (2005)

34.114 INSPEC: *Properties of Silicon*, EMIS Data Reviews Series No. 4 (Institution of Electrical Engineers, London 1988)

34.115 MEMS materials database, http://www.memsnet.org/material (2002)

34.116 T. Hoshino: Adsorption of atomic and molecular oxygen and desorption of silicon monoxide on Si(111) surfaces, Phys. Rev. B. **59**, 2332–2340 (1999)

34.117 T.L. Cottrell: *The Strength of Chemical Bonds*, 2nd edn. (Butterworths Scientific, London 1958)

34.118 F. Tian, X. Xiao, M.M.T. Loy, C. Wang, C. Bai: Humidity and temperature effect on frictional properties of mica and alkylsilane monolayer self-assembled on mica, Langmuir **15**, 244–249 (1999)

34.119 R.D. Chambers: *Fluorine in Organic Chemistry* (Wiley, New York 1973)
34.120 B.A. Sexton, A.E. Hughes: A comparison of weak molecular adsorption of organic-molecules on clean copper and platinum surfaces, Surf. Sci. **140**, 227–248 (1984)
34.121 L.H. Dubois, B.R. Zegarski, R.G. Nuzzo: Fundamental studies of microscopic wetting on organics surfaces. 2. Interaction of secondary adsorbates with chemically textured organic monolayers, J. Am. Chem. Soc. **112**, 570–579 (1990)
34.122 M.C. McMaster, S.L.M. Schroeder, R.J. Madix: Molecular propane adsorption dynamics on Pt(110)-(1x2), Surf. Sci. **297**, 253–271 (1993)
34.123 G.J. Kluth, M. Sander, M.M. Sung, R. Maboudian: Study of the desorption mechanism of alkylsiloxane self-assembled monolayers through isotopic labeling and high resolution electron energy-loss spectroscopy experiments, J. Vac. Sci. Technol. A **16**, 932–936 (1998)
34.124 H.K. Kim, J.P. Lee, C.R. Park, H.T. Kwak, M.M. Sung: Thermal decomposition of alkylsiloxane self-assembled monolayers in air, J. Phys. Chem. B **107**, 4348–4351 (2003)

35. Nanoscale Boundary Lubrication Studies

Bharat Bhushan

Boundary films are formed by physisorption, chemisorption, and chemical reaction. A good boundary lubricant should have a high degree of interaction between its molecules and the solid surface. As a general rule, liquids are good lubricants when they are polar and thus able to grip solid surfaces (or be adsorbed). In this chapter, we present an overview of various perfluoropolyethers (PFPEs) and ionic liquid films. PFPEs exhibit the highest chemical and thermal stability. Ionic liquid films possess efficient heat transfer properties and some electrical conductivity properties of interest. We present a summary of nanodeformation, molecular conformation, and lubricant spreading studies, followed by an overview of the nanotribological properties of polar and nonpolar PFPEs and ionic liquid films studied by atomic force microscopy (AFM), and chemical degradation studies using a high vacuum tribotest apparatus.

35.1 **Nanodeformation, Molecular Conformation, Spreading, Nanotribological and Electrical Studies, and Environmental Effects of Commonly Used PFPE Lubricant Films** 1216
35.1.1 Nanodeformation, Molecular Conformation, and Spreading 1217
35.1.2 Nanotribological Studies and Environmental Effects 1219
35.2 **Nanotribological, Electrical, and Chemical Degradations Studies and Environmental Effects in Novel PFPE Lubricant Films** 1233
35.2.1 Nanotribological Studies 1234
35.2.2 Wear Detection by Surface Potential Measurements 1234
35.2.3 Wear Detection by Electrical Resistance Measurements of Z-TETRAOL and the Effect of Cycling 1237
35.2.4 Chemical Degradation and Environmental Studies 1239
35.3 **Nanotribological and Electrical Studies of Ionic Liquid Films** 1242
35.3.1 Monocationic Liquid Films 1244
35.3.2 Dicationic Ionic Liquid Films 1249
35.4 **Conclusion** 1257
References 1258

Boundary films are formed by physisorption, chemisorption, and chemical reaction. With physisorption, no exchange of electrons takes place between the molecules of the adsorbate and those of the adsorbent [35.1, 2]. The physisorption process typically involves van der Waals forces, which are relatively weak. In chemisorption, there is an actual sharing of electrons or an electron interchange between the chemisorbed species and the solid surface. The solid surfaces very strongly bond to the adsorption species through covalent bonds. Chemically reacted films are formed by the chemical reaction of the solid surface with the environment. Physisorbed films can be either monomolecularly or polymolecularly thick. Chemisorbed films are monomolecular, but stoichiometric films formed by chemical reaction can have a large film thickness. In general, the stability and durability of surface films decrease in the following order: chemically reacted films, chemisorbed films, and physisorbed films. A good boundary lubricant should have a high degree of interaction between its molecules and the solid surface. As a general rule, liquids are good lubricants when they are polar and thus able to grip solid surfaces (or be adsorbed). Polar lubricants contain reactive functional groups with low ionization potential or groups having high polarizability [35.1–5]. Boundary lubrication properties of lubricants are also dependent upon the molecular conformation and lubricant spreading [35.6–9].

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_35

Perfluoropolyether (PFPE) films, self-assembled mono layers (SAMs), and Langmuir–Blodgett (L–B) films can be used as boundary lubricants [35.1–5, 10–14]. PFPEs exhibit the following properties: low surface tension and a low contact angle, which allow easy spreading on surfaces and provide hydrophobic properties; chemical and thermal stability, which minimize degradation under use; low vapor pressure, which provides low outgassing; high adhesion to substrate via organofunctional bonds; and good lubricity, which reduces the interfacial friction and wear [35.4, 5]. PFPE films are used exclusively for lubrication of magnetic rigid disks and metal evaporated magnetic tapes to reduce friction and wear of the head-medium interface [35.4]. The structure of the lubricants employed at the head-medium interface has not changed substantially over the past decade, except with the variations in the end groups. However, the thickness of the PFPE film used to lubricate the disk has steadily decreased from multilayer thicknesses to the submonolayer thickness regime [35.5, 13].

Molecularly thick PFPE films are also being considered for lubrication purposes for micro/nanoelectromechanical systems (MEMS/NEMS) [35.5, 15–17]. It is well known that the properties of molecularly thick liquid films confined to solid surfaces can be dramatically different from those of the corresponding bulk liquid. In order to efficiently develop lubrication systems that meet the requirements of advanced rigid disk drives and MEMS/NEMS applications, the nanotribology of thin films of PFPE lubricants should be fully understood [35.12, 18, 19]. It is also important to understand lubricant-substrate interfacial interactions and the influence of operating environment on the nanotribological performance of molecularly thick PFPE liquid films.

Ionic liquids have been considered as potential lubricants [35.16, 20, 21]. They possess efficient heat transfer properties. They are also electrically conducting, which is of interest in various MEMS/NEMS applications such as atomic force microscope probe-based data recording [35.22–24].

In this chapter, we present an overview of PFPE and ionic liquid films. We first present a summary of nanodeformation, molecular conformation, and lubricant spreading studies and an overview of nanotribological and electrical studies of commonly used polar and nonpolar PFPE and ionic liquid films using atomic force microscopy (AFM) and chemical degradation studies using a high vacuum tribotest apparatus.

35.1 Nanodeformation, Molecular Conformation, Spreading, Nanotribological and Electrical Studies, and Environmental Effects of Commonly Used PFPE Lubricant Films

The molecular structures of two commonly used typical PFPE lubricants (Z-15 and Z-DOL 2000) are schematically shown in Fig. 35.1 (Fomblin, Solvay Solexis Inc., Milan, Italy). Z-15 has nonpolar $-CF_3$ end groups, whereas Z-DOL is a polar lubricant with hydroxyl(–OH) end groups. Their typical properties are summarized in Table 35.1. It shows that Z-15 and Z-DOL have almost the same density and surface tension, but Z-15 has a larger molecular weight and higher viscosity. Both of them have low surface tension, low vapor pressure, low evaporation weight loss, and good oxidative stability [35.3, 4]. For nanotribological characterization, a single crystal Si(100) wafer with a native oxide layer has generally been used as a substrate for deposition of molecularly thick lubricant films. Z-15 and Z-DOL films can be directly deposited on the Si(100) wafer by a dip-coating technique. The clean silicon wafer is vertically submerged into a dilute solution of lubricant in a hydrocarbon solvent (HT-70, Solvay Solexis Inc., Milan, Italy). The silicon wafers are vertically pulled up from the solution with a motorized stage at a constant speed for deposition of desired thicknesses of Z-15 and Z-DOL lubricants [35.5, 13, 25]. The lubricant film thickness obtained in dip coating is a function of concentration and pulling-up speed, among other factors. The Z-DOL film is bonded to the silicon substrate by heating the as-deposited Z-DOL samples in an oven at 150 °C for about 30 min. The native oxide layer

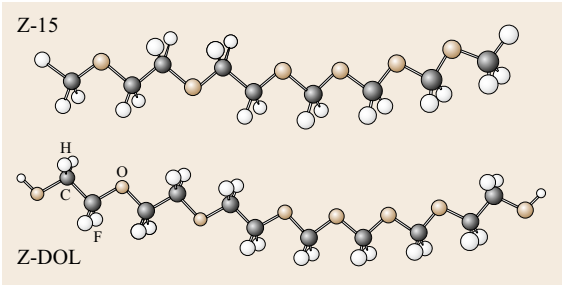


Fig. 35.1 Schematics of the molecular structures of Z-15 and Z-DOL 2000. Z-15 has nonpolar $-CF_3$ end groups, whereas Z-DOL is a polar lubricant with hydroxyl (−OH) end groups. In this figure, the m/n value, shown in Table 35.1, equals 2/3

Table 35.1 Typical properties of Z-15 and Z-DOL (Fomblin Z, Solvay Solexis Inc., Milan, Italy)

	Z-15	Z-DOL (2000)
Formula	$CF_3-O-(CF_2-CF_2-O)_m-(CF_2-O)_n-CF_3$[a]	$HO-CH_2-CF_2-O-(CF_2-CF_2-O)_m-(CF_2-O)_n-CF_2-CH_2-OH$[a]
Molecular weight (Da)	9100	2000
Density (ASTM D891) 20 °C (g/cm^3)	1.84	1.81
Kinematic viscosity (ASTM D445) (cSt)		
20 °C	148	85
38 °C	90	34
99 °C	25	–
Viscosity index (ASTM D2270)	320	–
Surface tension (ASTM D1331) (dyn/cm)		
20 °C	24	24
Vapor pressure (Torr)		
20 °C	1.6×10^{6}	2×10^{-5}
100 °C	1.7×10^{-5}	6×10^{-4}
Pour point (ASTM D972) (°C)	−80	–
Evaporation weight loss (ASTM D972) (%)		
149 °C, 22 h	0.7	–
Oxidative stability (°C)	–	320
Specific heat (cal/g°C) 38 °C	0.21	–

[a] $m/n \approx 2/3$.

of Si(100) wafer reacts with the –OH groups of the lubricants during thermal treatment [35.25–28]. Subsequently, fluorocarbon solvent (FC-72, 3M) washing of the thermal treated specimen removes loosely absorbed species leaving the chemically bonded phase on the substrate. The chemical bonding between Z-DOL molecules and the silicon substrate is illustrated in Fig. 35.2. The bonded and washed Z-DOL film is referred to as Z-DOL (fully bonded) in this chapter. The as-deposited Z-15 and Z-DOL films are mobile phase lubricants (i. e., liquid-like lubricant), whereas the Z-DOL (fully bonded) films are fully bonded soft solid phase (i. e., solid-like) lubricants.

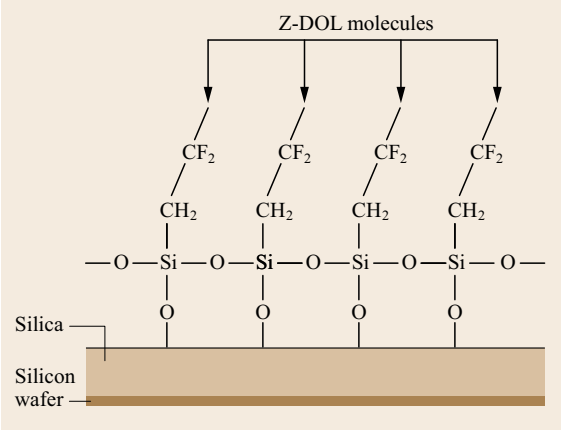


Fig. 35.2 Schematic of Z-DOL molecules that chemically bonded onto a Si(100) substrate surface (which has native oxide) after thermal treatment at 150 °C for 30 min

35.1.1 Nanodeformation, Molecular Conformation, and Spreading

Nanodeformation behavior of Z-DOL lubricants was studied using an AFM by *Blackman* et al. [35.29, 30]. Before bringing a tungsten tip into contact with a molecular overlayer, it was first brought into contact with a bare clean-silicon surface (Fig. 35.3). As the sample approaches the tip, the force initially is zero, but at point A the force suddenly becomes attractive (*top curve*), which increases until point B where the sample and tip come into intimate contact and the force becomes repulsive. As the sample is retracted, a pull-off force of 5×10^{-8} N (point D) is required to overcome adhesion between the tungsten tip and the silicon surface. When an AFM tip is brought into contact with an unbonded Z-DOL film, a sudden jump in adhesive contact is also observed. A much larger pull-off force is required to overcome the adhesion. The adhesion is initiated by the formation of a lubricant meniscus surrounding the tip. This suggests that the unbonded Z-DOL lubricant shows liquid-like behavior. However, when the tip was brought into contact with a lubricant film that was firmly bonded to the surface, the liquid-like behavior disappears. The initial attractive force (point A) is no longer sudden as with the liquid film, but rather gradually increases as the tip penetrates the film.

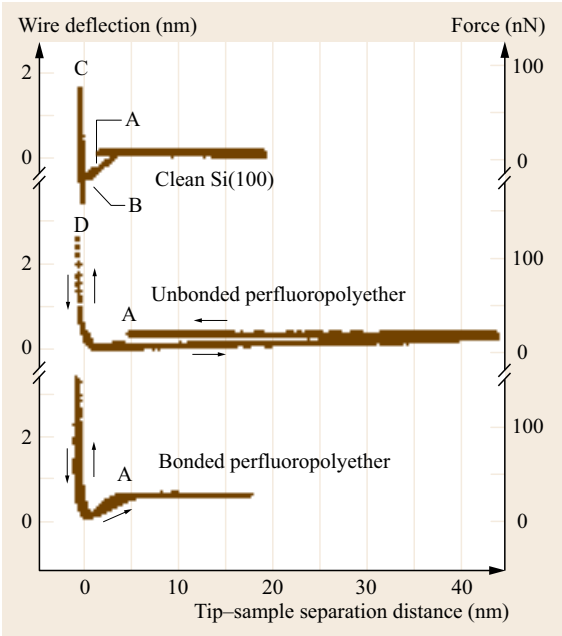


Fig. 35.3 Wire deflection (normal load) as a function of tip-sample separation distance curves comparing the behavior of a clean Si(100) surface to a surface lubricated with free and unbonded PFPE lubricant, and a surface where the PFPE lubricant film was thermally bonded to the surface (after [35.29])

According to *Blackman* et al. [35.29, 30], if the substrate and tip were infinitely hard with no compliance and/or deformation in the tip and sample supports, the line from B to C would be vertical with an infinite slope. The tangent to the force-distance curve at a given point is referred to as the stiffness at that point and was determined by fitting a least-squares line through the nearby data points. For silicon, the deformation is reversible (elastic) since the retracting (outgoing) portion of the curve (C to D) follows the extending (ingoing) portion (B to C). For bonded lubricant film, at the point where slope of the force changes gradually from attractive to repulsive, the stiffness changes gradually, indicating compression of the molecular film. As the load is increased, the slope of the repulsive force eventually approaches that of the bare surface. The bonded film was found to respond elastically up to the highest loads that could be applied, of 5 μN. Thus, bonded lubricant behaves as a soft polymer solid.

Figure 35.4 illustrates two extremes for the conformation on a surface of a linear liquid polymer without any reactive end groups and at submonolayer coverages [35.6, 8]. At one extreme, the molecules lie flat on the surface, reaching no more than their chain diameter δ above the surface. This would be the case if a strong attractive interaction exists between the molecules and the solid. On the other extreme, when a weak attraction exists between polymer segments and the solid, the molecules adopt conformation close to that of the molecules in the bulk, with the microscopic thickness equal to about the radius of gyration R_g. *Mate* and *Novotny* [35.8] used AFM to study conformation of 0.5–1.3 nm-thick Z-15 molecules on clean Si(100) surfaces. They found that the thickness measured by AFM is thicker than that measured by ellipsometry with the offset ranging from 3 to 5 nm. They found that the offset was the same for very thin submonolayer coverages. If the coverage is submonolayer and inadequate to make a liquid film, the relevant thickness is then the height (h_e) of the molecules extended above the solid surface. The offset should be equal to $2h_e$, assuming that the molecules extend the same height above both the tip and silicon surfaces. They thus concluded that the molecules do not extend more than 1.5–2.5 nm above a solid or liquid surface, much smaller than the radius of gyration of the lubricants ranging between 3.2 and 7.3 nm, and similar to the approximate cross-sectional diameter of 0.6–0.7 nm for the linear polymer chain. Consequently, the height that the molecules extend above the surface is considerably less than the diameter of gyration of the molecules and only a few molecular diameters in height, implying that the physisorbed molecules on a solid surface have an extended, flat conformation. They also determined the disjoining pressure of these liquid films from AFM measurements of the distance needed to break the liquid meniscus that forms between the solid surface and the AFM tip (also see [35.9]). For a monolayer thickness of about 0.7 nm, the disjoining pressure is about 5 MPa, indicating strong attractive interaction between the liquid molecules and the solid surface. The disjoining pressure decreases with increasing film thickness in a manner consistent with a strong attractive van der Waals interaction between the liquid molecules and the solid surface.

Rheological characterization shows that the flow activation energy of PFPE lubricants is weakly dependent on chain length, and is strongly dependent on the functional end groups [35.31]. PFPE lubricant films that contain polar end groups have lower mobility than those with nonpolar end groups of similar chain length [35.32]. The mobility of PFPE also depends on the surface chemical properties of the substrate. The spreading of Z-DOL on an amorphous carbon surface has been studied as a function of hydrogen or nitrogen content in the carbon film, using scanning microellipsometry [35.33]. The diffusion coefficient data presented in Fig. 35.5 are thickness dependent. It shows that the surface mobility of Z-DOL increased as hy-

drogen content increased, but decreased as nitrogen content increased. The enhancement of Z-DOL surface mobility by hydrogenation may be understood from the fact that the interactions between Z-DOL molecules and the carbon surface can be significantly weakened due to a reduction of the number of high-energy binding sites on the carbon surface. The stronger interactions between Z-DOL molecules and the carbon surface, as nitrogen content in the carbon coating increases, leads to the lowering of Z-DOL surface mobility.

Molecularly thick films may be sheared at very high shear rates, on the order of 10^8–10^9 s^{-1} during sliding, such as during magnetic disk drive operation. During such shear, lubricant stability is critical to the protection of the interface. For proper lubricant selection, viscosity at high shear rates and associated shear thinning need to be understood. Viscosity measurements of eight different types of PFPE films show that all eight lubricants display Newtonian behavior and their viscosity remains constant at shear rates up to 10^7 s^{-1} [35.34, 35].

35.1.2 Nanotribological Studies and Environmental Effects

AFM techniques have been used to investigate the nanotribological performance of PFPEs. *Mate* [35.36, 37], *O'Shea* et al. [35.38, 39], *Bhushan* et al. [35.18, 40], *Koinkar* and *Bhushan* [35.28, 41], *Bhushan* and *Sundararajan* [35.42], *Bhushan* and *Dandavate* [35.43], *Liu* and *Bhushan* [35.25], and *Palacio* and *Bhushan* [35.44, 45] used an AFM to provide insight into how PFPE lubricants function at the molecular level. *Mate* [35.36, 37] conducted friction experiments on bonded and unbonded Z-DOL, and he found that the coefficient of friction of the unbonded Z-DOL is about two times larger than the bonded Z-DOL (also see [35.38, 39]). *Koinkar* and *Bhushan* [35.28, 41] and *Liu* and *Bhushan* [35.25] studied the friction and wear performance of a Si(100) sample lubricated with Z-15, Z-DOL, and Z-DOL (fully bonded) lubricants. They found that using Z-DOL (fully bonded) could significantly improve the adhesion, friction, and wear performance of Si(100). They also discussed the lubrication mechanisms on the molecular level. *Bhushan* and *Sundararajan* [35.42] and *Bhushan* and *Dandavate* [35.43] studied the effect of tip radius and relative humidity on the adhesion and friction properties of Si(100) coated with Z-DOL (fully bonded).

In this section, we review the adhesion, friction, and wear properties of Z-15 and Z-DOL at various operating conditions (rest time, velocity, relative humidity, temperature, and tip radius). The experiments were carried out using a commercial AFM system with pyramidal Si_3N_4 and diamond tips. An environmentally controlled chamber and a thermal stage were used to perform relative humidity and temperature effect studies.

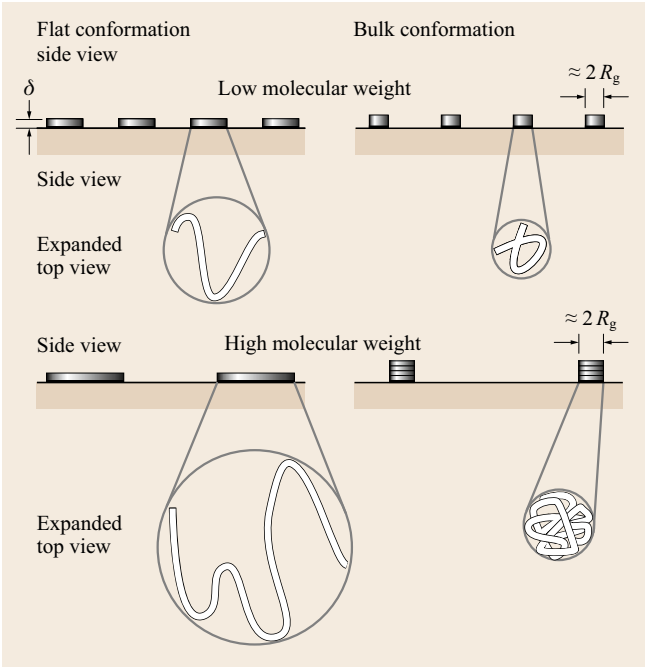


Fig. 35.4 Schematic representation of two extreme liquid conformations at the surface of the solid for low and high molecular weights at low surface coverage. δ is the cross-sectional diameter of the liquid chain and R_g is the radius of gyration of the molecules in the bulk (after [35.8])

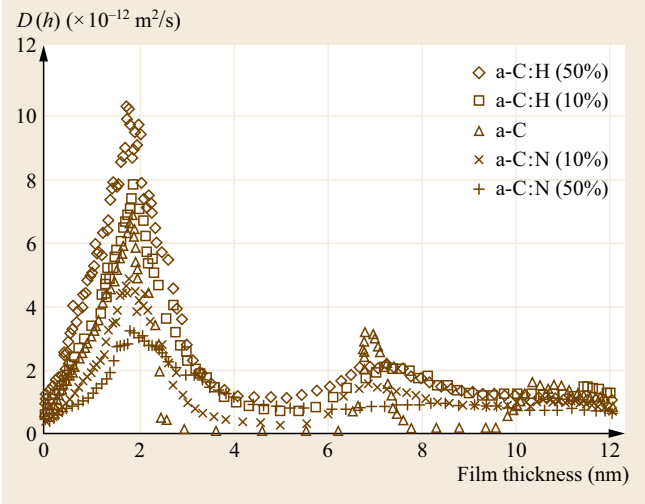


Fig. 35.5 Diffusion coefficient $D(h)$ as a function of lubricant film thickness for Z-DOL on different carbon films (after [35.33])

Friction and Adhesion

To investigate the friction properties of Z-15 and Z-DOL (fully bonded) films on Si(100), the friction force versus normal load curves were measured by making friction measurements at increasing normal loads [35.25]. The representative results of Si(100), Z-15, and Z-DOL (fully bonded) are shown in Fig. 35.6. An approximately linear response of all three samples is observed in the load range of 5−130 nN. The friction force of solid-like Z-DOL (fully bonded) is consistently smaller than that for Si(100), but the friction force of liquid-like Z-15 lubricant is higher than that of Si(100). *Sundararajan* and *Bhushan* [35.46] have studied the static friction force of silicon micromotors lubricated with Z-DOL by AFM. They also found that liquid-like lubricant of Z-DOL significantly increased the static friction force; on the contrary a solid-like Z-DOL (fully bonded) coating can dramatically reduce the static friction force. This is in good agreement with the results of *Liu* and *Bhushan* [35.25]. In Fig. 35.6, the nonzero value of the friction force signal at zero external load is due to the adhesive forces. It is well known that the following relationship exists between the friction force F and external normal load W [35.1, 2]

$$F = \mu(W + W_a)\,, \tag{35.1}$$

where μ is the coefficient of friction, and W_a is the adhesive force. Based on this equation and the data in Fig. 35.6, we can calculate the μ and W_a values. The coefficients of friction of Si(100), Z-15, and Z-DOL are 0.07, 0.09 and 0.04 respectively. Based on (35.1), the adhesive force values are obtained from the horizontal intercepts of the friction force versus normal load curves at a zero value of friction force. Adhesive force values of Si(100), Z-15, and Z-DOL are 52 nN, 91 nN, and 34 nN respectively.

The adhesive forces of these samples were also measured by using a force calibration plot (FCP) technique to obtain force-versus-distance curves. In this technique, the tip is brought into contact with the sample, and the maximum force needed to pull the tip and sample apart is measured as the adhesive force. Figure 35.7 shows the typical FCP curves of Si(100),

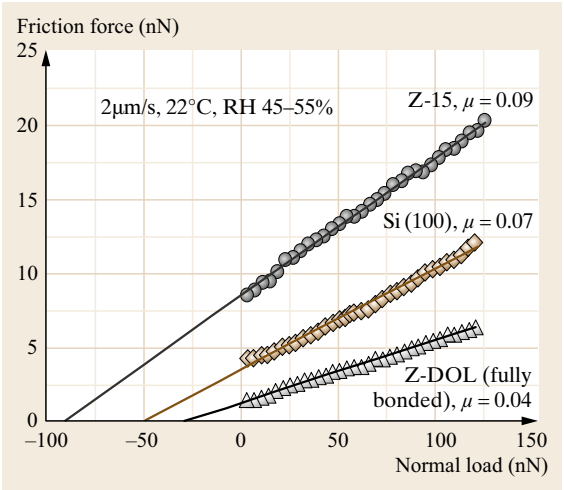


Fig. 35.6 Friction force versus normal load curves for Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film at 2 μm/s, and in ambient air sliding against a Si_3N_4 tip. Based on these curves, coefficient of friction μ and adhesion force of W_a can be calculated (after [35.25])

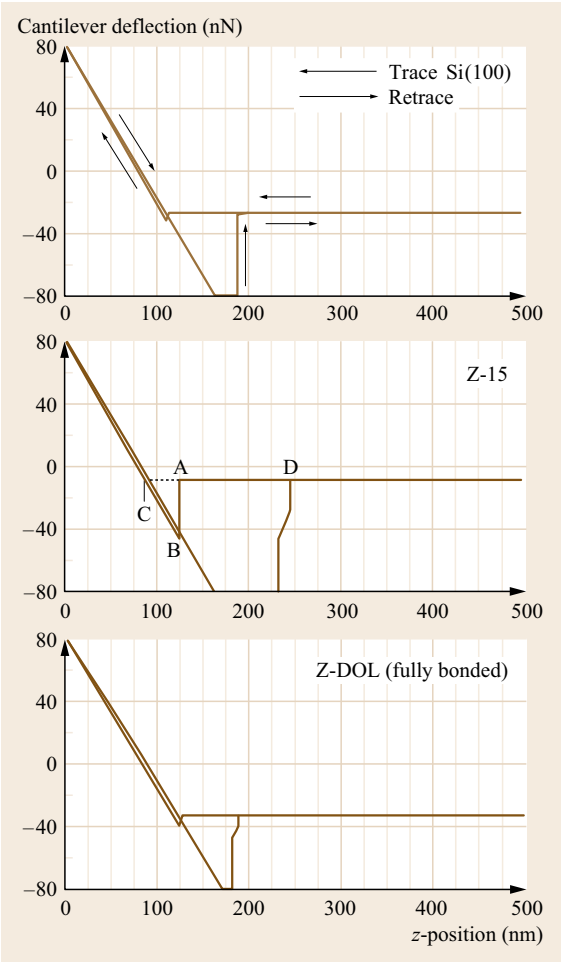


Fig. 35.7a–c Typical force calibration plots of **(a)** Si(100), **(b)** 2.8 nm-thick Z-15 film, and **(c)** 2.3 nm-thick Z-DOL (fully bonded) film in ambient air. The adhesive forces can be calculated from the horizontal distance between points C and D, and the cantilever spring constant of 0.58 N/m (after [35.25])

Z-15, and Z-DOL (fully bonded) [35.25]. As the tip approaches the sample within a few nanometers (A), an attractive force exists between the tip and the sample surfaces. The tip is pulled toward the sample, and contact occurs at point B on the graph. The adsorption of water molecules and/or presence of liquid lubricant molecules on the sample surface can also accelerate this so-called *snap-in*, due to the formation of a meniscus of the water and/or liquid lubricant around the tip. From this point on, the tip is in contact with the surface, and as the piezoelectric tube scanner extends further, the cantilever is further deflected. This is represented by the slope portion of the curve. As the piezoelectric tube scanner retracts, at point C the tip goes beyond the zero deflection (flat) line, because of the attractive forces, into the adhesive force regime. At point D, the tip snaps free of the adhesive forces and is again in free air. The adhesive force (pull-off force) is determined by multiplying the cantilever spring constant (0.58 N/m) by the horizontal distance between points C and D, which corresponds to the maximum cantilever deflection toward the samples before the tip is disengaged. Incidentally, the horizontal shift between the loading and unloading curves results from the hysteresis of the piezoelectric tube.

The adhesive forces of Si(100), Z-15, and Z-DOL (fully bonded) measured by FCP and friction force versus normal load plot are summarized in Fig. 35.8 [35.25]. The results measured by these two methods are in good agreement. Figure 35.8 shows that the presence of mobile Z-15 lubricant film increases the adhesive force as compared to that of Si(100). In contrast, the presence of solid phase Z-DOL (fully bonded) film reduces the adhesive force as compared that of Si(100). This result is in good agreement with the results of *Blackman* et al. [35.29] and *Bhushan* and *Ruan* [35.47]. Sources of adhesive forces between the tip and the sample surfaces are van der Waals attraction and long-range meniscus force [35.1, 2, 12]. Relative magnitudes of the forces from the two sources are dependent upon various factors including the distance between the tip and the sample surface, their surface roughness, their hydrophobicity, and relative humidity [35.48]. For most surfaces with some roughness, the meniscus contribution dominates at moderate to high humidities.

The schematic (*bottom*) in Fig. 35.8 shows the relative size and source of the meniscus. The native oxide layer (SiO_2) on the top of a Si(100) wafer exhibits hydrophilic properties, and some water molecules can be adsorbed on this surface. The condensed water will form a meniscus as the tip approaches the sample surface. In the case of a sphere (such as a single asperity

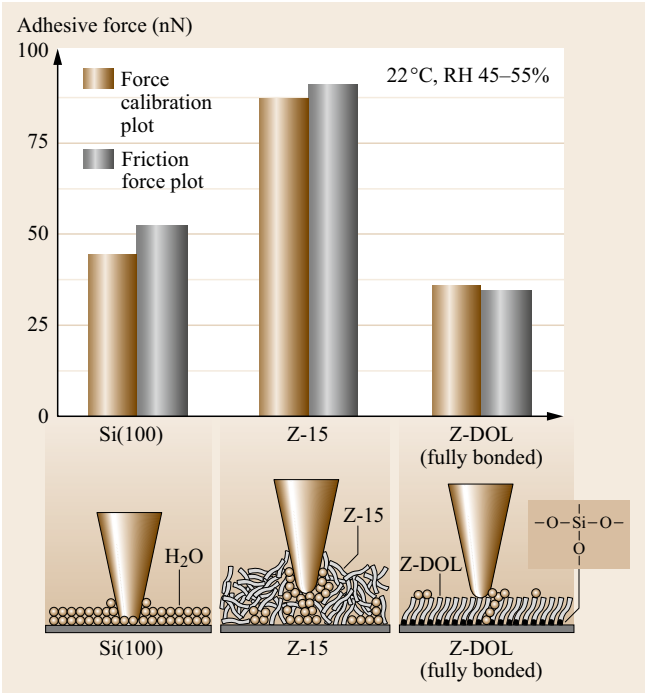


Fig. 35.8 Summary of the adhesive forces of Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film measured by force calibration plots and friction force versus normal load plots in ambient air. The schematic (*bottom*) showing the effect of a meniscus, formed between the AFM tip and the sample surface, on the adhesive and friction forces (after [35.25])

AFM tip) in contact with a flat surface, the attractive Laplace force F_L caused by the capillary is

$$F_L = 2\pi R\gamma(\cos\theta_1 + \cos\theta_2)\,, \qquad (35.2)$$

where R is the radius of the sphere, γ is the surface tension of the liquid against air, and θ_1 and θ_2 are the contact angles between liquid and flat and spherical surfaces respectively [35.1, 2, 49]. As the surface tension value of Z-15 (24 dyn/cm) is smaller than that of water (72 dyn/cm), the larger adhesive force in Z-15 cannot only be caused by the Z-15 meniscus. The nonpolar Z-15 liquid does not have complete coverage and strong bonding with Si(100). In the ambient environment, the condensed water molecules will permeate through the liquid Z-15 lubricant film and compete with the lubricant molecules present on the substrate. The interaction of the liquid lubricant with the substrate is weakened, and a boundary layer of the liquid lubricant forms puddles [35.28, 41]. This dewetting allows water molecules to be adsorbed on the Si(100) surface as aggregates along with Z-15 molecules, and both of them can form a meniscus while the tip approaches the surface. In ad-

dition, as the Z-15 film is pretty soft compared to the solid Si(100) surface, and penetration of the tip in the film occurs while pushing tip down, this leads to a large area of the tip involved in the formation of the meniscus at the tip-liquid interface (water aggregates along with Z-15). These two factors of the liquid-like Z-15 film result in higher adhesive force. It should also be noted that Z-15 has a higher viscosity compared to that of water, therefore Z-15 film provides higher resistance to sliding motion and results in a larger coefficient of friction. In the case of Z-DOL (fully bonded) film, both active groups of Z-DOL molecules are strongly bonded on the Si(100) substrate through the thermal and washing treatment, thus the Z-DOL (fully bonded) film has a relatively low free surface energy and cannot be displaced readily by water molecules or readily adsorb water molecules. Thus, the use of Z-DOL (fully bonded) can reduce the adhesive force. We further believe that the bonded Z-DOL molecules can be orientated under stress (behaves as a soft polymer solid), which facilitates sliding and reduces the coefficient of friction.

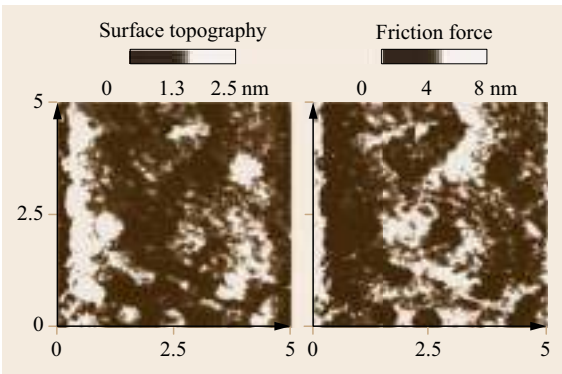


Fig. 35.9 Grayscale plots of the surface topography and friction force obtained simultaneously for unbonded 2.3 nm-thick Demnum-type PFPE lubricant film on silicon (after [35.28])

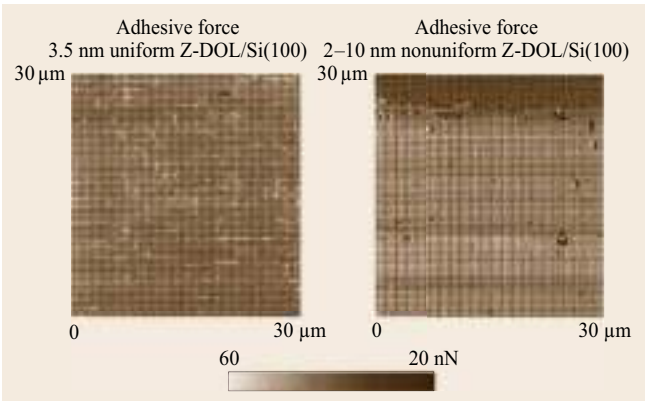


Fig. 35.10 Grayscale plots of the adhesive force distribution of a uniformly coated, 3.5 nm-thick unbonded Z-DOL film on silicon and a 3–10 nm-thick unbonded Z-DOL film on silicon that was deliberately coated nonuniformly by vibrating the sample during the coating process (after [35.43])

These studies suggest that if the lubricant films exist in a liquid-like phase, such as Z-15 films, they easily form menisci (by themselves and the adsorbed water molecules), and thus have higher adhesive force and higher friction force, whereas if the lubricant film exists in a solid-like phase, such as Z-DOL (fully bonded) films, they are hydrophobic with low adhesion and friction.

In order to study the uniformity of a lubricant film and its influence on friction and adhesion, friction force mapping and adhesive force mapping of PFPE have been carried out by *Koinkar* and *Bhushan* [35.41] and *Bhushan* and *Dandavate* [35.43] respectively. Figure 35.9 shows grayscale plots of surface topography and friction force images obtained simultaneously for an unbonded Demnum-type PFPE lubricant film on silicon [35.41]. Demnum-type PFPE lubricant (Demnum, Daikin, Japan) chains have $-CF_2-CH_2-OH$, a reactive end group on one end, whereas Z-DOL chains have the hydroxyl groups on both ends, as described earlier [35.5]. The friction force plot shows well distinguished low and high friction regions roughly corresponding to high and low surface height regions in the topography image (thick and thin lubricant regions). A uniformly lubricated sample does not show such a variation in the friction. Figure 35.10 shows the grayscale plots of the adhesive force distribution for silicon samples coated uniformly and nonuniformly with Z-DOL lubricant. It can be clearly seen that there exists a region that has an adhesive force distinctly different from the other region for the nonuniformly coated sample. This implies that the liquid film thickness is nonuniform giving rise to a difference in the meniscus forces.

Rest Time Effect

It is well known that in the computer rigid disk drive, the stiction force increases rapidly with an increase in rest time between the head and magnetic medium disk [35.4, 5]. Considering that stiction and friction are among the major issues that lead to the failure of computer rigid disk drives and MEMS, it is very important to find out if the rest time effect also exists on the nanoscale. First the rest time effect on the friction force, adhesive force, and coefficient of Si(100) sliding against a Si_3N_4 tip was studied (Fig. 35.11a) [35.25]. It was found that the friction and adhesive forces logarithmically increase up to a certain equilibrium time after which they remain constant. Figure 35.11a also

shows that the rest time does not affect the coefficient of friction. These results suggest that the rest time can result in growth of the meniscus, which causes a higher adhesive force, and in turn a higher friction force. However, in the whole testing range the friction mechanisms do not change with the rest time. Similar studies were also performed on Z-15 and Z-DOL (fully bonded) films. The results are summarized in Fig. 35.11b [35.25]. It is seen that a similar time effect has been observed on Z-15 film but not on Z-DOL (fully bonded) film.

An AFM tip in contact with a flat sample surface can be treated as a single asperity contact. Therefore, a Si_3N_4 tip in contact with Si(100) or Z-15/Si(100) can be modeled as a sphere in contact with a flat surface covered by a layer of liquid (adsorbed water and/or liquid lubricant), Fig. 35.12a. A meniscus forms around the contacting asperity and grows with time until equilibrium occurs [35.50]. The meniscus force, which is the product of meniscus pressure and meniscus area, depends on the flow of liquid phase toward the contact zone. The flow of the liquid towards the contact zone is governed by the capillary pressure P_c, which draws liquid into the meniscus, and the disjoining pressure Π, which tends to draw the liquid away from the meniscus. Based on the Young–Laplace equation, the

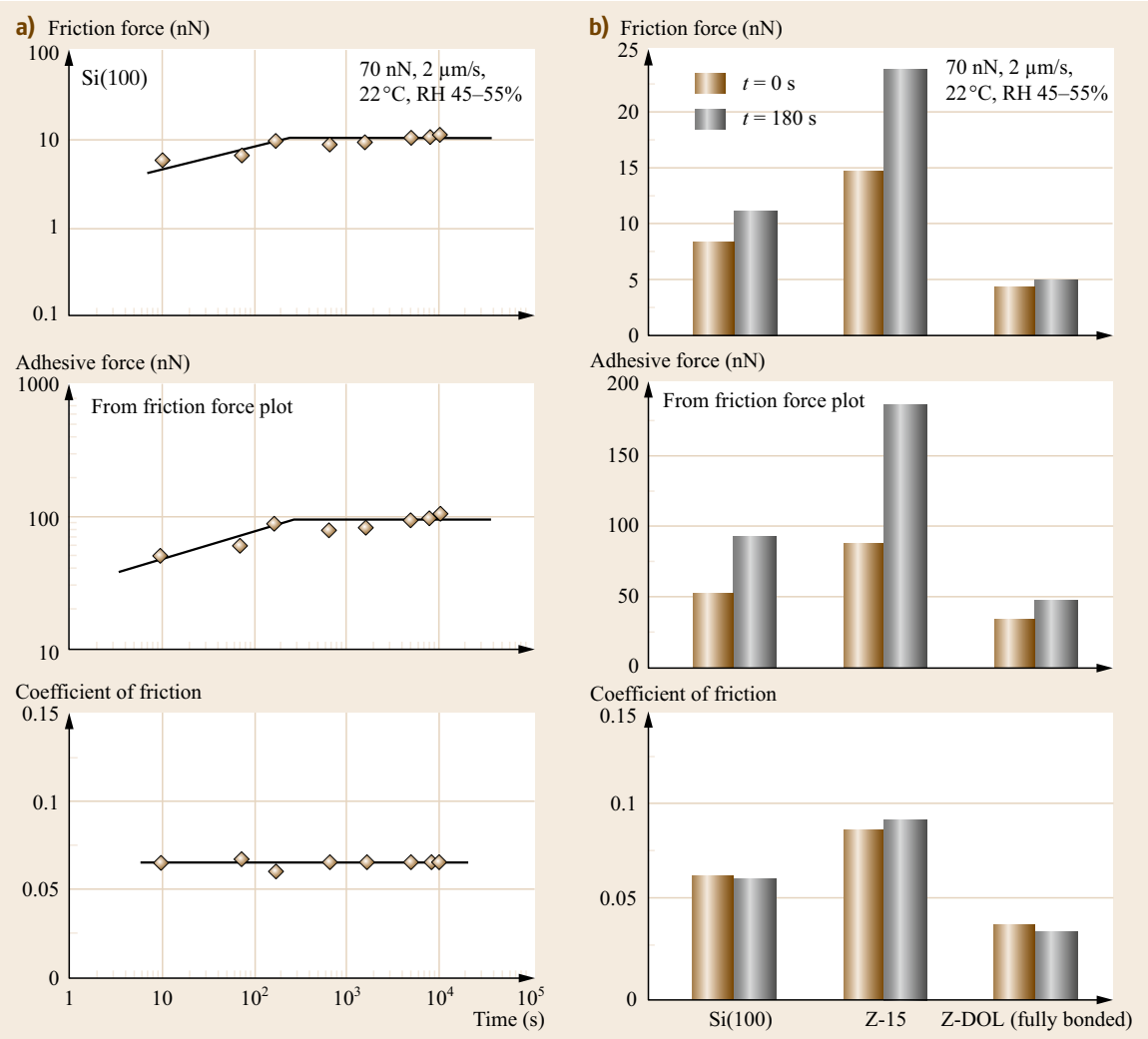


Fig. 35.11 **(a)** Rest time effect on friction force, adhesive force, and coefficient of friction of Si(100). **(b)** Summary of the rest time effect on friction force, adhesive force, and coefficient of friction of Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film. All of the measurements were carried out at 70 nN, 2 μm/s, and in ambient air (after [35.25])

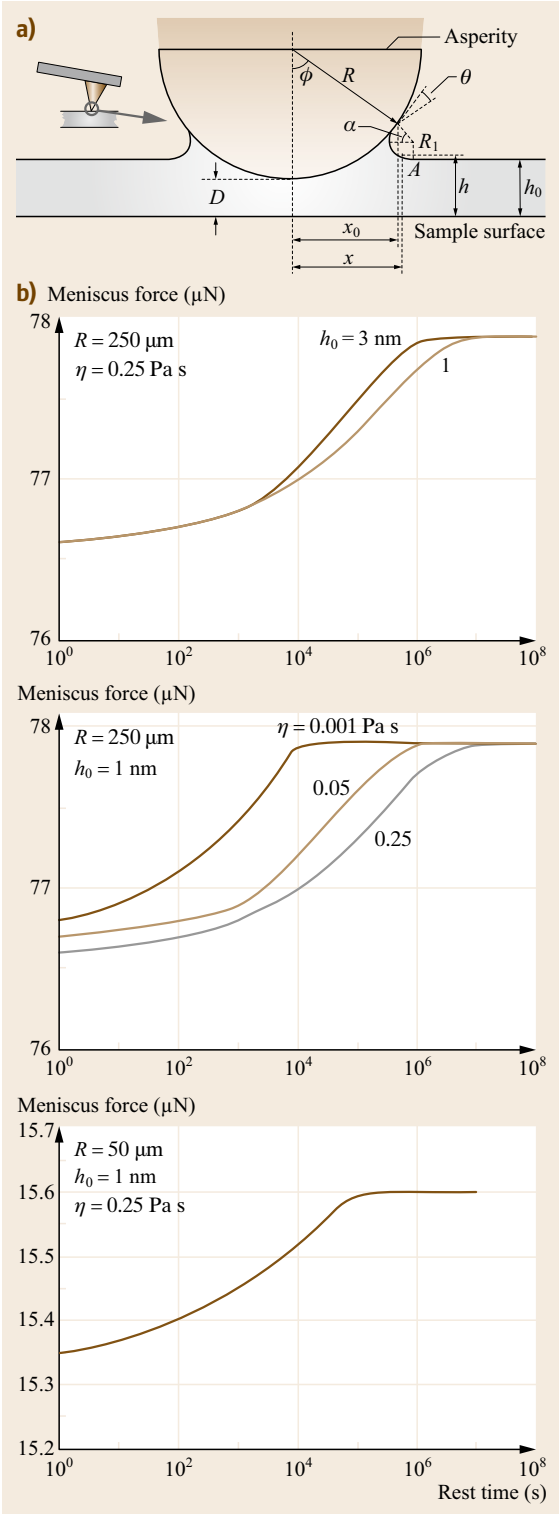


Fig. 35.12 **(a)** Schematic of a single asperity in contact with a smooth flat surface in the presence of a continuous liquid film when ϕ is large. **(b)** Results of the single asperity model. Effect of viscosity of the liquid, radius of the asperity and film thickness is studied with respect to the time-dependent meniscus force (after [35.50]) ◀

capillary pressure, P_c is

$$P_c = 2\kappa\gamma \,, \tag{35.3}$$

where 2κ is the mean meniscus curvature ($= \kappa_1 + \kappa_2$, where κ_1 and κ_2 are the curvatures of the meniscus in the contact plane and perpendicular to the contact plane). *Mate* and *Novotny* [35.8] have shown that the disjoining pressure decreases rapidly with increasing liquid film thickness in a manner consistent with a strong van der Waals attraction. The disjoining pressure, Π for these liquid films can be expressed as

$$\Pi = \frac{A}{6\pi h^3} \,, \tag{35.4}$$

where A is the Hamaker constant, and h is the liquid film thickness. The driving forces that cause the lubricant flow that result in an increase in the meniscus force are the disjoining pressure gradient due to a gradient in film thickness and capillary pressure gradient due to curved liquid-air interface. The driving pressure, P, can then written as

$$P = -2\kappa\gamma - \Pi \,. \tag{35.5}$$

Based on these three basic relationships, the following differential equation has been derived by *Chilamakuri* and *Bhushan* [35.50], which can describe the meniscus at time t

$$2\pi x_0 \left(D + \frac{x_0^2}{2R} - h_0\right) \frac{\mathrm{d}x_0}{\mathrm{d}t} = \frac{2\pi h_0^3 \gamma}{3\eta} \frac{(1+\cos\theta)}{D + a - h_0} - \frac{A x_0}{3\eta h} \cot\alpha \,, \tag{35.6}$$

where η is the viscosity of the liquid, and a is given as

$$a = R(1-\cos\phi) \approx \frac{R\phi^2}{2} \approx \frac{x_0^2}{2R} \,. \tag{35.7}$$

The differential (35.6) was solved numerically using Newton's iteration method. The meniscus force at any time t less than the equilibrium time is proportional to

the meniscus area and meniscus pressure ($2\kappa\gamma$), and it is given by

$$f_m(t) = 2\pi R\gamma(1+\cos\theta)\left(\frac{x_0}{(x_0)_{eq}}\right)^2\left(\frac{\kappa}{\kappa_{eq}}\right), \quad (35.8)$$

where $(x_0)_{eq}$ is the value of x_0 at the equilibrium time

$$\left[(x_0)_{eq}\right]^2 = 2R\left[\frac{-6\pi h_0^3(1+\cos\theta)}{A} + |h_0 - D|\right]. \quad (35.9)$$

This modeling work (at the microscale) showed that the meniscus force initially increases logarithmically with the rest time up to a certain equilibrium time, after which it remains constant. Equilibrium time decreases with an increase in liquid film thickness, a decrease in viscosity, and a decrease in the tip radius, Fig. 35.12b. This early numerical modeling work and the data at the nanoscale in Fig. 35.11a are in good agreement.

Velocity Effect

To investigate the velocity effect on friction and adhesion, the friction force versus normal load relationships of Si(100), Z-15, and Z-DOL (fully bonded) at different velocities were measured (Fig. 35.13) [35.25]. Based on these data, the adhesive force and coefficient of friction values can be calculated by (35.1). The variation of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of velocity are summarized in Fig. 35.14. It indicates that for Si(100) the friction force decreases logarithmically with increasing velocity. For Z-15, the friction force decreases with increasing velocity up to 10 μm/s, after which it remains almost constant. The velocity has very little effect on the friction force of Z-DOL (fully bonded); it reduced slightly only at very high velocity. Figure 35.14 also indicates that the adhesive force of Si(100) is increased when the velocity is higher than 10 μm/s. The adhesive force of Z-15 is reduced dramatically with a velocity increase up to 20 μm/s, after which it is reduced slightly, and the adhesive force of Z-DOL (fully bonded) is also decreased at high velocity. In the testing range of velocity, only the coefficient of friction of Si(100) decreases with velocity, but the coefficients of friction of Z-15 and Z-DOL (fully bonded) remain almost constant. This implies that the friction mechanisms of Z-15 and Z-DOL (fully bonded) do not change with the variation of velocity.

The mechanisms of the effect of velocity on adhesion and friction are explained based on schematics

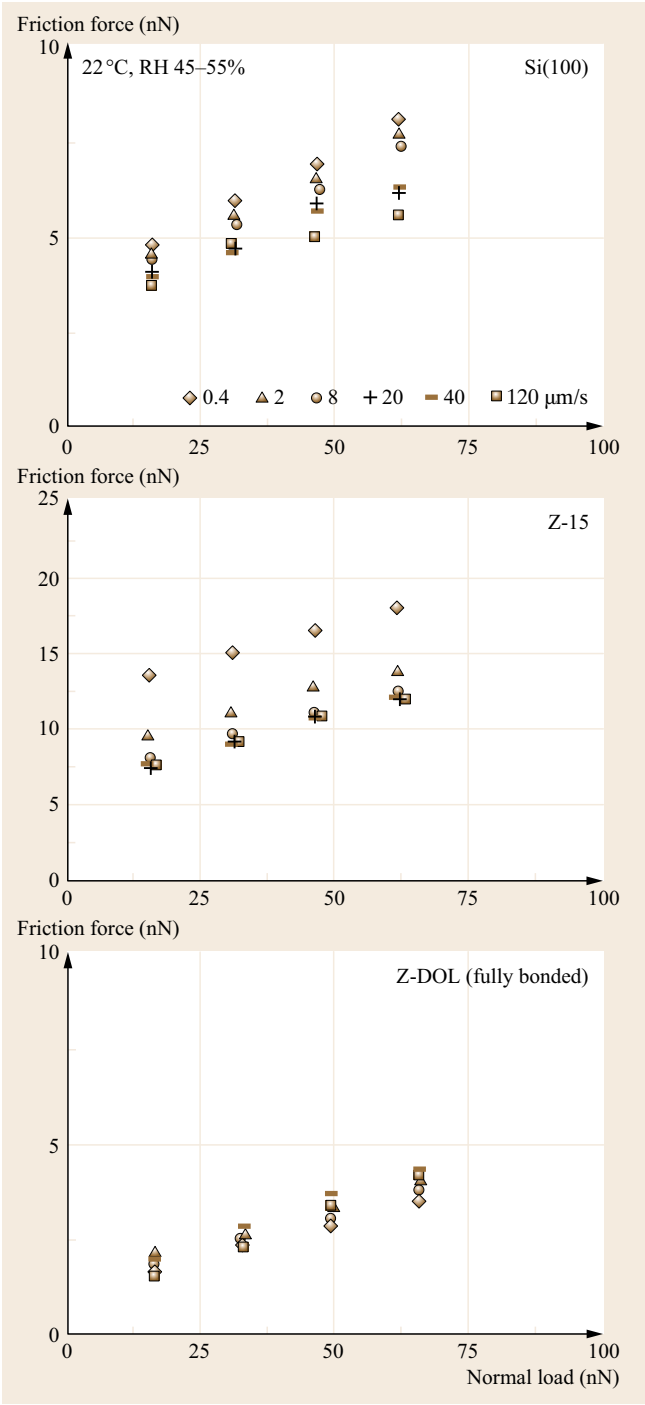


Fig. 35.13a–c Friction forces versus normal load data of **(a)** Si(100), **(b)** 2.8 nm-thick Z-15 film, and **(c)** 2.3 nm-thick Z-DOL (fully bonded) film at various velocities in ambient air (after [35.25])

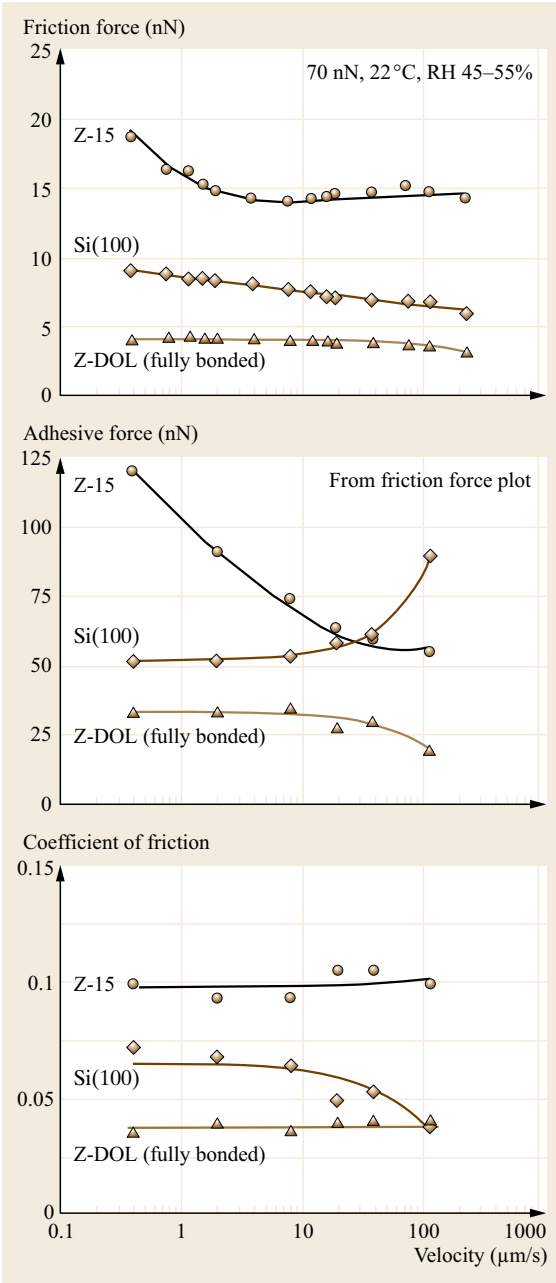


Fig. 35.14 The influence of velocity on the friction force, adhesive force and coefficient of friction of Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film at 70 nN, in ambient air

shown in Fig. 35.15. For Si(100), tribochemical reaction plays a major role. Although, at high velocity, the meniscus is broken and does not have enough time to rebuild, the contact stresses and high velocity lead to tribochemical reactions of the Si(100) wafer and Si_3N_4 tip, which have native oxide (SiO_2) layers with water molecules. The following reactions occur

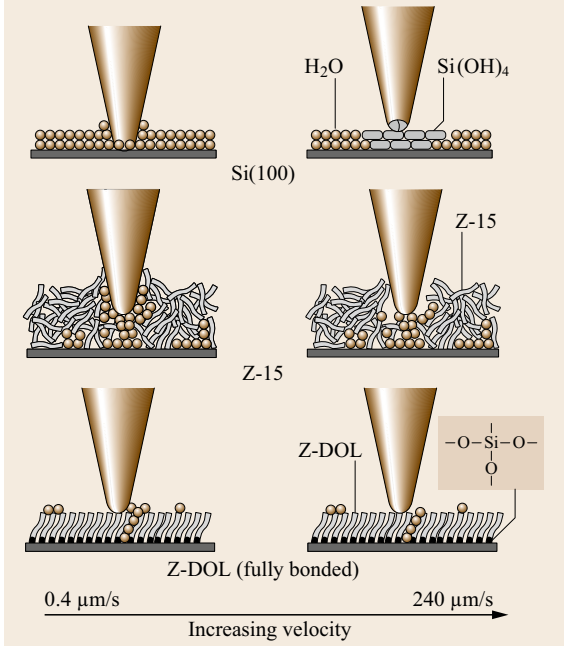


Fig. 35.15 The schematic shows the change of surface composition (by tribochemical reaction) and change of meniscus while increasing the velocity (after [35.25])

$$SiO_2 + 2H_2O \rightarrow Si(OH)_4 \quad (35.10)$$

$$Si_3N_4 + 16H_2O \rightarrow 3Si(OH)_4 + 4NH_4OH \quad (35.11)$$

The $Si(OH)_4$ is removed and continuously replenished during sliding. The $Si(OH)_4$ layer between the tip and Si(100) surface is known to be of low shear strength and causes a decrease in friction force and coefficient of friction in the lateral direction [35.51–55]. The chemical bonds of Si–OH between the tip and Si(100) surface induce a large adhesive force in the normal direction. For Z-15 film, at high velocity the meniscus formed by condensed water and Z-15 molecules is broken and does not have enough time to rebuild; therefore, the adhesive force and consequently friction force is reduced. For Z-DOL (fully bonded) film, the surface can adsorb a few water molecules in the ambient condition, and at high velocity these molecules are displaced, resulting in a slight decrease in friction force and adhesive force. Even at a high velocity range, the friction mechanisms for Z-15 and Z-DOL (fully bonded) films still are shearing of the viscous liquid and molecular orientation respectively. Thus the coefficients of friction of Z-15 and Z-DOL (fully bonded) do not change with velocity.

Koinkar and *Bhushan* [35.28, 41] have suggested that in the case of samples with mobile films such as condensed water and Z-15 films, alignment of liquid molecules (shear thinning) is responsible for the drop in friction force with an increase in scanning velocity. This could be another reason for the decrease in friction force for Si(100) and Z-15 film with velocity in this study.

Relative Humidity and Temperature Effect

The influence of relative humidity on friction and adhesion were studied in an environmentally controlled chamber. The friction force was measured by making measurements at increasing relative humidity, and the results are presented in Fig. 35.16 [35.25]. It shows that for Si(100) and Z-15 film, the friction force increases with a relative humidity increase up to relative humidity (RH) 45%, and then it shows a slight decrease with further increase in relative humidity. Z-DOL (fully bonded) has a smaller friction force than Si(100) and Z-15 in the whole testing range, and its friction force shows a relative apparent increase when the relative humidity is higher than RH 45%. For Si(100), Z-15, and Z-DOL (fully bonded), their adhesive forces increase with relative humidity, and their coefficients of friction increase with a relative humidity up to RH 45%, after which they decrease with a further increase of relative humidity. It is also observed that the humidity effect on Si(100) really depends on the history of the Si(100) sample. As the surface of Si(100) wafer readily adsorbs water in air, without any pretreatment the Si(100) used in our study almost reaches its saturated stage of adsorbing water, and is responsible for a smaller effect during increasing relative humidity. However, once the Si(100) wafer was thermally treated by baking at 150 °C for 1 h, a larger effect was observed.

The schematic in Fig. 35.17 shows that Si(100), because of its high free surface energy, can adsorb more water molecules with an increase in the relative humidity. As discussed earlier, for Z-15 film in a humid environment, the condensed water from the humid environment competes with the lubricant film present on the sample surface. Obviously, more water molecules also can be adsorbed on a Z-15 surface while increasing relative humidity. The more adsorbed water molecules in the case of Si(100), along with lubricant molecules in the Z-15 film case, form a larger water meniscus, which leads to an increase of friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 with humidity. However, at a very high humidity of RH 70%, large quantities of adsorbed water can form a continuous water layer that separates the tip and sample surface and acts as a kind of lubricant, which causes a decrease in the friction force and coefficient of friction. For Z-DOL (fully bonded) films, because of their hydrophobic surface properties, water molecules can be adsorbed and cause an increase in the adhesive force and friction force only at high humidity (≥ RH 45%).

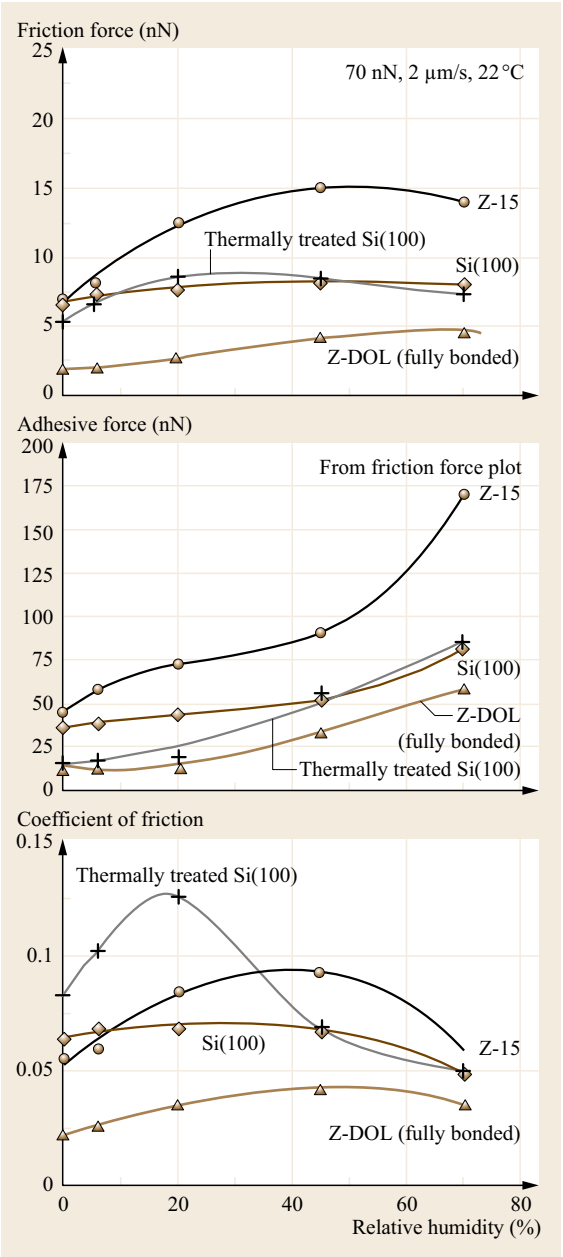


Fig. 35.16 The influence of relative humidity (RH) on the friction force, adhesive force, and coefficient of friction of Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film at 70 nN, 2 μm/s, and in 22 °C air.

The effect of temperature on friction and adhesion was studied using a thermal stage attached to the AFM.

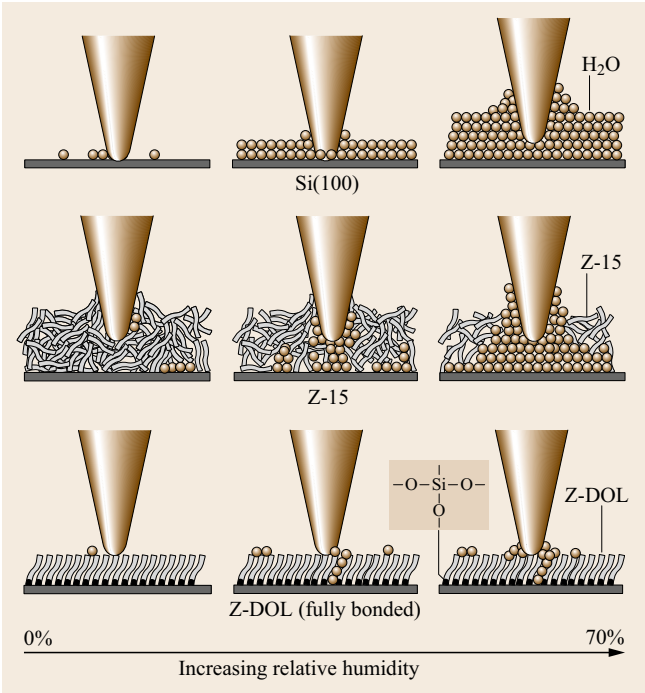


Fig. 35.17 Schematic shows the change of meniscus while increasing the relative humidity. In this figure, the thermally treated Si(100) represents the Si(100) wafer that was baked at 150 °C for 1 h in an oven (in order to remove the adsorbed water) just before it was placed in the 0% RH chamber (after [35.25])

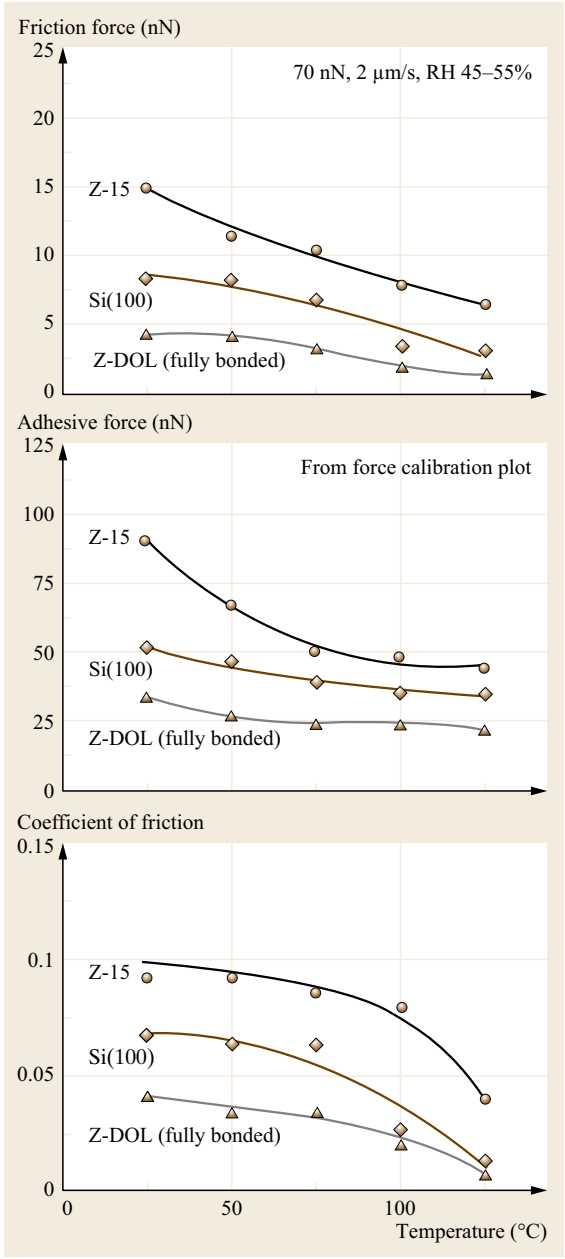


Fig. 35.18 The influence of temperature on the friction force, adhesive force, and coefficient of friction of Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film at 70 nN, at 2 μm/s, and in RH 40–50% air

The friction force was measured by making measurements at increasing temperature from 22 to 125 °C. The results are presented in Fig. 35.18 [35.25]. It shows that the increasing temperature causes a decrease of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded). The schematic in Fig. 35.19 indicates that at high temperature, desorption of water leads to a decrease of friction force, adhesive force and coefficient of friction for all of the samples. Besides that, the reduction of the surface tension of water also has a contribution to the decrease of friction and adhesion. For Z-15 film, the reduction of viscosity at high temperature has an additional contribution to the decrease of friction. In the case of Z-DOL (fully bonded) film, molecules are easier to orient at high temperature, which may also be responsible for the low friction.

Using a surface force apparatus, *Yoshizawa* and *Israelachvili* [35.56, 57] have shown that a change in the velocity or temperature induces phase transformations (from crystalline solid-like to amorphous, then to liquid-like) in surfactant monolayers, which are responsible for the observed changes in the friction force. Stick slip is observed in the low velocity regime of a few μm/s, and adhesion and friction first increase followed by a decrease in the temperature range of 0–50 °C. Stick slip at low velocity and adhesion and friction curves peaking at some particular temperature (observed in their study) have not been observed in the AFM study. This suggests that the phase transforma-

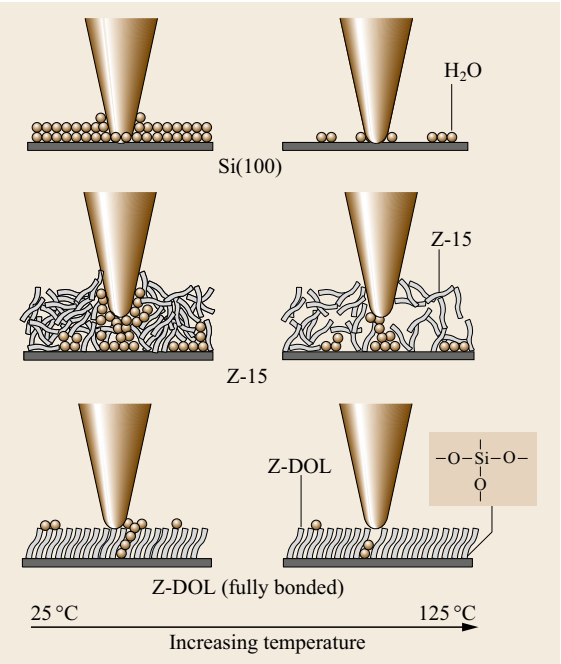


Fig. 35.19 The schematic shows that at high temperature, desorption of water decreases the adhesive forces. The reduced viscosity of Z-15 leads to the decrease of the coefficient of friction. High temperature facilitates orientation of molecules in Z-DOL (fully bonded) film, which results in a lower coefficient of friction (after [35.25])

tions may not happen in this study. This is because PFPEs generally have very good thermal stability [35.3, 4].

As a brief summary, the influence of velocity, relative humidity, and temperature on the friction force of Z-15 film is presented in Fig. 35.20. The changing trends are also addressed in this figure.

Tip Radius Effect

The tip radius and relative humidity affect adhesion and friction for unlubricated and lubricated surfaces [35.42, 43]. Figure 35.21a shows the variation of single-point adhesive force measurements as a function of tip radius on a Si(100) sample for several humidities. The adhesive force data are also plotted as a function of relative humidity for various tip radii. Figure 35.21a indicates that the tip radius has little effect on the adhesive forces at low humidities, but the adhesive force increases with tip radius at high humidity. Adhesive force also increases with an increase in humidity for all tips. The trend in adhesive forces as a function of tip radii and relative humidity, in Fig. 35.21a, can be explained by the presence of meniscus forces, which arise from the capillary condensation of water vapor from the environment.

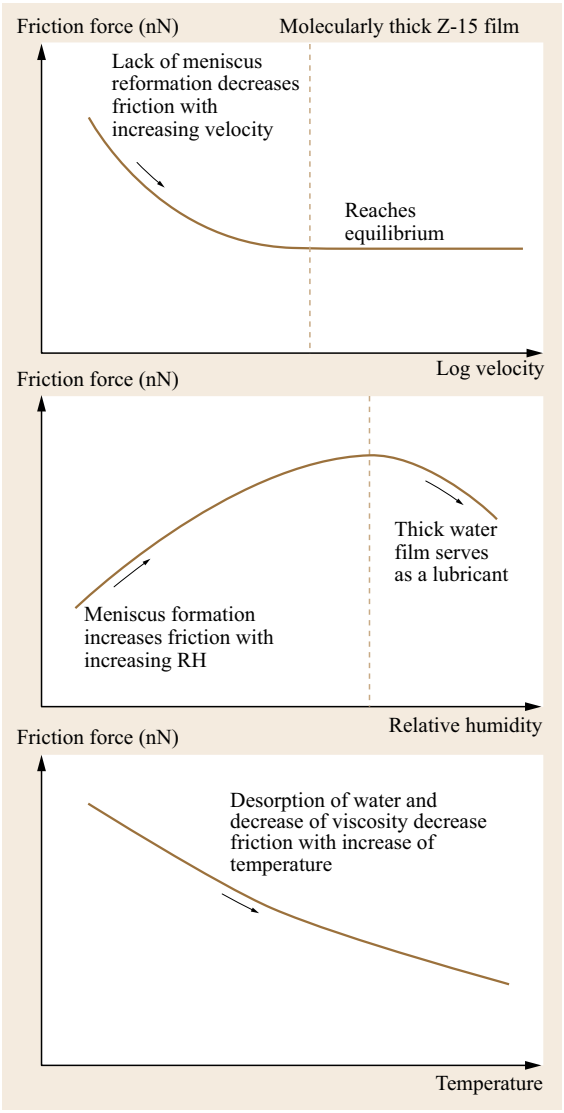


Fig. 35.20 Schematic showing the change of friction force of molecularly thick Z-15 films with log velocity, relative humidity, and temperature (after [35.25])

If enough liquid is present to form a meniscus bridge, the meniscus force should increase with an increase in tip radius based on (35.2). This observation suggests that the thickness of the liquid film at low humidities is insufficient to form continuous meniscus bridges and to affect adhesive forces in the case of all tips.

Figure 35.21a also shows the variation in coefficient of friction as a function of tip radius at a given humidity and as a function of relative humidity for a given tip radius on the Si(100) sample. It can be observed that for RH 0%, the coefficient of friction is about the same for the tip radii except for the largest tip, which shows

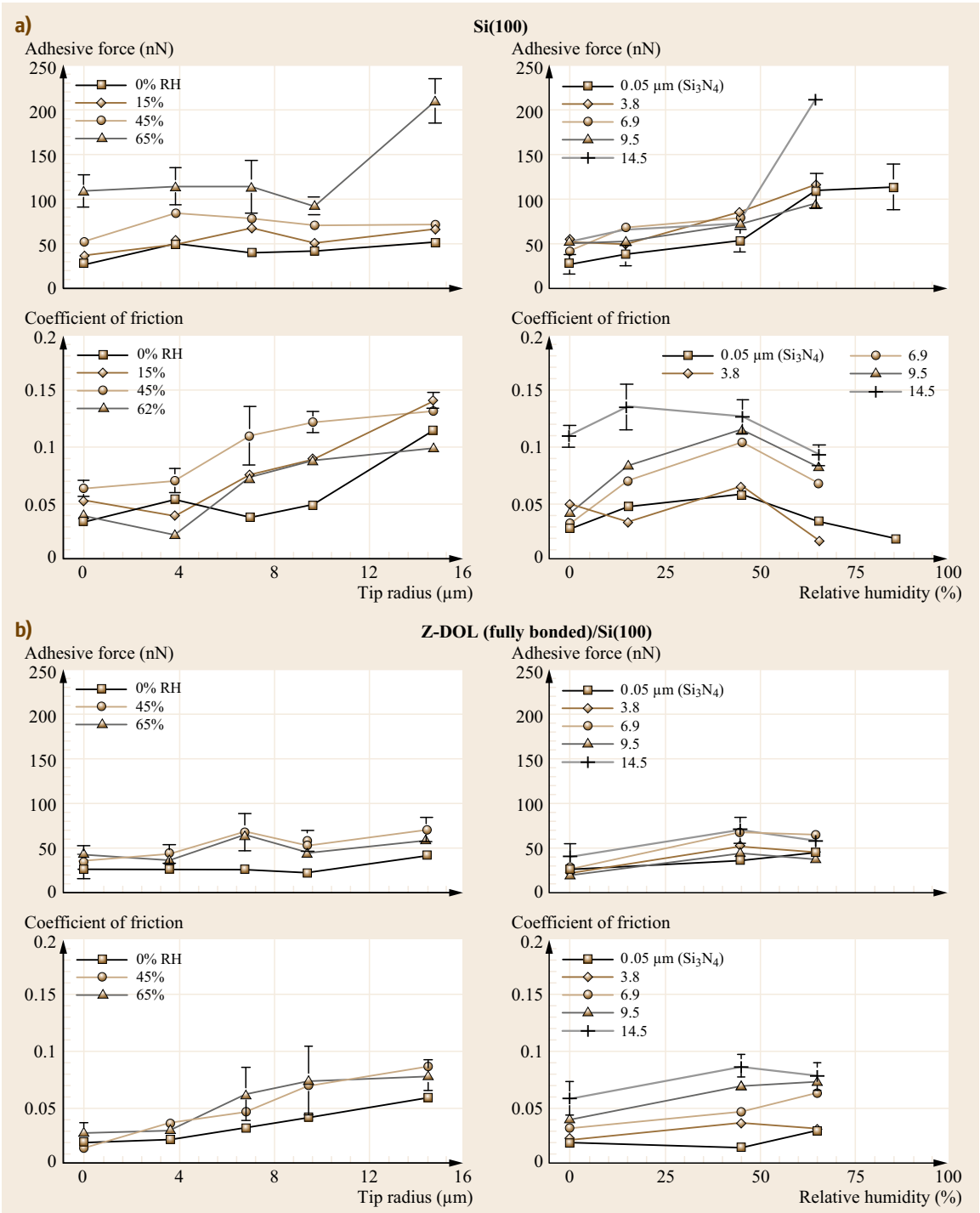


Fig. 35.21 Adhesive force and coefficient of friction as a function of tip radius at several humidities and as a function of relative humidity at several tip radii on **(a)** Si(100), and **(b)** 0.5 nm Z-DOL (fully bonded) films (after [35.42])

a higher value. At all other humidities, the trend consistently shows that the coefficient of friction increases with tip radius. An increase in friction with tip radius at low to moderate humidities arises from increased contact area (i. e., higher van der Waals forces) and higher values of shear forces required for larger contact area. At high humidities, similar to adhesive force data, an increase with tip radius occurs because of both contact area and meniscus effects. It can be seen that for all tips, the coefficient of friction increases with humidity to about RH 45%, beyond which it starts to decrease. This is attributed to the fact that at higher humidities, the adsorbed water film on the surface acts as a lubricant between the two surfaces [35.25]. Thus the interface is changed at higher humidities, resulting in lower shear strength and hence lower friction force and coefficient of friction.

Figure 35.21b shows adhesive forces as a function of tip radius and relative humidity on Si(100) coated with a 0.5 nm-thick Z-DOL (fully bonded) film. Adhesive forces for all the tips with the Z-DOL (fully bonded) lubricated sample are much lower than those experienced on unlubricated Si(100) shown in Fig. 35.21a. The data also show that even at a monolayer thickness of the lubricant, there is very little variation in adhesive forces with tip radius at a given humidity. For a given tip radius, the variation in adhesive forces with relative humidity indicates that these forces slightly increase from RH 0 to RH 45%, but remain more or less the same with a further increase in humidity. This is seen even with the largest tip, which indicates that the lubricant is indeed hydrophobic; there is some meniscus formation at humidities higher than RH 0%, but the formation is very minimal and does not increase appreciably even up to RH 65%. Figure 35.21b also shows the coefficient of friction for various tips at different humidities for the Z-DOL (fully bonded) lubricated sample. Again, all the values obtained with the lubricated sample are much lower than the values obtained on unlubricated Si(100) shown in Fig. 35.21a. The coefficient of friction increases with tip radius for all humidities as was seen on unlubricated Si(100) due to an increase in the contact area. Similar to the adhesive forces, there is an increase in friction from RH 0 to RH 45% due to a contribution from an increased number of menisci bridges. However, there is very little additional water film forming due to the hydrophobicity of the Z-DOL (fully bonded) layer thereafter, and consequently, the coefficient of friction does not change appreciably, even with the largest tip. These findings show that even a monolayer of Z-DOL (fully bonded) offers good hydrophobic performance of the surface.

Wear

To study the durability of lubricant films at the nanoscale, the friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of the number of scanning cycles was measured (Fig. 35.22) [35.25]. As observed earlier, the friction force of Z-15 is higher than that of Si(100), and Z-DOL (fully bonded) has the lowest value. During cycling, the friction force of Si(100) shows a slight variation during the initial few cycles then remains constant. This is related to the removal of the top adsorbed layer. In the case of Z-15 film, the friction force shows an increase during the initial few cycles and then approaches higher and stable values. This is believed to be caused by the attachment of the Z-15 molecules onto the tip. The molecular interaction between these attached molecules to the tip and molecules on the film surface is responsible for an increase in the friction. However, after several scans, this molecular interaction

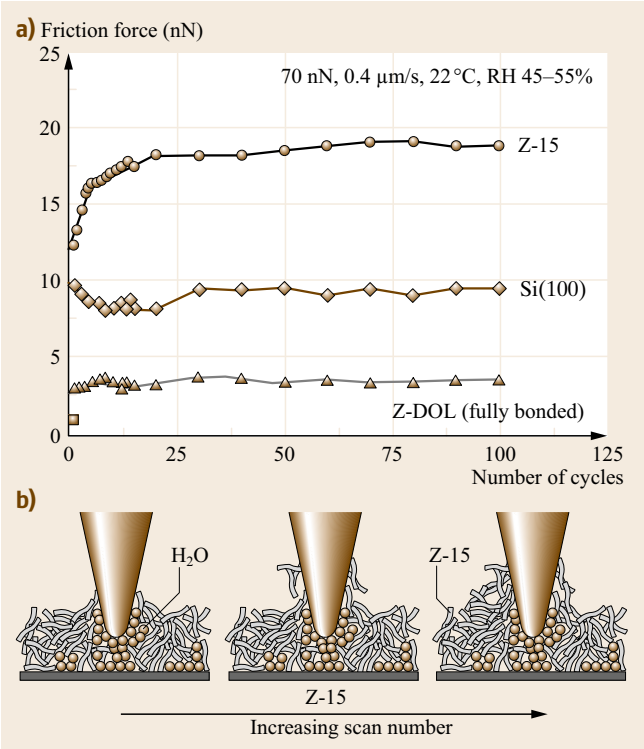


Fig. 35.22 Friction force versus number of sliding cycles for Si(100), 2.8 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film at 70 nN, 0.8 µm/s, and in ambient air. Schematic (*bottom*) shows that some liquid Z-15 molecules can be attached to the tip. The molecular interaction between the molecules attached to the tip with the Z-15 molecules in the film results in an increase of the friction force with multiscanning (after [35.25])

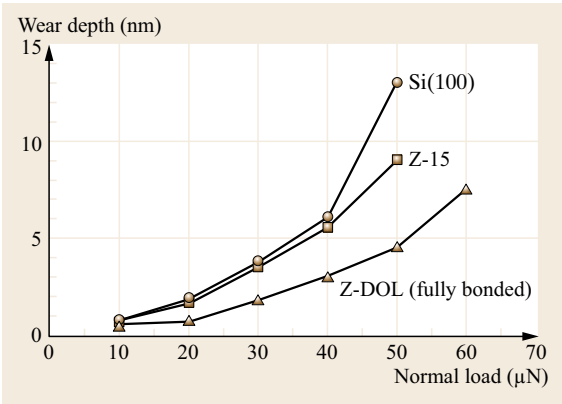


Fig. 35.23 Wear depth as a function of normal load using a diamond tip for Si(100), 2.9 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) after one cycle (after [35.28])

reaches equilibrium, and after that the friction force and coefficient of friction remain constant. In the case of Z-DOL (fully bonded) film, the friction force starts out low and remains low during the entire test for 100 cycles. It suggests that Z-DOL (fully bonded) molecules do not get attached or displaced as readily as in Z-15.

Koinkar and *Bhushan* [35.28, 41] conducted wear studies using a diamond tip at high loads. Figure 35.23 shows the plots of wear depth as a function of normal force, and Fig. 35.24 shows the wear profiles of the worn samples at 40 μN normal load. The 2.3 nm-thick Z-DOL (fully bonded) lubricated sample exhibits better wear resistance than unlubricated and 2.9 nm-thick Z-15 lubricated silicon samples. Wear resistance of the Z-15 lubricated sample is little better than that of the unlubricated sample. The Z-15 lubricated sample shows debris inside the wear track. Since Z-15 is a liquid lubricant, the debris generated is held by the lubricant, and it becomes sticky, which moves inside the wear track and does damage (Fig. 35.23). These results suggest that Z-DOL (fully bonded) exhibits higher wear resistance of the substrate as compared to that for Z-15.

To study the effect of the degree of chemical bonding, durability tests were conducted on both fully bonded and partially bonded Z-DOL films. Durability results for Z-DOL (fully bonded) and Z-DOL bonded and unwashed Z-DOL (partially bonded) (which is a partially bonded film, and contains both bonded and mobile phase lubricants) with different film thicknesses are shown in Fig. 35.25 [35.41]. Thicker films, such as Z-DOL (partially bonded) with a thickness of 4.0 nm (bonded/mobile = 2.3 nm/1.7 nm) exhibit behavior similar to 2.3 nm-thick Z-DOL (fully bonded) film. Figure 35.25 also indicates that Z-DOL (fully bonded) and Z-DOL (partially bonded) films with a thinner film thickness exhibit a higher friction value. Compar-

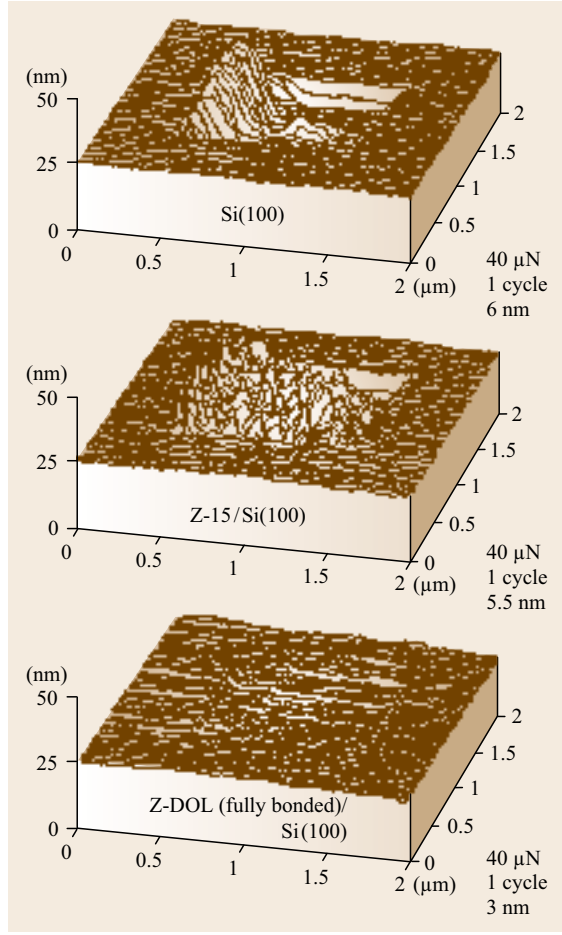


Fig. 35.24 Wear profiles for Si(100), 2.9 nm-thick Z-15 film, and 2.3 nm-thick Z-DOL (fully bonded) film after wear studies using a diamond tip. Normal force used and wear depths are listed in the figure (after [35.28])

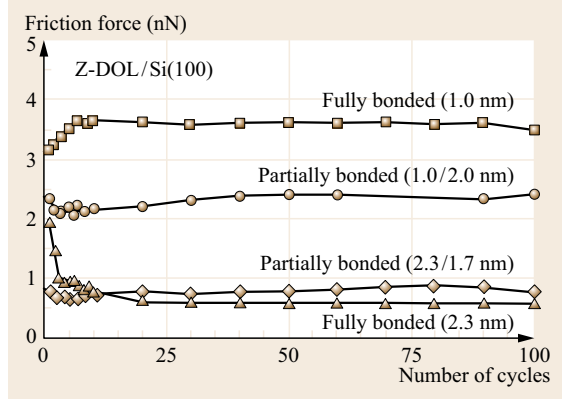


Fig. 35.25 Friction force as a function of number of cycles using a Si_3N_4 tip at a normal load of 300 nN for Z-DOL (fully bonded) and Z-DOL (partially bonded) films with different film thicknesses (after [35.41])

ing 1.0 nm-thick Z-DOL (fully bonded) with 3.0 nm (bonded/mobile = 1.0 nm/2.0 nm)-thick Z-DOL (partially bonded), the Z-DOL (partially bonded) film exhibits lower and stable friction values. This is because the mobile phase on a surface acts as a source of lubricant replenishment. Similar conclusions have also been reported by *Ruhe* et al. [35.58], *Bhushan* and *Zhao* [35.13], and *Eapen* et al. [35.59]. All of them indicate that using partially bonded Z-DOL films can dramatically reduce friction and improve the wear life.

35.2 Nanotribological, Electrical, and Chemical Degradations Studies and Environmental Effects in Novel PFPE Lubricant Films

Electrical properties of lubricant films are of interest in various MEMS/NEMS applications. Changes in the surface potential and electrical resistance can be measured during sliding using an AFM [35.20, 24, 44, 45, 60–62]. These techniques are also useful for wear detection and for studying the initiation of wear [35.15, 19, 60]. *Palacio* and *Bhushan* [35.44, 45, 61] carried out nanotribological studies on various novel PFPE lubricant films and monitored the electrical properties as well. Chemical degradation studies and environmental effects on various PFPE lubricant films on a Si(100) wafer and amorphous carbon-coated (diamond-like carbon) magnetic tapes were carried out in a macroscale configuration by *Tao* and *Bhushan* [35.63], *Bhushan* and *Tao* [35.64], and *Bhushan* et al. [35.65].

The structure and properties of several novel PFPE lubricants commonly used in the lubrication of magnetic rigid disks – Z-TETRAOL 2000 and A20H-2000 – are presented in Table 35.2 [35.63]. Z-DOL 2000 is also included for comparison. Z-TETRAOL (Solvay Solexis Inc.) is a derivate of PFPE. The backbone of Z-TETRAOL is the same as the conventional PFPE lubricant Z-DOL (Solvay Solexis Inc.) as described earlier. The difference between Z-TETRAOL and Z-DOL is that Z-TETRAOL has two hydroxyl groups at each end while Z-DOL has one hydroxyl group at each end. It is believed that two hydroxyl bonds are expected to lead to stronger interaction with the substrate. However, the Z-TETRAOL lubricant film is less mobile because of larger viscosity, which may lead to shorter durability. A20H (Moresco, Japan) is a PFPE lubricant with a cyclotriphosphazene at one end and a hydroxyl at the other end. The backbone of A20H is also the same as that of Z-DOL. Phosphazene lubricants (such as X1-P) have been used as additives in the data storage industry because they exhibit better durability in high-humidity environments [35.5, 13]. It is believed that X1-P coats the mating surface and makes it hydrophobic and minimizes stiction and improves durability. Studies have shown that A20H exhibits less thinning or reduced mobility during drive rotation [35.63]. The durability of these less mobile lubricants could be less than that of the highly mobile Z-DOL + X1-P systems. A mixture of Z-DOL and A20H is known to provide longer durability and good performance at high humidity.

The lubricants were applied on single-crystal Si(100) with a native oxide layer on the surface using the dip coating technique by *Tao* and *Bhushan* [35.63] and *Palacio* and *Bhushan* [35.44]. The wafer was ultrasonicated in acetone followed by methanol for 10 min each. This was then followed by soaking in the solvent HFE 7100 (3M, St. Paul, MN), which consists of isomers of methoxynonafluorobutane ($C_4F_9OCH_3$). The cleaned wafer was vertically submerged into a beaker containing a dilute solution of the lubricant in HFE 7100 for 10 min and then pulled off. The lubricated sample used without postthermal treatment is referred

Table 35.2 Chemical structures and selected properties of several PFPE lubricants (data obtained from manufacturers' data sheets)

Lubricant	End group (X[a])	Molecular weight (amu)	Density (10^3 kg/m^3 at 20 °C)	Kinematic viscosity (mm^2/s)	Surface tension (mJ/m^2)	Vapor pressure (Torr)	
						20 °C	100 °C
Z-DOL 2000[b]	$-CF_2-CH_2-OH$	2000	1.81	85	24	2×10^{-5}	6×10^{-4}
Z-TETRAOL 2000[b]	$-CF_2-CH_2-O-CH_2-CH(OH)-CH_2-OH$	2300	1.75	2000	–	5×10^{-7}	2×10^{-4}
A20H-2000[c]	$-CH_2-OH$ and $-CH_2-N_3P_3\ (OC_6H_4-CF_3)_5$	3000	1.7	65	22	–	–

[a] The whole molecular chain: $X-CF_2-O-(CF_2-CF_2-O)_m-(CF_2-O)_n-CF_2-X$ ($m/n = 2/3$)
[b] Solvay Solexis, Inc., Thorofare, NJ
[c] Moresco Matsumura Oil Research Corp., Kobe-city, Hyogo, Japan

to as untreated. Partially bonded samples were prepared by heating at 150 °C for 30 min after dip coating, while the fully bonded samples were heated at 150 °C for 30 min and washed in the HFE 7100 solvent to remove the mobile fraction. The lubricant-coated silicon samples were then measured with an ellipsometer and the coating thickness was found to be about 1, 3, and 7 nm for the fully bonded, partially bonded and untreated samples respectively. A schematic of the bonding of the Z-DOL, Z-TETRAOL, and A20H end groups to the silicon substrate is shown in Fig. 35.26.

35.2.1 Nanotribological Studies

Adhesive force and coefficient of friction measurements were made using an AFM and the results are presented in Fig. 35.27 [35.44]. The lubricant-coated samples reduce adhesion compared to the uncoated silicon. The adhesive forces measured on the partially bonded lubricant films are higher than the data from their fully bonded counterparts. As discussed earlier, the mobile fraction on the surface of the partially bonded sample facilitates the formation of a meniscus, which increases the tip-surface adhesion. All of the lubricant-coated samples exhibit a reduction in the coefficient of friction relative to the uncoated silicon. The partially bonded samples have a lower coefficient of friction compared to the fully bonded samples, implying that the mobile lubricant molecules in the former facilitate sliding of the tip on the surface. This will be analyzed in more detail in the following subsection in the context of wear, which was monitored using surface potential and resistance measurements.

Palacio and *Bhushan* [35.45] studied the effect of relative humidity and temperature on the friction and adhesion of novel PFPE films deposited on amorphous carbon-coated (diamond-like carbon) magnetic tapes. The trends were similar to those as reported earlier in Sect. 35.1.2, *Relative Humidity and Temperature Effect*.

35.2.2 Wear Detection by Surface Potential Measurements

Wear experiments were performed on the lubricated surfaces using a diamond tip over a $5 \times 5\,\mu m^2$ region at a load of 10 μN for 20 cycles. Figure 35.28a shows the surface height and surface potential images for the fully bonded and partially bonded lubricants [35.44]. The corresponding images for the uncoated silicon substrate are shown for comparison. It can be seen from the images that debris is generated around the wear region for all coatings tested. More wear debris is observed in

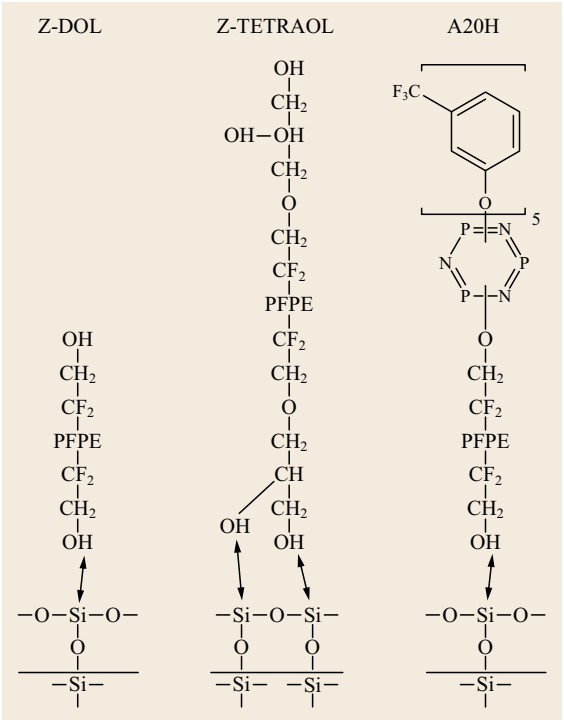


Fig. 35.26 Schematics of Z-DOL, Z-TETRAOL, and A20H molecules bonded on Si substrates

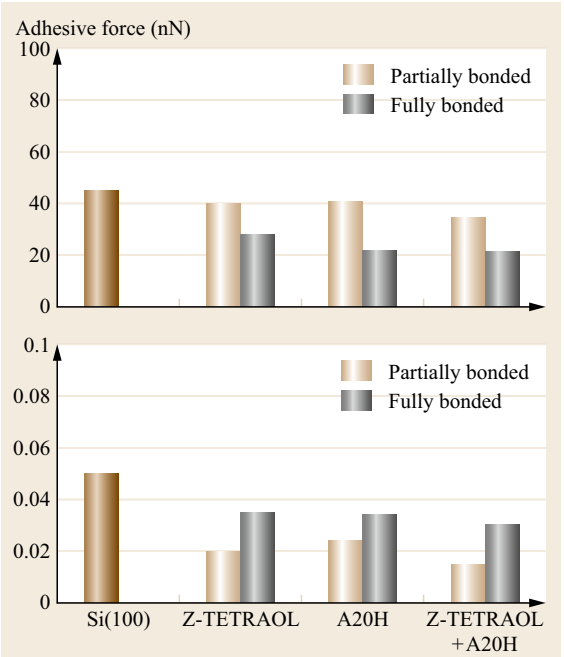


Fig. 35.27 Summary of the adhesive force and coefficient of friction of Si(100), Z-TETRAOL, A20H and Z-TETRAOL+A20H films at room temperature (22 °C) and in ambient air (45–55% RH) (after [35.44])

the fully bonded samples than in the partially bonded samples. Samples with the fully bonded lubricant have limited wear protection compared to those with partially bonded lubricants because the former has only immobile molecules, while the latter has both immobile and mobile molecules. The mobile fraction of the lubricant can replenish the surface, i. e., it can move to the worn area and protect it after the immobile lubricant molecules have been displaced from the surface following repeated rubbing. Since the partially bonded films are thicker than the fully bonded films, additional wear protection is attributed to a thickness effect.

Figure 35.28b presents a bar chart showing surface potential changes for various samples. The surface potential of the area subjected to the wear test increased, which is more prominent in the fully bonded samples. The partially bonded samples exhibited a smaller change in surface potential, indicating less wear. These findings should be correlated with the result of the wear test on the uncoated silicon sample, where the increase in surface potential is well understood. The Kelvin probe method measures the surface potential difference between the tip and sample, which pertains to differences in the work functions between these two materials. For conducting and semiconducting materials, the mechanism is as follows. The surface potential is altered during physical wear because the Fermi energy level is altered. This is the energy required to

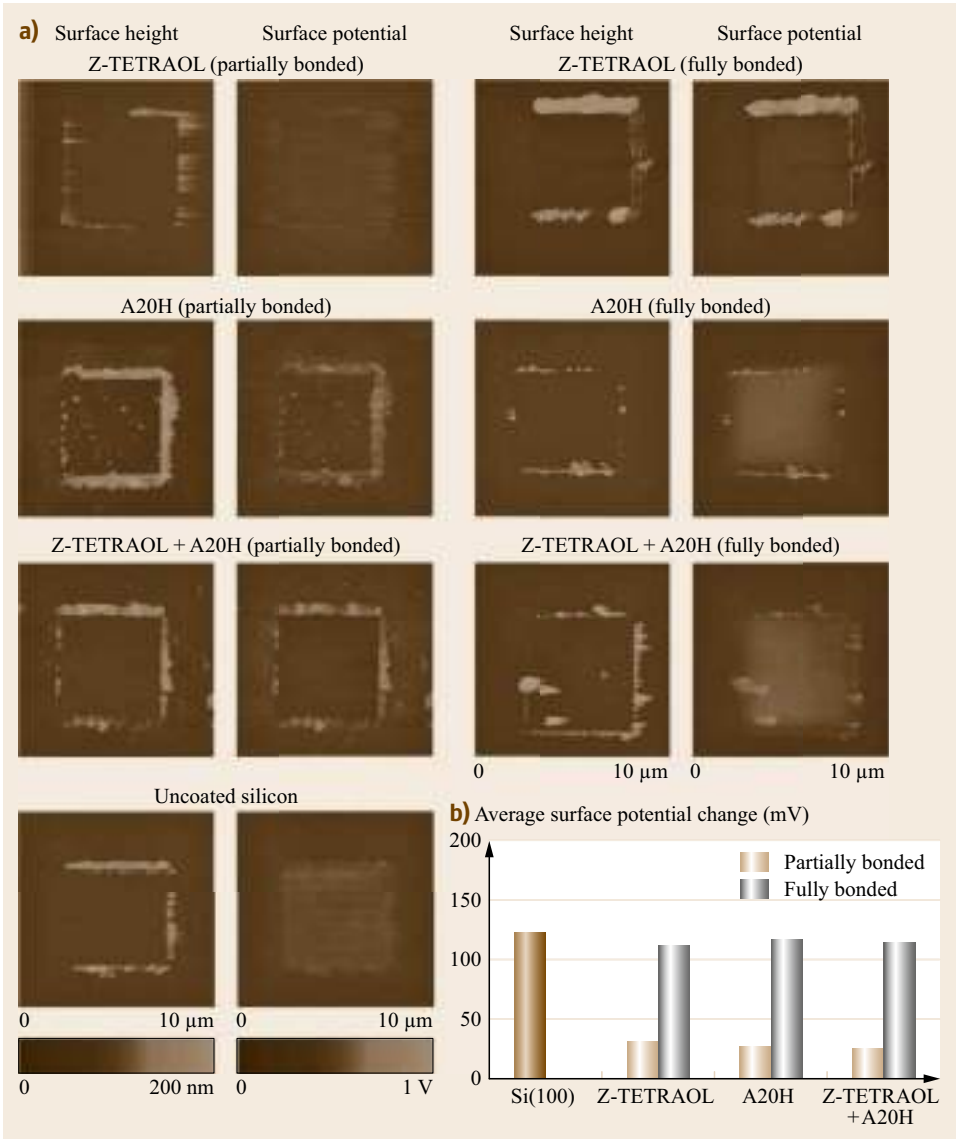


Fig. 35.28 (a) Surface height and surface potential maps of the coatings under investigation after wear testing. Brighter areas correspond to higher values of both the height and surface potential change. (b) Bar chart showing surface potential change (after [35.44])

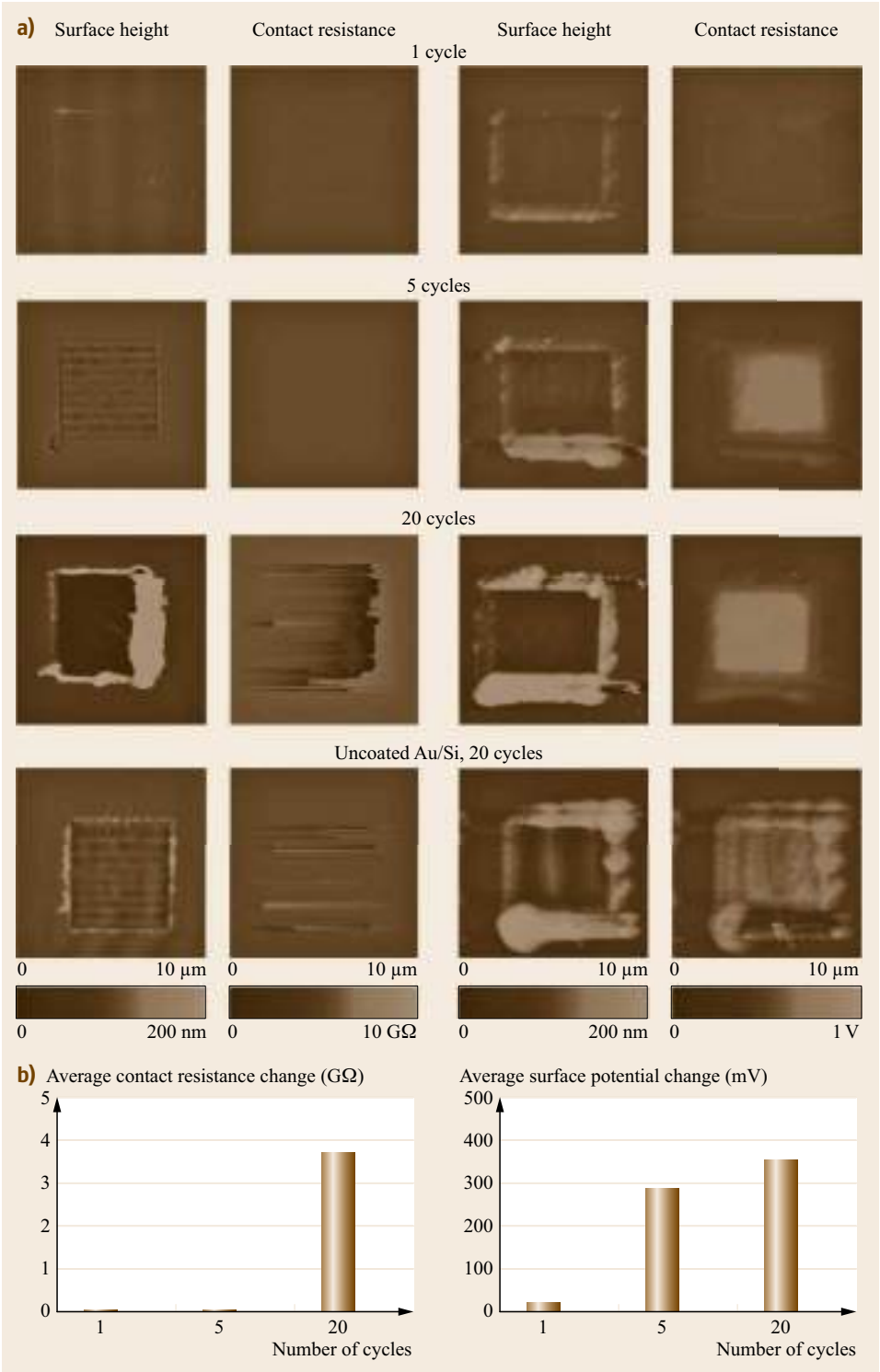


Fig. 35.29 **(a)** Surface height, contact resistance and surface potential maps of fully bonded Z-TETRAOL coating on Au as a function of the number of wear cycles, and **(b)** bar chart showing contact resistance and surface potential change (after [35.44])

remove an electron to a point just outside the material surface. Thermodynamic equilibrium is disrupted with a change in the Fermi level, and can only be restored by the flow of electrons either into or from the area subjected to wear. This mechanism does not apply to materials such as SiO_2 (naturally present as a thin layer) and the lubricants, which are both insulators. Physical wear on these materials would not cause a change in the surface potential because charge dissipation is poor. Therefore, a considerable surface potential change would be observed only when:

1. The lubricant has been fully removed from the substrate.
2. The native SiO_2 layer has been abraded from the surface.
3. Wear has caused subsurface structural changes.

However, for insulators such as the lubricant, electrostatic charges are introduced as they come into contact with a material with a dissimilar electron affinity (the diamond tip) during the wear test. The charges on the insulating lubricant surface have low mobility and would eventually dissipate into the ambient environment. In the presence of debris, the electrostatic charges become localized and may get trapped (as debris particles are mostly isolated), causing an increase in the surface potential of the debris that formed around the worn area.

35.2.3 Wear Detection by Electrical Resistance Measurements of Z-TETRAOL and the Effect of Cycling

For electrical resistance measurements, lubricant films were applied on evaporated Au film deposited on a Si substrate. The silicon wafer used in the experiments is lightly doped, and it was coated with Au so that a metal-metal contact would be attained once the lubricant was removed. PFPE is insulating. Figure 35.29 presents surface height and resistance images for a fully bonded Z-TETRAOL coating on Au by using a scanning spreading resistance microscopy (SSRM) attachment in AFM. For comparison, another set of tests was conducted on the same sample, and the postwear surface potential images were obtained as shown in the same figure. The Au surface is expected to have a low resistance (corresponding to darker areas in the resistance map), while the lubricant should have high resistance (lighter areas) as it is an insulator. The surface height images show a small amount of wear after one and five cycles. In both cases, the resistance images are featureless, which implies that the lubricant is still present, and the Au is not yet exposed. A surface potential increase is observed, which is attributed to electrostatic charge buildup, but this does not imply full lubricant removal since the resistance of the test area has remained unchanged to this point. After 20 cycles, the contact resistance in the tested area has suddenly decreased, which indicates exposure of the underlying Au. The wear debris has the same resistance compared to

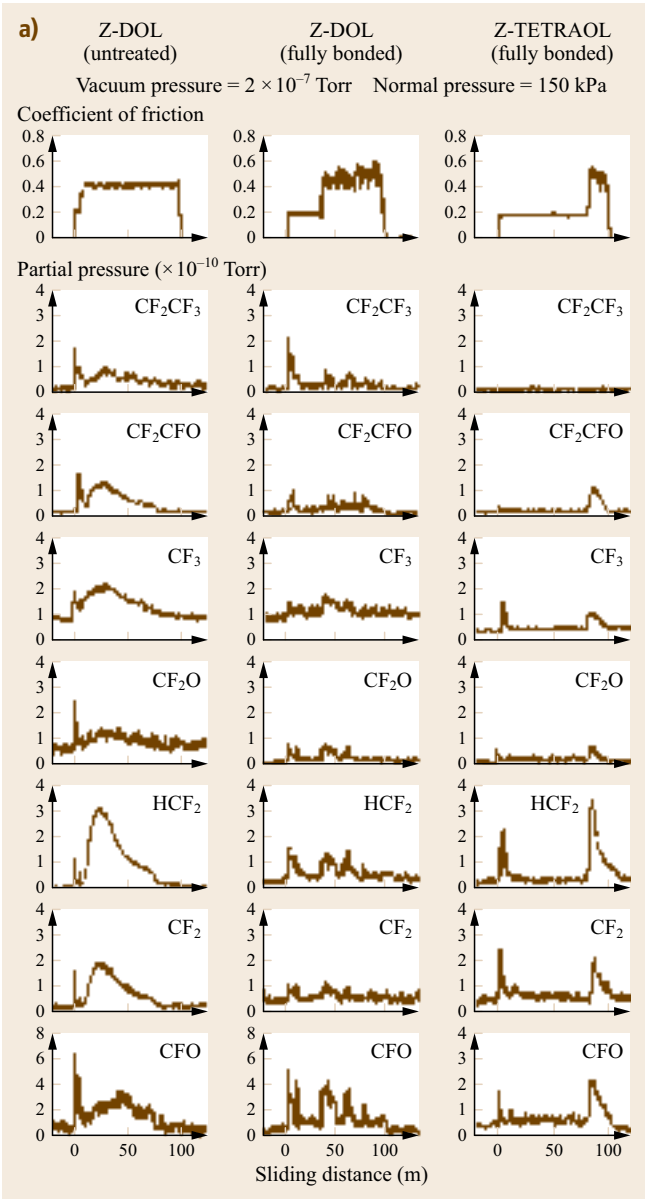


Fig. 35.30 Coefficients of friction and mass spectra data on **(a)** untreated (2.2 nm) and fully bonded (2.3 nm) Z-DOL, fully bonded Z-TETRAOL (2.6 nm), **(b)** fully bonded (0.9 nm) and partially bonded (2.3 nm) A20H, and Z-DOL+A20H (2.4 nm) in high vacuum and under 150 kPa normal pressure, **(c)** partially bonded A20H in high vacuum and under 200 kPa normal pressure (after [35.63])

the unworn lubricant, indicating that the current flow is influenced by the inherent difference in conductivity between the probe tip and the material it is in contact with.

During the scan of the area subjected to the wear test, the current measured by the SSRM sensor corresponds to the contact resistance that exists between the metal-coated tip and the sample. The spreading resistance of the electric current flowing within the semiconductor sample is not measured because this is only present when there is direct contact between the metal-coated tip and a highly doped semiconductor sample.

Because the lubricant is very soft, some signal instability could occur. This comes from possible tip contamination by the lubricant as well as plowing of wear debris during the scan. Since this AFM-based resistance measurement is a contact technique, contamination and plowing are more likely to occur compared to the surface potential measurement, which is a tapping technique. However, this technique is of interest because it provides complementary information to the Kelvin probe method in measuring the extent of wear of conducting films and lubricants, which are materials that can potentially be used in MEMS/NEMS devices.

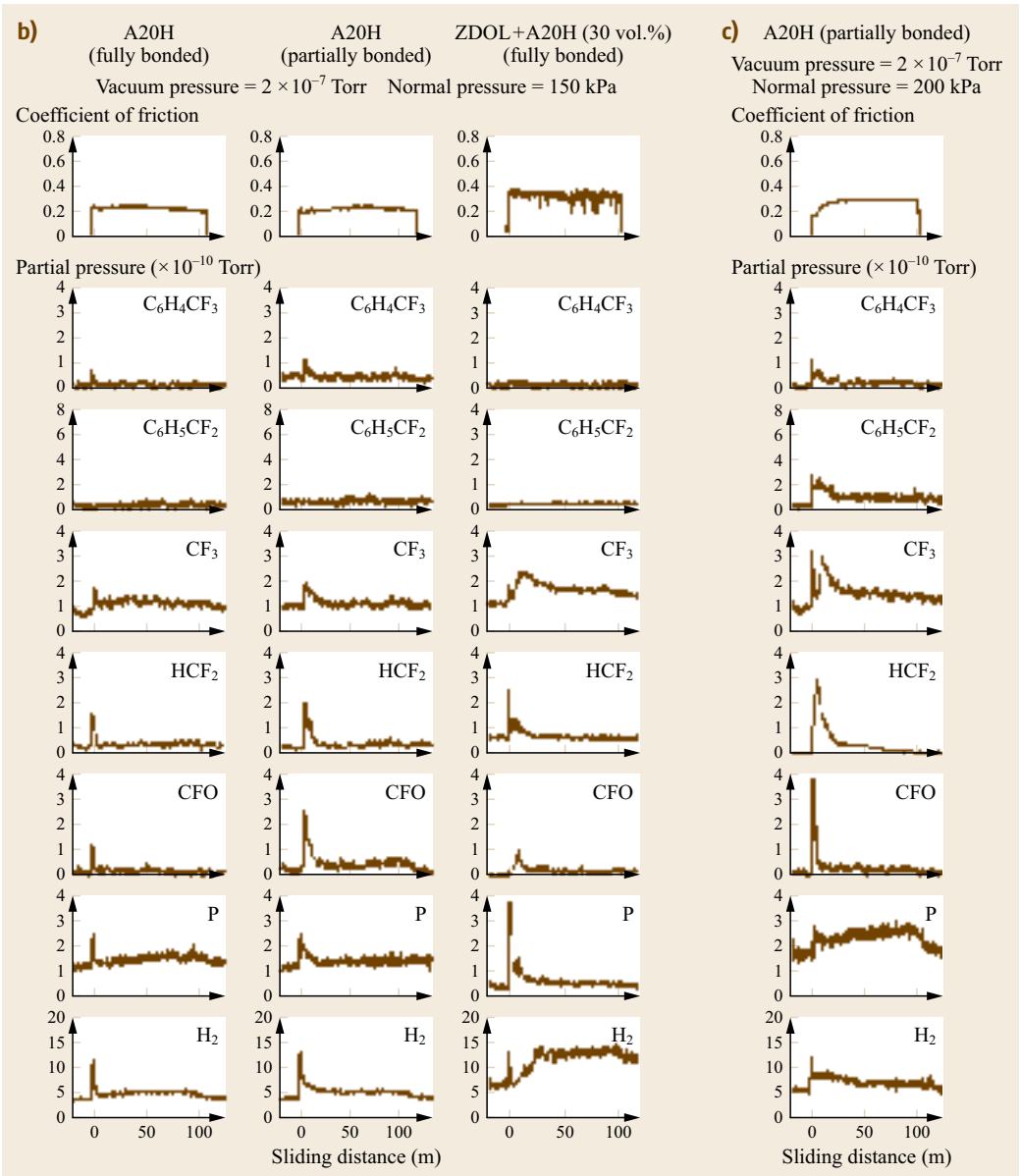


Fig. 35.30 (continued)

35.2.4 Chemical Degradation and Environmental Studies

Tao and *Bhushan* [35.63] carried out chemical degradation studies in a high-vacuum tribotest apparatus [35.40, 66]. In this apparatus, a lubricated wafer mounted on a flexible cantilever beam was slid against an uncoated Si(100) wafer in a macroscale configuration. The system was equipped with a mass spectrometer so that gaseous emissions from the interface could be monitored in situ during the sliding in high vacuum and other controlled environments. The normal load and friction force at the contacting interface were measured using resistive-type strain gage transducers. For the sliding tests, the lubricated Si(100) sample was glued onto a flat surface at the end of a rotating shaft. The sample was slid against a Si(100) wafer mounted on the flat surface of a slider integrated with a flexible cantilever used in magnetic rigid disk drives. The sliding speed was 0.3 m/s and applied pressure was 150 kPa. The environmental effects were investigated in high vacuum (2×10^{-7} Torr), argon, dry air (less than 2% RH), ambient air (30% RH) and high-humidity air (70% RH).

Chemical Degradation Studies

The coefficient of friction and partial pressure of the gaseous products for fully bonded Z-DOL, Z-TETRAOL, A20H, Z-DOL+A20H (30 vol.%), and partially bonded A20H during sliding in high vacuum are shown in Fig. 35.30a–c. The result for untreated Z-DOL is also shown in the figure as a comparison.

In the tests, the sharp increase of friction indicates the failure of the lubricant film. Therefore, the durability of the lubricant film can be obtained from the friction curve. Under the normal pressure of 150 kPa, the untreated Z-DOL failed immediately after sliding. The fully bonded Z-DOL began to fail after sliding of about 40 m. The fully bonded Z-TETRAOL, began to fail at about 80 m. Fully bonded and partially bonded A20H, however, did not fail during the 100 m sliding. The fully bonded Z-DOL+A20H (30 vol.%) failed immediately after sliding. The results show that fully bonded Z-TETRAOL and A20H are more durable than fully bonded Z-DOL, while the untreated Z-DOL is less durable than the bonded films. The PFPE lubricants can be bonded to the applied surface through the hydroxyl group. Z-TETRAOL, with two hydroxyl groups at each end of the molecular chain, can be bonded more tightly on the silicon surface than Z-DOL. The high durability of A20H, however, is beyond expectation. To further confirm the result, the normal pressure of 200 kPa was applied on partially bonded A20H. At the normal pressure of 200 kPa, the coefficient of friction was found to increase. At the same time, gaseous products were detected (Fig. 35.30c). The final coefficient of friction was about 0.3, which is lower than the values (about 0.4) of the other films after failure. This could indicate that the A20H film was only partially worn. The reason for lower durability of Z-DOL+A20H (30 vol.%) in high vacuum is not clear.

In the tests on Z-DOL and Z-TETRAOL films, CF_2CF_3, CF_2CFO, CF_3, CF_2O, HCF_2, CF_2, and CFO were found to increase when friction increased. In the test of A20H films, $C_6H_4CF_3$, $C_6H_5CF_2$, CF_3, HCF_2, CFO, P, and H_2 were detected. For Z-DOL+A20H (30 vol.%), $C_6H_4CF_3$, $C_6H_5CF_2$, CF_3, HCF_2, CFO, P, and H_2 were also detected during sliding. Degradation mechanisms for Z-DOL and various model lubricants have been studied by *Zhao* and *Bhushan* [35.67–69]. Based on their work, Z-DOL starts to decompose above 350 °C. In the sliding conditions used here, it is not likely that the frictional heat can generate such a high temperature. The possibility of catalytic degradation is very low because of the small contact area. Triboelectrical reaction and mechanical scission are considered to be the dominant mechanisms during the sliding of Si on PFPE films. Electron emission is known to

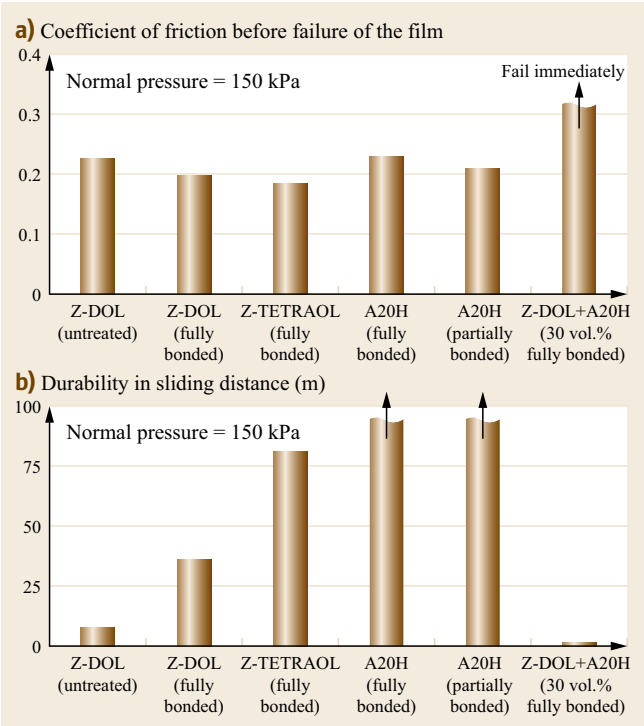


Fig. 35.31 Coefficient of friction and durability comparison of Z-DOL, Z-TETRAOL, A20H, and Z-DOL+A20H in high vacuum (after [35.63])

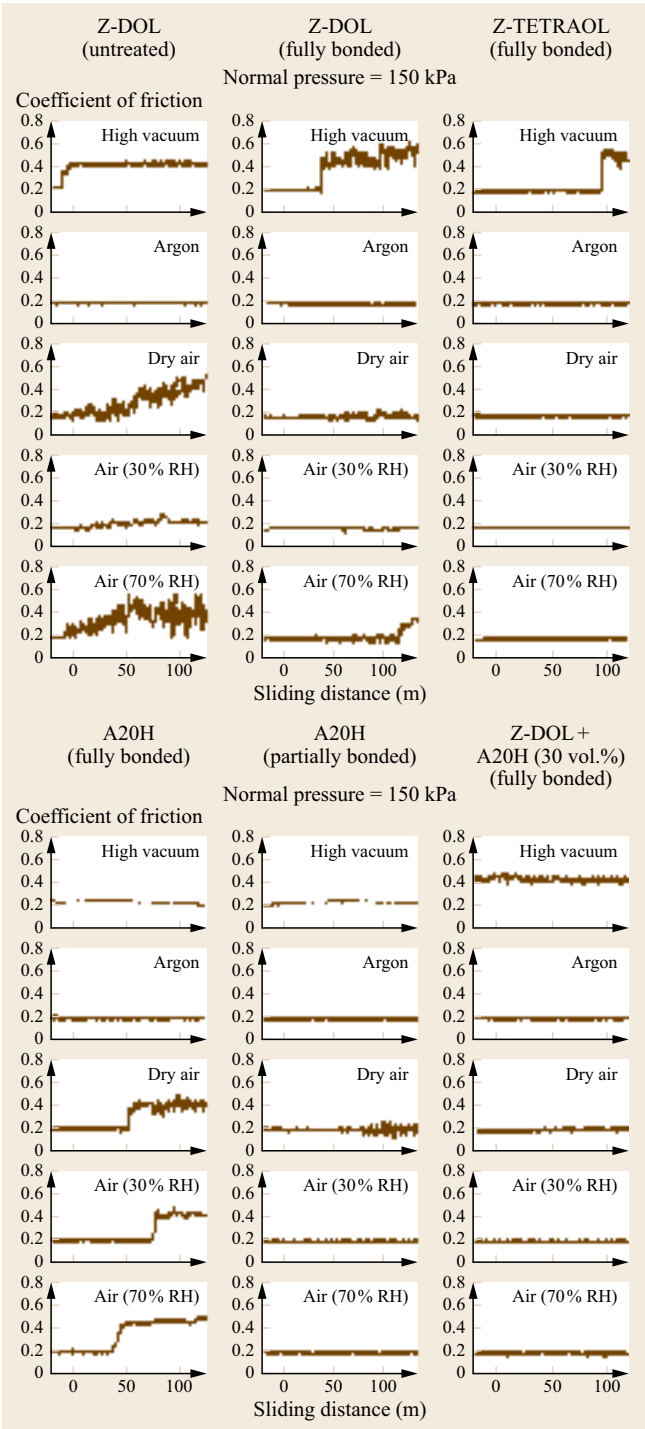


Fig. 35.32a,b Coefficient of friction data in high vacuum, argon, and air with different humidity levels of untreated (2.2 nm) and fully bonded (2.3 nm) Z-DOL, fully bonded Z-TETRAOL (2.6 nm), fully bonded (0.9 nm) and partially bonded (2.3 nm) A20H, and Z-DOL+A20H (2.4 nm) (after [35.63])

occur during sliding for both metal and nonmetal surfaces. The interaction of electrons with PFPE molecules could create various radicals such as $^{\bullet}CF_2{-}O{-}CF_2{-}$, $^{\bullet}CF_2{-}CF_2{-}O{-}CF_2{-}$, and/or $^{\bullet}CF_2{-}CF_3$. These radicals can decompose or react with each other and/or the remaining PFPE molecules. A detailed description of triboelectrical reactions can be found in *Zhao* et al. [35.69] and *Zhao* and *Bhushan* [35.68].

Mechanical scission is another mechanism capable of causing the degradation of PFPE films. The PFPE molecule has a long linear chain structure. The C−C and C−O bonds in the molecular chain are easily subjected to cleaving by the microasperities on the rubbing surfaces, which results in decomposition of the lubricant in the sliding process.

A summary of the coefficients of friction and durability for tested lubricants in high vacuum is shown in Fig. 35.31. The coefficients of friction of fully bonded Z-DOL and Z-TETRAOL are lower than untreated Z-DOL and A20H in high vacuum. Z-TETRAOL shows higher durability than Z-DOL, while A20H (both fully bonded and partially bonded) even shows high durability in high vacuum.

Environmental Studies

The wear tests were conducted in high vacuum, argon, dry air (less than 2% RH), air with 30% RH, and air with 70% RH [35.63]. The applied normal pressure was 150 kPa for all tested films, which was the same as in the degradation tests. Test results are presented in Fig. 35.32. The coefficients of friction show that friction and durability vary with the environment. Z-DOL films fail more rapidly in high vacuum than in other environments. As described in the previous section, there are very few foreign molecules on the contacting surfaces in high vacuum. This enables an intimate contact between the lubricant film and the counterpart surface. A tendency towards chemical bonding occurs between the lubricant film and the counterpart surface. In argon, the Z-DOL films exhibited lower friction and higher durability than in dry air. The chemical effects of oxygen can influence the friction and durability of the lubricant films [35.66]. The water molecules at a moderate humidity level (ambient air) can act as a lubricant between the contacting surfaces. However, in a high humidity environment, the water molecules can penetrate the Z-DOL film and result in a nonuniform distribution of the Z-DOL molecules. The fully bonded Z-DOL film is less influenced by environment than untreated Z-DOL. The reason is that the untreated Z-DOL has more free hydroxyl end groups. The water molecules in the environment can interact with the hydroxyl group of Z-DOL via hydrogen bonds.

Z-TETRAOL is more durable than Z-DOL. In particular, a high-humidity environment does not seem to have negative effects on Z-TETRAOL, although it has one more hydroxyl end group than Z-DOL and may attract more water molecules in high humidity. This could be because the two-hydroxyl end group provides better attachment than Z-DOL on the surface. The tight bonding reduces the probability of nonuniform distribution of the film. The coefficient of friction of fully bonded Z-TETRAOL is very close to fully bonded Z-DOL in argon and air with various humidity levels. In high vacuum, Z-TETRAOL shows a lower coefficient of friction than Z-DOL.

Both fully bonded and partially bonded A20H were very durable in high vacuum. In argon, fully bonded and partially bonded A20H did not fail during the 100 m sliding. In air with various humidity levels, however, the fully bonded A20H failed. The durability is lower compared to Z-DOL. The low durability can be from the low thickness of the film, which is only 0.9 nm. For the 2.3 nm thick partially bonded A20H film, the durability in air with various humidity levels is apparently improved. In particular, at high-humidity levels, partially bonded A20H exhibits high durability. A20H consists of a phosphazene group that is large in size and protects the surface. Phosphazene lubricant has been used as an additive and is known to have high durability in a high-humidity environment. The coefficient of friction of A20H, however, is higher (in the range of 5–15%) than that of Z-DOL in various environments.

Z-DOL+A20H (30 vol.%) shows low durability in high vacuum. In argon and air with various humidity levels, however, the film did not fail during the 100 m sliding. The durability of the mixture is higher than fully bonded Z-DOL. The coefficient of friction of the mixture is higher than that of Z-DOL while lower than that of A20H.

A summary of the coefficients of friction before the failure of the lubricant films in various environments is presented in Fig. 35.33. To summarize the highlights, Z-TETRAOL exhibits higher durability than Z-DOL. A20H exhibits high durability in high vacuum, in argon, and in air with various humidity levels. The mixture of Z-DOL+A20H (30 vol.%) shows low durability in high

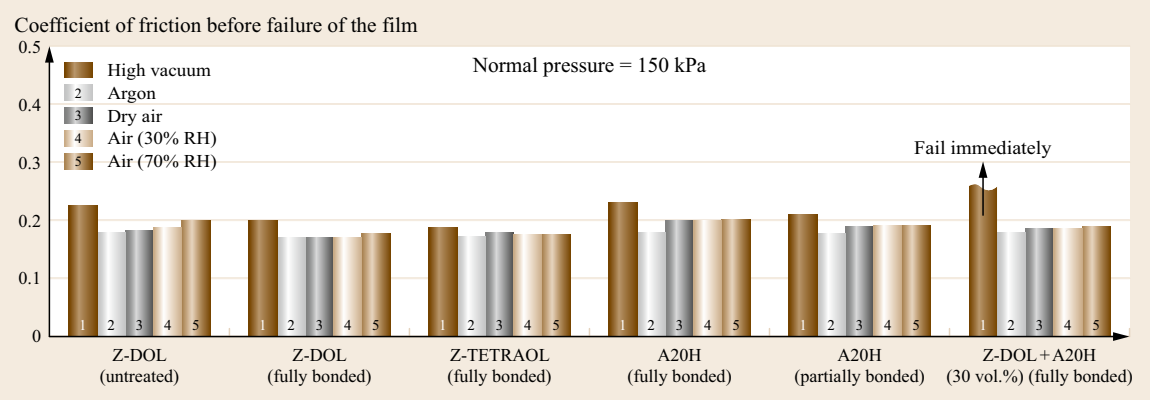


Fig. 35.33a–c Comparison of coefficient of friction data for Z-DOL, Z-TETRAOL, A20H, and Z-DOL+A20H in high vacuum, argon, and air with different humidity levels (after [35.63])

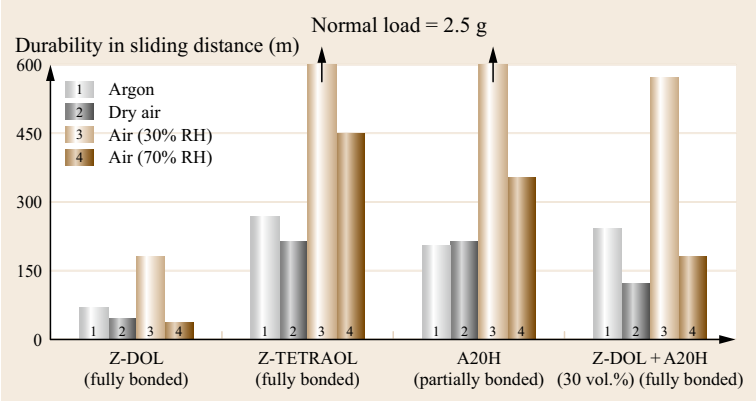


Fig. 35.34 Durability comparison for Z-DOL, Z-TETRAOL, A20H, and Z-DOL+A20H in argon and air with different humidity levels. A single crystal Si(100) ball (1 mm in diameter) was used as a slider on the films (after [35.63])

vacuum while it shows high durability in argon and air with various humidity levels.

In order to further investigate the durability of the lubricant films, a single-crystal Si(100) ball (1 mm in diameter, 5×10^{-7} atoms/cm^3 boron doped) was used as a slider on the films to accelerate wear. Tests were performed in argon, dry air, ambient air, and high-humidity air. The applied load on the silicon ball was 2.5 g, and the sliding speed was 0.3 m/s. The sliding distance was up to 600 m. The results are shown in Fig. 35.34. From Fig. 35.34, in all environments, Z-TETRAOL exhibits higher durability than the other films. The durability of A20H is slightly lower than that of Z-TETRAOL except in ambient air, where both are comparable. Tests on Z-TETRAOL and A20H were terminated at 600 m. Z-DOL+A20H shows higher durability than Z-DOL.

35.3 Nanotribological and Electrical Studies of Ionic Liquid Films

An ionic liquid (IL) is a synthetic salt with a melting point below 100 °C. A room-temperature ionic liquid is a synthetic molten salt with a melting point at or below room temperature. One or both of the ions are an organic species. At least one ion has a delocalized charge such that the formation of a stable crystal lattice is prevented, and the ions are held together by strong electrostatic forces. As a result of the poor coordination of the ions, these compounds are liquid below 100 °C or even at room temperature [35.20, 62].

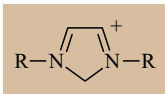
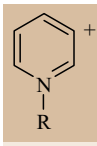
The number of combinations of anions and cations that can be used to produce ionic liquids is in the range of one million. Typical cations include imidazolium, pyridinium, ammonium, phosphonium and sulfonium as shown below, where R stands for an organic group (Table 35.3).

Typical anions are tetrafluoroborate, BF_4^-, hexafluorophosphate, PF_6^-, bis(trifluorosulfonyl) imide, $(CF_3SO_2)_2N$ (*triflamide*) and toluene-4-sulfonate, $C_7H_7O_3S$ (*tosylate*) [35.70]. Dependent upon the substrate wettability and other functional requirements, a set of cations and anions can be combined. The ionic liquids were initially developed for use as electrolytes in batteries and for electrodeposition. Recent applications have geared these compounds as environmentally friendly solvents for chemical synthesis (*green chemistry*) where these liquids are used as substitutes for conventional organic solvents.

Ionic liquids are considered as potential lubricants. Their strong electrostatic bonding, compared to covalently bonded fluids, leads to very desirable lubrication properties. They also possess desirable properties such as negligible volatility, nonflammability, high thermal stability or high decomposition temperature, efficient heat transfer properties, low melting point, as well as compatibility with lubricant additives. Unlike conventional lubricants that are electrically insulating, ionic liquids can minimize the contact resistance between sliding surfaces because they are conducting, and conducting lubricants are needed for various electrical applications [35.22–24]. These liquids can also be used to mitigate arcing, which is a cause of electrical breakdown in sliding electrical contacts. In addition, ILs have high thermal conductivity, which helps to dissipate heat during sliding. The use of ionic liquids instead of hydrocarbon base oils (such as highly reformed mineral oils) has the potential to dramatically reduce air emissions. Perfluoropolyethers (PFPEs) are used in magnetic rigid disk and vacuum grease applications due to their high thermal stability and extremely low vapor pressure. However, from the commercial standpoint, ionic liquids are cheaper than PFPEs by a factor of two or so, providing the motivation for comparing the tribological properties of the former with the latter. ILs are being considered for MEMS/NEMS applications because of their high temperature stability, electrical conductivity, and desirable lubrication properties.

Bhushan et al. [35.62] evaluated ionic liquids with the hexafluorophosphate anion deposited on Si(100) wafers, and these were found to exhibit improved friction and wear properties compared to conventional lubricants. The films were evaluated as untreated, partially bonded (by heating at 150 °C for 30 min after dip coating), and fully bonded (thermal treated and washed). The ionic liquid containing the octyl sulfate anion has also been developed and is of interest due to

Table 35.3 The number of combinations of anions and cations to produce ionic liquids

		NR_4^+	PR_4^+	SR_3^+
Imidazalium	Pyridinium	Ammonium	Phosphonium	Sulfonium

its resistance to hydrolysis. Based on experience, anions are observed to affect tribological performance. Table 35.4 lists the physical and thermal properties of a selected ionic liquid and its properties are compared to the PFPE lubricant Z-TETRAOL.

The durability of ionic liquid films on various metal and ceramic substrates has been investigated from the standpoint of film formation (wettability) and film removal (friction and wear), where it was found that certain cations and anions exhibit better wetting, friction reduction, and wear resistance properties [35.62]. In general, ionic liquids exhibit better wettability on noble metal and ceramic surfaces than on nonnoble metal surfaces [35.70]. The flat imidazolium cation shows poorer wettability compared to bulkier cations such as ammonium and sulfonium. Among salts with the imidazolium cation, the presence of longer organic side chains lead to a reduction of the coefficient of friction. An anion effect is also observed, where oxygen-rich anions show better substrate wettability and lower wear compared to other imidazolium salts. Based on these findings, the ionic liquids of interest are 1-butyl-3-methylimidazolium hexafluorophosphate (BMIM-PF_6) and 1-butyl-3-methylimidazolium octyl sulfate (BMIM-$OctSO_4$). These were studied by *Bhushan* et al. [35.62], and the former was found to be superior in tribological performance. The chemical structures of BMIM-PF_6 and Z-TETRAOL for comparison are shown in Fig. 35.35, and a summary of their properties is presented in Table 35.4.

Some dicationic ILs are thermally stable up to 400 °C [35.75]. The adhesion and friction properties of two dicationic IL films on Si(100) substrate, based on the imidazolium cation and the bis(trifluoromethanesulfonyl)imide (or *triflamide*) anion, were studied by *Palacio* and *Bhushan* [35.21] and compared to the monocationic IL 1-butyl-3-methyl-1 *H*-imidazolium hexafluorophosphate (BMIM-PF_6). AFM experiments were also performed at various humidity and temperature conditions in order to investigate the effect of the environment on the nanolubrication properties of these ILs. Microscale friction and wear experiments using the ball-on-flat tribometer and wear at ultralow loads using an AFM were carried out. Fourier transform

Table 35.4 Physical, thermal and electrical properties of BMIM-PF_6 and Z-TETRAOL

	1-Butyl-3-methylimidazolium hexafluorophosphate (BMIM-PF_6)	Z-TETRAOL
Cation	$C_8H_{15}N_2^+$	–
Anion	PF_6^-	–
Molecular weight (g/mol)	284[a]	2300[b]
$T_{melting}$ (°C)	10[c]	–
$T_{decomposition}$ (°C)	300[c]	≈ 320[b]
Density (g/cm^3)	1.37[a]	1.75[b]
Kinematic viscosity (mm^2/s)	281[a] (20 °C) 78.7[d] (40 °C)	2000[b] (20 °C)
Pour point (°C)	< −50[e]	−67[b]
Specific heat (J/g K)	1.44[f] (25 °C)	≈ 0.20[b] (50 °C)
Thermal conductivity at 25 °C (W/m K)	0.15[g]	≈ 0.09[b]
Dielectric strength at 25 °C (kV/mm)	–	≈ 30[b]
Volume resistivity (Ω cm)	714[h]	≈ 10^{13}[b]
Vapor pressure at 20 °C (Torr)	< 10^{-9}	5 × 10^{-7}[b]
Wettability on Si	Moderate[c]	–
Water contact angle	46° (untreated) 39° (partially bonded) 41° (fully bonded)	54° (untreated) 83° (partially bonded) 88° (fully bonded)
Miscibility with isopropanol	Total[a]	–
Miscibility with water	None[h]	–

[a] Merck Ionic Liquids Database, Darmstadt, Germany, also: http://ildb.merck.de/ionicliquids/en/startpage.htm
[b] Z-TETRAOL Data Sheet, Solvay Solexis Inc., Thorofare, NJ
[c] *Kinzig* and *Sutor* [35.70]
[d] *Reich* et al. [35.27]
[e] *Wang* et al. [35.71]
[f] *Kabo* et al. [35.72]
[g] *Frez* et al. [35.73]
[h] *Carda-Broch* et al. [35.74].

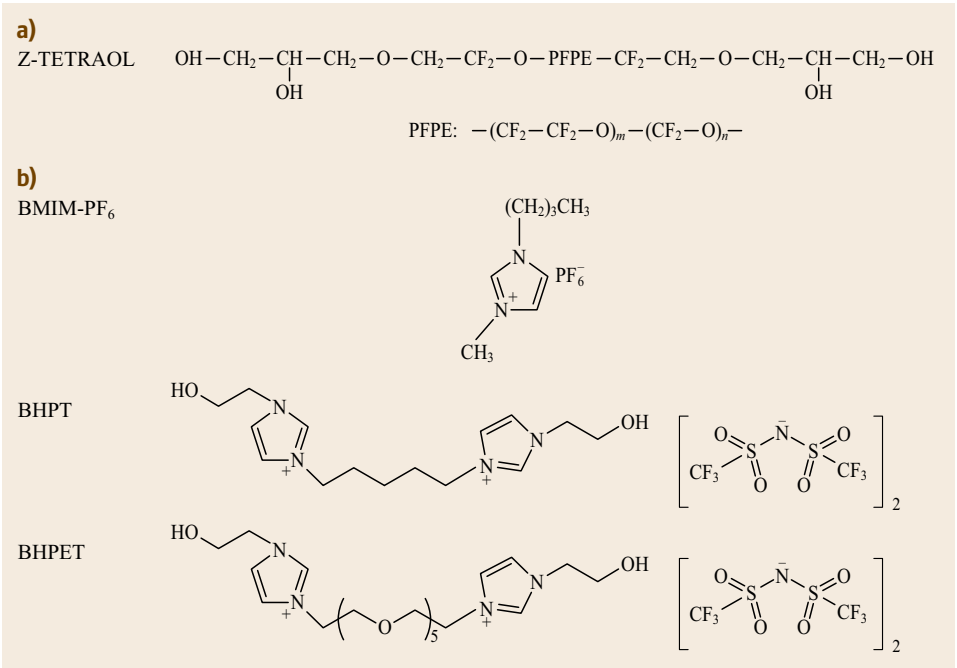


Fig. 35.35 **(a)** Chemical structures of the Z-TETRAOL molecule, and **(b)** chemical structures of the BMIM-PF_6, BHPT and BHPET molecules

infrared (FTIR) spectroscopy and x-ray photoelectron spectroscopy (XPS) were used to determine the chemical species that affect intermolecular bonding and also to elucidate the effect of the environment on the IL film surface in the case of FTIR data (Table 35.4).

The dicationic ionic liquids used in this study were 1,1′-(Pentane-1,5-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethanesulfonyl)imide)] (abbreviated as BHPT) and 1,1′-(3,6,9,12,15-Pentaoxapentadecane-1,15-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethanesulfonyl)imide)] (abbreviated as BHPET) [35.76] deposited on Si(100) wafers. The common name *triflamide* will be used here when referring to the anion. The properties of these dicationic ILs were compared with the conventional monocationic ionic liquid 1-Butyl-3-methyl-1*H*-imidazolium hexafluorophosphate, abbreviated as BMIM-PF_6 (Merck, Germany). Their chemical structures are shown in Fig. 35.35b. These compounds have been applied on single-crystal Si(100) (phosphorus doped) with a native oxide layer on the surface using the dip coating technique. The method and the apparatus used have been described earlier in Sect. 35.1.2. The films were heat-treated at 150 °C for 30 min after dip coating, such that they were partially bonded [35.62].

35.3.1 Monocationic Liquid Films

In this section, nanotribological data on BMIM-PF_6 and Z-TETRAOL are presented [35.20, 24, 45, 62].

Nanotribological Studies

In Fig. 35.36a, the surface height images for the untreated sample (air dried) are compared with the two chemically bonded samples (partially bonded and fully bonded). Aggregates of varying sizes are observed on the untreated Z-TETRAOL and on the ionic liquid coating. These aggregates could have formed initially during preparation of the dilute solution. For Z-TETRAOL, the long PFPE chains can orient in various configurations (such as coils), leading to aggregate formation [35.62]. In ionic liquids containing the 1-butyl-3-methylimidazolium cation, it is believed that aggregate formation in solvents that are less polar compared to water aids in minimizing the charge density (charge delocalization) within the ions. These aggregates are subsequently deposited on the silicon surface during the dip coating procedure. It is observed that the untreated lubricant surface has more prominent aggregates compared to the two chemically bonded samples, implying that the heat treatment promotes bonding to the Si substrate. Without the chemical bonding procedure, the lubricant molecules are less likely to attach to the substrate and would tend to attract each other instead, such that dewetting is more likely. The immobilization of the ionic liquid, which is promoted by thermal treatment, occurs by reaction of the anion with the hydroxyl groups present on the silicon surface, as shown in Fig. 35.36b for BMIM-PF_6.

Figure 35.37a is a summary of the adhesive force and coefficient of friction measurements on the ionic

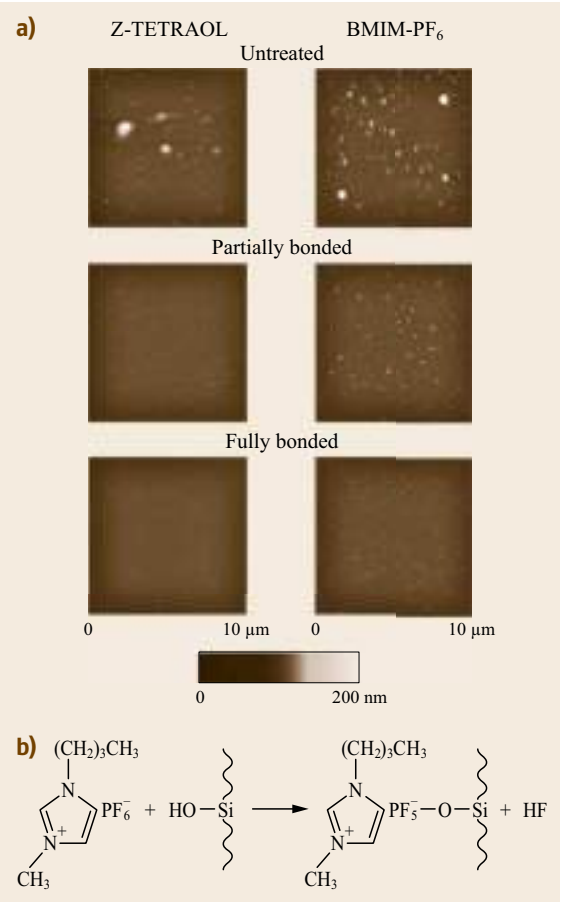


Fig. 35.36 **(a)** Surface height images for the untreated, partially bonded and fully bonded films of BMIM-PF_6 on silicon substrate, and **(b)** schematic for the attachment of BMIM-PF_6 to the silicon substrate (after [35.62])

liquid. Z-TETRAOL and Si(100) data are provided for comparison. The adhesive force has been observed to decrease in the following order: untreated > partially bonded > fully bonded. This mobile fraction on the untreated sample facilitates the formation of a meniscus, which increases the tip-sample adhesion. The adhesive force is highest in the untreated coating since it has the greatest amount of the mobile fraction among the three samples. Conversely, the sample with no mobile lubricant fraction available (fully bonded) has the lowest adhesive force.

A different trend is observed in the coefficient of friction (μ) data. Both the fully bonded and partially bonded samples have lower μ values compared to the uncoated silicon. Friction forces are lower on the latter, implying that the mobile lubricant fraction present in the partially bonded samples facilitates sliding of the tip on the surface. However, μ values for the untreated samples are higher than the data for the heat-treated coatings. Due to the lack of chemical bonding, the interaction of the lubricant to the substrate is weakened and dewetting can occur. Water and lubricant molecules are more likely to form a meniscus as the tip approaches the surface. This provides greater resistance to tip sliding, leading to higher coefficient of friction values. The lower portion of Fig. 35.37a is a schematic illustrating the role of meniscus formation in the adhesive and friction forces obtained for the uncoated Si and the untreated, partially bonded and fully bonded lubricant-coated Si surfaces.

Figure 35.37b contains plots of the coefficient of friction as a function of the number of sliding cycles at 70 nN normal load. Only a small rise in the coefficient of friction was observed for both Z-TETRAOL and the BMIM-PF_6 surfaces, indicating low surface wear. In the case of untreated Z-TETRAOL, a crossover is observed, where the coefficient of friction increases from its initial value and exceeds the μ of silicon after a certain number of cycles. This is attributed to the transfer of lubricant molecules to the AFM tip and the interaction of the transferred molecules with the lubricant still attached on the Si substrate, which will increase the friction force.

In order to compare friction and wear properties on the nanoscale with those on the microscale, conventional ball-on-flat tribometer experiments were conducted on the same samples. Images and profile traces of the wear scars are shown in Fig. 35.38a. The coefficient of friction and number of cycles to failure are summarized in Fig. 35.38b. Ionic liquid shows enhanced durability compared to both the Z-TETRAOL-coated and the uncoated Si. The nanoscale data presented in Fig. 35.37 can be compared to μ values obtained from ball-on-flat tests (Fig. 35.38b). The μ values of the untreated lubricant samples obtained by using AFM are lower than the μ obtained from the ball-on-flat tests. This is attributed to the difference in the length scales of the test techniques. An AFM tip simulates a single asperity contact while the conventional friction test involves the contact of multiple asperities present in the test system [35.1, 2]. With regards to wear, the interface contact of the AFM and ball-on-flat techniques are different from each other. In an AFM, the contact stress is very high, such that material can be displaced more easily. For a ball-on-flat test, the ball exerts a lower pressure on the surface, and the coating is in a confined geometry. As a consequence, displacement of the coating is not as easy as in AFM, leading to enhanced wear resistance.

Wear Detection by Surface Potential and Electrical Resistance Measurements

Figure 35.39a is a summary of the wear tests and the corresponding surface potential measurements on the

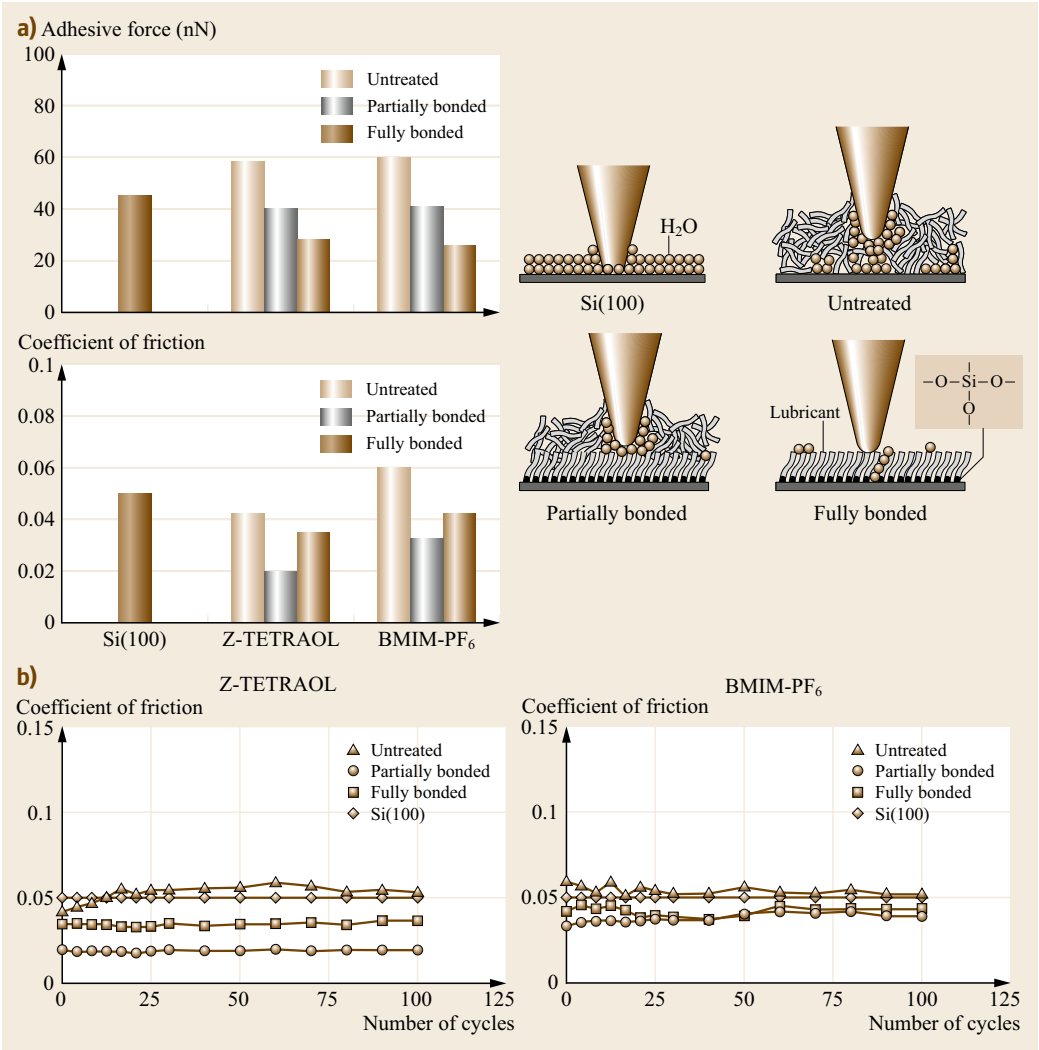


Fig. 35.37 **(a)** Summary of the adhesive force and coefficient of friction, and **(b)** durability data after 100 cycles for BMIM-PF_6 at room temperature (22 °C) and ambient air (45–55% RH). Data for the uncoated Si and Z-TETRAOL are shown for comparison. Schematic in **(a)** shows the effect of chemical bonding treatment and meniscus formation between the AFM tip and sample surface on the adhesive and friction forces (after [35.62])

ionic liquid. A bar chart summarizing the average surface potential change on the tested area is shown in Fig. 35.39b. In all cases, a smaller amount of debris was generated compared to the uncoated silicon surface, indicating that the ionic liquid provides wear protection. In general, the samples containing the mobile lubricant fraction (i. e., untreated and partially bonded surfaces) exhibit a lower surface potential change compared to the fully bonded sample, which only has immobile lubricant molecules. This is attributed to lubricant replenishment by the mobile fractions, which can occur in the untreated and partially bonded samples [35.44, 63]. From the bar chart in Fig. 35.39b, it is also observed that the change in surface potential is generally lower in the ionic liquid coating compared to the Z-TETRAOL coating and the uncoated silicon. This indicates that any built-up surface charges arising from the wear test were immediately dissipated onto the conducting ionic lubricant coating surface. In the case of Z-TETRAOL and the uncoated silicon, the charges remained trapped in the test area, since both these materials are insulators. Based on these findings, a considerable surface potential change will be observed on the wear region when:

1. The lubricant has been fully removed from the substrate.

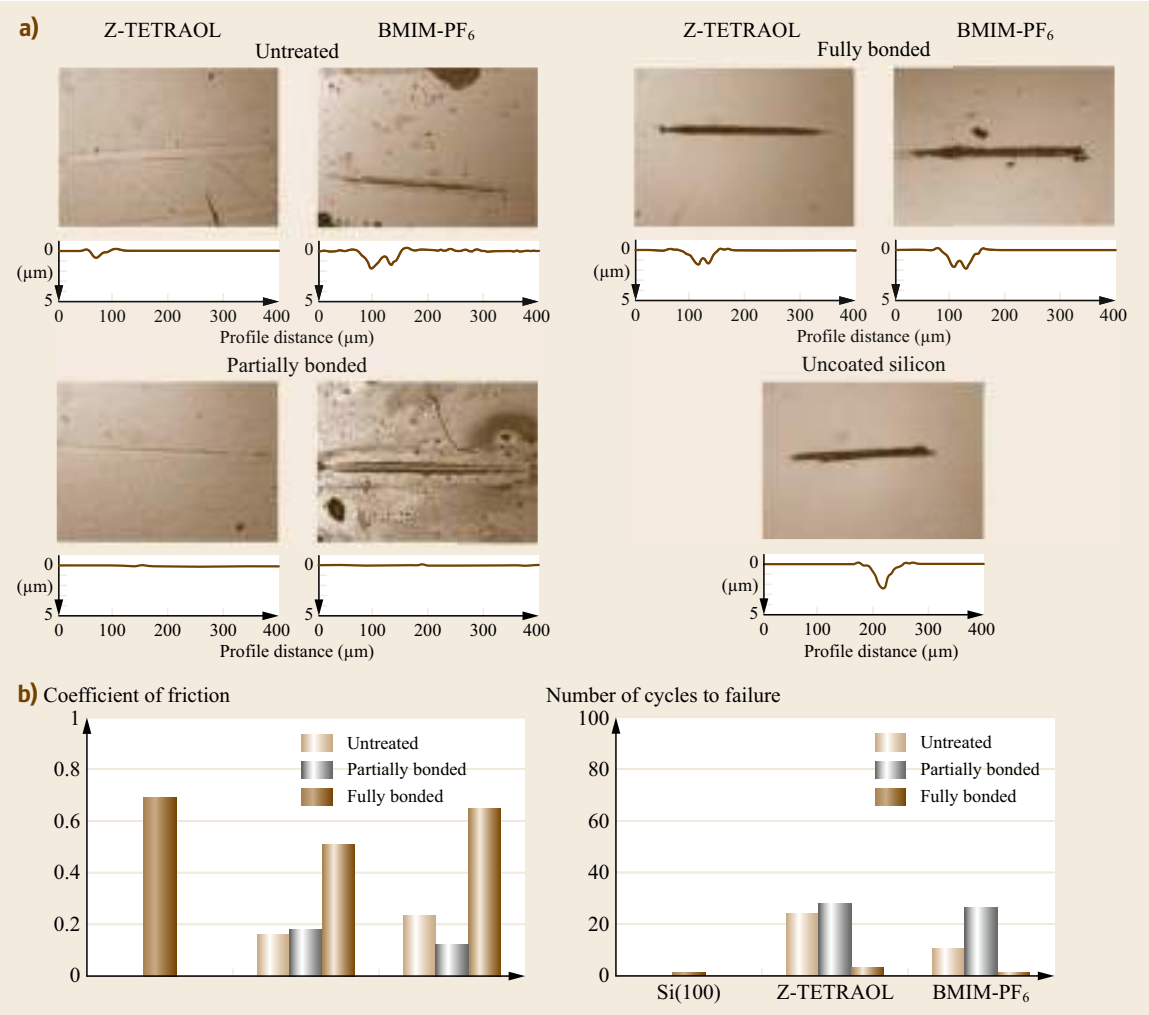


Fig. 35.38 **(a)** Optical images and height profiles taken after 20 cycles, and **(b)** summary of the coefficient of friction and number of cycles to failure from ball-on-flat tests on various BMIM-PF_6 coatings. Data for the uncoated Si and Z-TETRAOL are shown for comparison (after [35.62])

2. The native SiO_2 layer has been abraded from the surface.
3. Wear has caused subsurface structural changes.
4. Charges build up as they are unable to dissipate into the surrounding material.

Contact resistance images for the surfaces subjected to the wear tests are presented in Fig. 35.40a. The average change in the contact resistance of the wear region relative to the untested area is summarized in Fig. 35.40b. The fully bonded BMIM-PF_6 has an appreciable contact resistance increase in the wear region. Since silicon is a semiconductor, it has much higher resistance compared to the surrounding ionic liquid. The resistance increase in the worn area implies that the substrate is exposed after the wear test. Partially bonded films did not get worn out from the substrate after the test, as evidenced by the lack of contact resistance change in the tested area. The untreated Z-TETRAOL exhibited an observable resistance change, while BMIM-PF_6 did not. This can be correlated to the durability data in Fig. 35.37b, where the untreated Z-TETRAOL sample exhibited an increase in the friction force with time due to the transfer of lubricant molecules to the tip. Easier lubricant removal means that the diamond tip (in the case of Fig. 35.40) can cause substrate wear much sooner, leading to the observed resistance increase in the tested area. However, the resistance image does not provide a clear contrast between Z-TETRAOL and the newly exposed

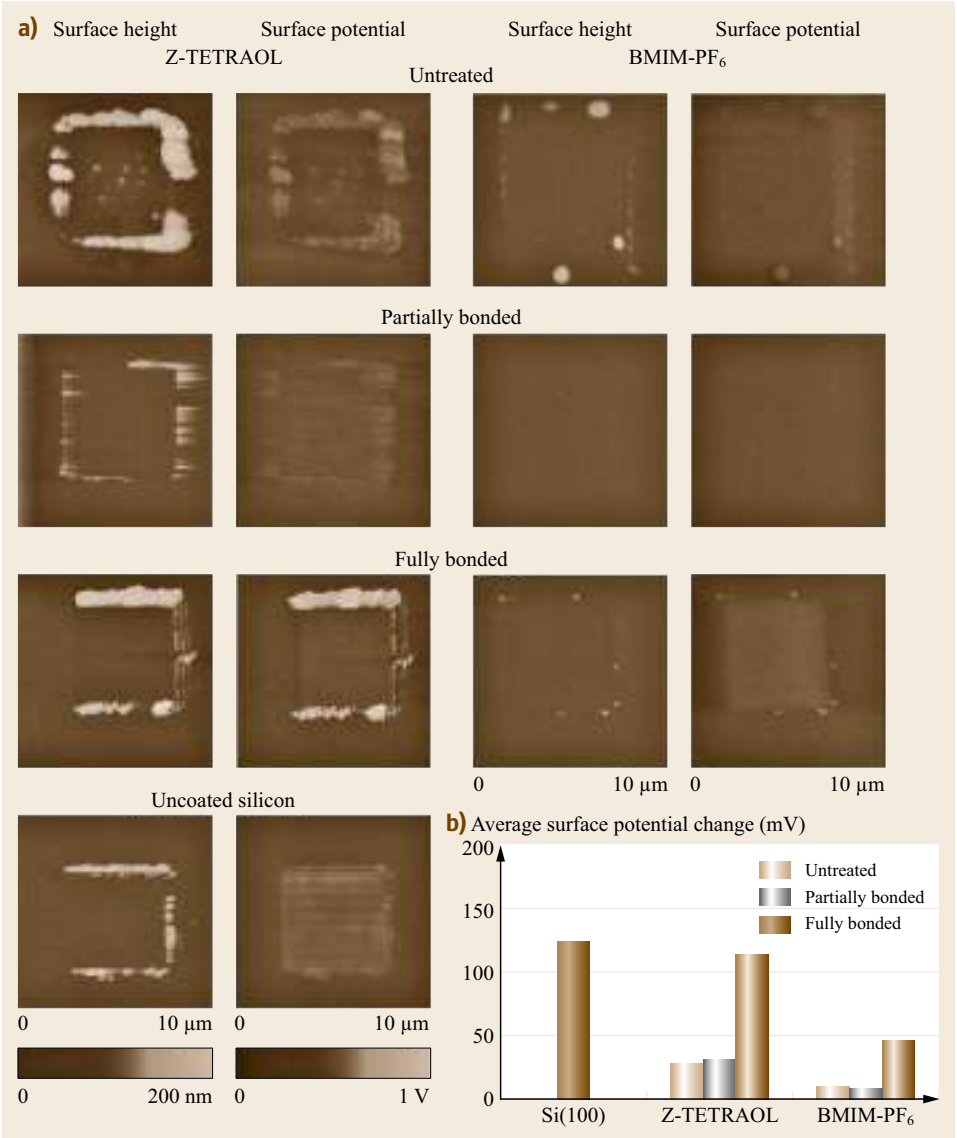


Fig. 35.39 **(a)** Surface height and surface potential maps after wear tests, and **(b)** bar chart showing surface potential change for various $BMIM\text{-}PF_6$ coatings. Data for the uncoated Si and Z-TETRAOL are shown for comparison (after [35.62])

substrate since both materials have high resistance values.

Microscale contact resistance obtained from ball-on-flat tribometer testing is shown in Fig. 35.41, along with the corresponding coefficient of friction data. For the ionic liquid films, the initial resistance is slightly lower than that of uncoated silicon, confirming their conductive nature. For the Z-TETRAOL samples, the contact resistance is about the same magnitude as the uncoated silicon. But for the conducting ionic liquid, an increase in resistance corresponds to an increase in the coefficient of friction, indicating wear of the lubricant and exposure of the silicon substrate, similar to observations in the nanoscale. These results are consistent with the adhesion, friction, and surface potential results with regards to wear detection and wear protection coming from the mobile and immobile lubricant fractions.

The durability data and trends for the ionic liquid obtained by using a steel ball are inferior to the results in Fig. 35.38, which was measured using a sapphire ball. In Fig. 35.41, the partially bonded samples still show the best durability, but in this case, Z-TETRAOL has the highest number of cycles to failure (opposite trend compared to Fig. 35.38b), as indicated by the point where the jump in the coefficient of friction is observed. This can be accounted for from the wetting properties of ionic liquids on different surfaces. It has been observed that ionic liquids have a tendency to wet

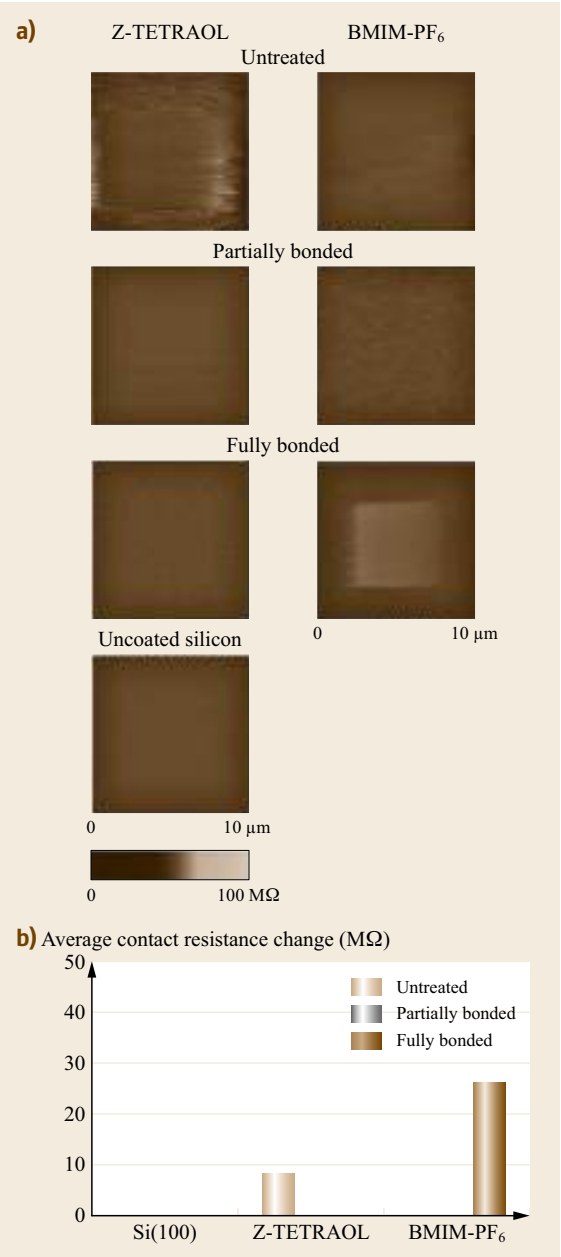


Fig. 35.40 **(a)** Nanoscale contact resistance images after wear tests and **(b)** bar chart showing contact resistance change for various BMIM-PF_6 coatings. Data for the uncoated Si and Z-TETRAOL are shown for comparison. Cases without the clear wear scar did not exhibit measurable change (after [35.62])

nonmetal surfaces (e.g., Si_3N_4, SiO_2, glass) better than conventional metal surfaces (such as 440C, M50 and 52100 steel) [35.62]. For wear tests with a steel ball, less wettability means less lubricant retention at the interface. This material wetting effect is possibly more significant for the ionic liquid than in Z-TETRAOL, but nonetheless, the durability of the partially bonded BMIM-PF_6 is still close to its Z-TETRAOL counterpart, such that ILs are still viable lubricants and comparable to PFPEs.

35.3.2 Dicationic Ionic Liquid Films

In this section, nanotribological data on BHPT and BHPET are presented. Data on BMIM-PF_6 and uncoated Si(100) are also presented for comparison [35.21].

Nanotribological Studies

Figure 35.42a is a summary of the contact angle, adhesive force, and coefficient of friction (μ) measurements for the coated and uncoated samples performed at ambient temperature and humidity conditions (22 °C and 50% RH respectively). The data shown in the bar charts are the averages of three measurements and the error bars represent $\pm 1\sigma$. BHPT is the least hydrophilic as it has the highest contact angle (81°) among the three IL coatings. This coating also exhibited the greatest reduction in the coefficient of friction relative to the uncoated surface. The high contact angle of BHPT leads to minimal meniscus formation between the tip and surface, leading to a large drop in the nanoscale friction force. In addition, BHPT has a pentyl chain that links the two imidazolium cations. This chain can orient the cation molecules on the substrate, thereby facilitating tip sliding on the film surface. In contrast, the polyether chain of BHPET is susceptible to interactions with water molecules, which can promote (instead of minimize) meniscus formation. The BMIM-PF_6 film also exhibited a reduction in the adhesive force and coefficient of friction. This comes from the combination of mobile and immobile lubricant fractions. Immobilization of this ionic liquid is possible as a result of the thermal treatment, which promotes the reaction between the hexafluorophosphate anion and the hydroxyl groups present on the silicon substrate surface [35.21, 62, 77].

Wear tests were conducted by monitoring the change in the friction force on a 2 μm line for 100 cycles. The data shown in Fig. 35.42b are representative of the three measurements made for each sample. The focus of this experiment is the wear of the lubricant film on the substrate. The μ value of the BHPT film changed minimally during the duration of the experiment, which indicates that the film was not being worn after 100 cycles. On the other hand, the BMIM-PF_6 and BHPET samples exhibited a gradual increase in μ, which means that these films could be undergoing some wear and that their interaction with the silicon substrate is weaker compared to that of BHPT.

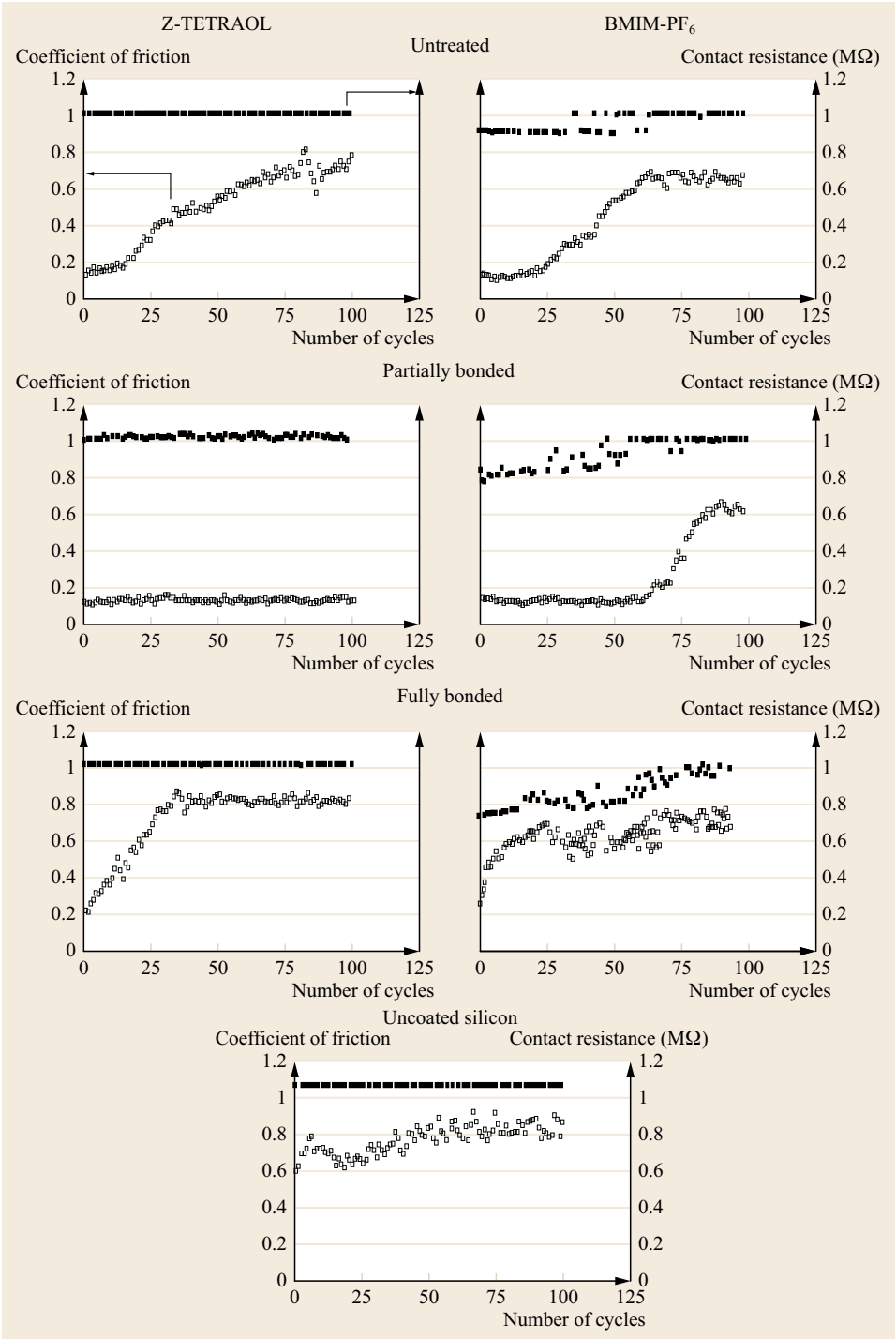


Fig. 35.41 Microscale contact resistance and coefficient of friction after ball-on-flat tests for 100 cycles on various BMIM-PF_6 coatings. Data for the uncoated Si and Z-TETRAOL are shown for comparison (after [35.62])

Figure 35.43 is a summary of the average contact potential change after the wear tests conducted by creating $5 \times 5\,\mu m^2$ wear scars with a diamond tip, where the height and surface potential maps were imaged afterwards. The data shown in the bar chart are an average of three measurements and the error bars represent $\pm 1\sigma$. A change in the surface potential in the wear region is observed when the following occur: the lubricant

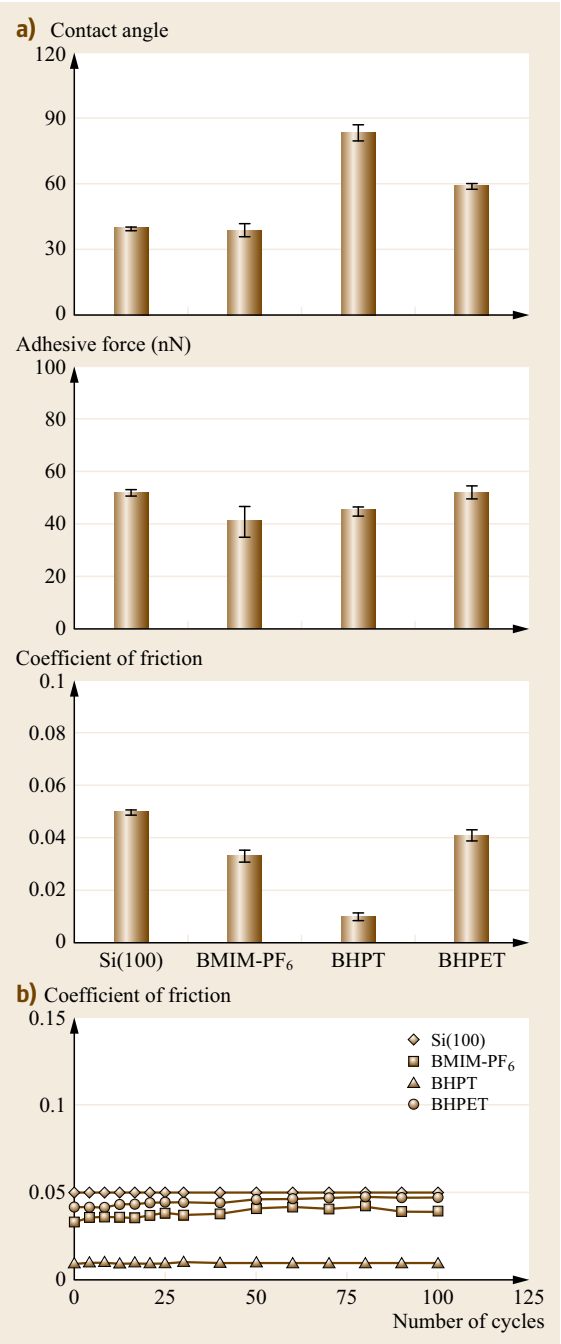


Fig. 35.42 **(a)** Summary of the contact angle, adhesive force and coefficient of friction, and **(b)** durability data after 100 cycles for BMIM-PF_6, BHPT and BHPET coatings at room temperature (22 °C) and ambient air (45–55% RH). Data for uncoated Si are shown for comparison. The error bars in **(a)** represent $\pm 1\sigma$ based on three measurements performed (after [35.21])

has been fully removed from the substrate, the native SiO_2 layer has been abraded from the surface, wear has caused subsurface structural changes, and charges build up as they are unable to dissipate into the surrounding material [35.21, 62]. As expected, the uncoated Si exhibited the greatest amount of wear (as evidenced by debris buildup around the edge of the wear test region) and highest increase in the surface potential. The surface potential image for BHPET film also showed an increase, indicating that the film was worn out after the test. This was not seen on tests with the BMIM-PF_6 and BHPT samples. The surface potential change could be absent in the test area if the lubricant was not removed completely, indicating that these two samples have a stronger interaction with the silicon substrate compared to BHPET.

Figure 35.44 shows a summary of the average contact resistance change after the wear test. As shown in Fig. 35.43, a significant change is observed in the wear region of the Si and BHPET samples. There is also a small amount of localized contact resistance increase in the BMIM-PF_6 sample. These results are consistent with the wear test presented in Fig. 35.42b, where the films are in this order of decreasing durability: BHPT > BMIM-PF_6 > BHPET.

Microscale Friction and Wear

In order to compare friction and wear properties at the microscale and the nanoscale, conventional ball-on-flat tribometer experiments were conducted on the same samples. The coefficient of friction data are summarized in Fig. 35.45. The data shown in the bar chart is an average of three measurements. All of the lubricated samples are reported to have less wear scars as a result of the ball having to displace the lubri-

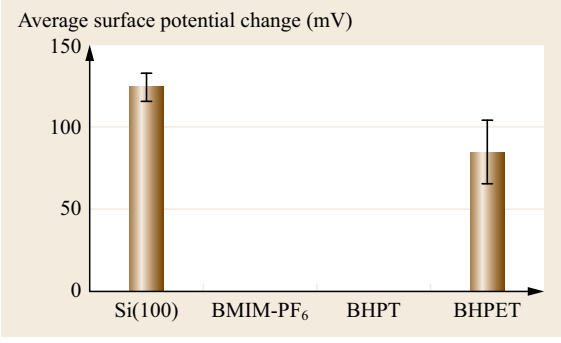


Fig. 35.43 Bar chart showing average surface potential change for BMIM-PF_6, BHPT and BHPET coatings. Data for uncoated Si are shown for comparison. The error bars represent $\pm 1\sigma$ based on three measurements performed (after [35.21])

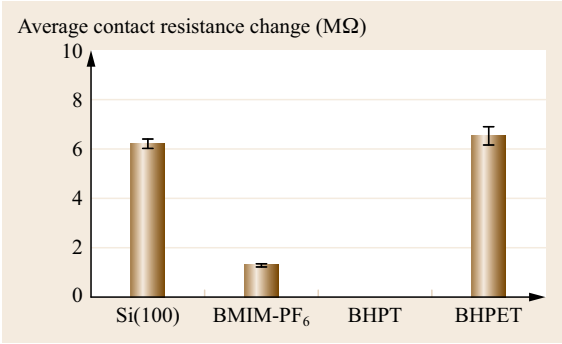


Fig. 35.44 Bar chart showing average contact resistance change for BMIM-PF_6, BHPT and BHPET coatings. Data for uncoated Si are shown for comparison. The error bars represent $\pm 1\sigma$ based on three measurements performed (after [35.21])

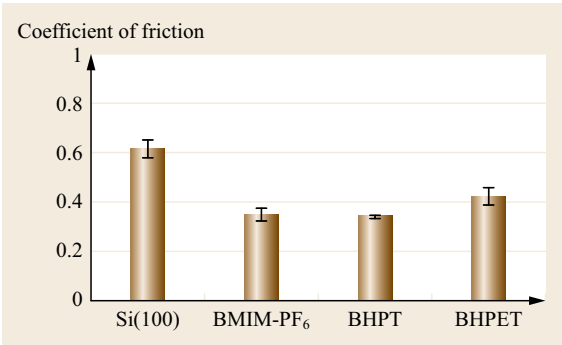


Fig. 35.45 Summary of the coefficient of friction from ball-on-flat tests on uncoated and coated Si samples. The error bars represent $\pm 1\sigma$ based on three measurements performed (after [35.21])

cant before damaging the silicon surface. A reduction in the coefficient of friction arising from application of the lubricant film is observed, which is consistent with the nanoscale adhesion, friction and wear results.

The μ values of the lubricant samples obtained by using AFM are lower than the μ obtained from the ball-on-flat tests. This is attributed to the difference in the length scales of the test techniques. An AFM tip simulates a single asperity contact while the conventional friction test involves the contact of multiple asperities present in the test system. With regards to wear, the interface contact of the AFM and ball-on-flat techniques are different from each other such that one cannot expect both tests to show the same trend. On an AFM, the tip stress is very high such that material can be displaced more easily. For a ball-on-flat test, the tip exerts a lower pressure on the surface and the coating is in a confined geometry.

Relative Humidity and Temperature Effect Measurements

The influence of relative humidity on adhesion and friction is summarized in Fig. 35.46. In general, the adhesive force increases with the relative humidity. The condensed water in the humid environment facilitates meniscus formation between the tip and sample and higher adhesive forces. Since Si(100) is hydrophilic, it readily adsorbs water molecules. For the three ionic liquids, an increase in the adhesive force is also due to increased water adsorption. This comes from attractive electrostatic interactions (ion-dipole forces) between the individual ions and water molecules.

Water adsorption affects the observed coefficient of friction as a function of the relative humidity. In Si, the coefficient of friction is uniform at 10–50% RH, then decreases at 70% RH. The adsorbed water at higher humidity can lead to the formation of a continuous water layer separating the tip and sample surface, which can act as a lubricant. Although the adhesive force increases, the reduction in interfacial strength accounts for the slight decrease of the coefficient of friction at the highest range of humidity level examined. However, the presence of more water molecules at higher humidity has an opposite effect on the ionic liquid surfaces, where the coefficient of friction increases with humidity. The attractive ion-dipole forces between the ions and water are amplified at higher humidity because more water molecules are available. A greater attractive force between tip and surface leads to greater resistance to sliding and a higher coefficient of friction. This is observed on both the monocationic (BMIM-PF_6) and the dicationic (BHPT and BHPET) ionic liquid-coated surfaces. For BHPET, polar interactions between water and the oxygen atoms in the polyether (C–O–C) chain is possible and increases the water adsorption to the surface. This could account for the larger rise in the coefficient of friction in the BHPET sample from 50 to 70% RH, compared to BMIM-PF_6 and BHPT.

The effect of temperature on the adhesion and friction properties of the ILs is summarized in Fig. 35.47. The adhesive and friction forces were measured from 22 to 125 °C. As shown in Fig. 35.47, the increase in test temperature leads to a decrease in the adhesive force and the coefficient of friction. The decrease in the adhesive force at higher temperatures is observed in all the samples, while the corresponding drop in the coefficient of friction is seen only for BMIM-PF_6, BHPET and the silicon substrate. At higher temperatures, the surface water molecules are desorbed, leading to the decrease in both the adhesive and friction forces. A reduction in the viscosity at higher temperatures can also facilitate the decrease in the friction force [35.25]. In BHPT, the coefficient of friction was not adversely

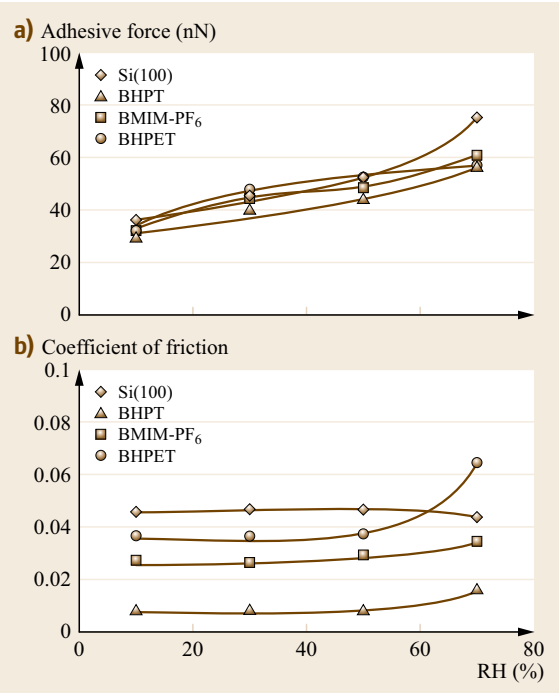


Fig. 35.46a,b Influence of the relative humidity (RH) on the adhesive force **(a)** and coefficient of friction **(b)** for unlubricated and lubricated tapes at 22 °C (after [35.21])

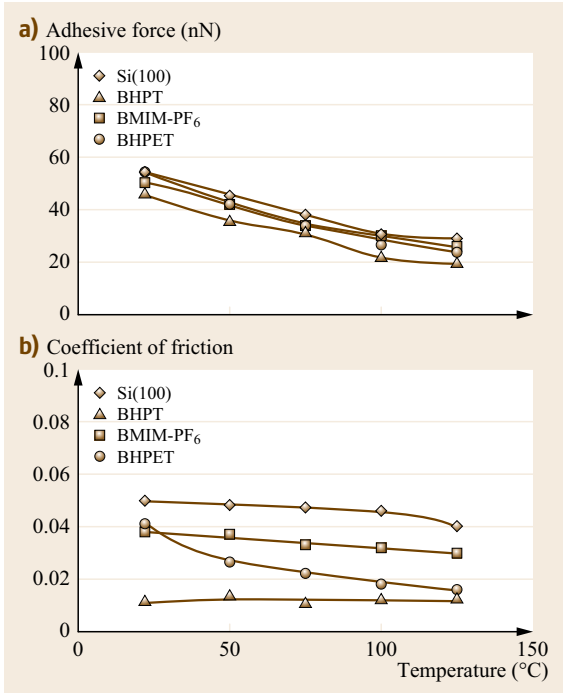


Fig. 35.47a,b Influence of temperature on the adhesive force **(a)** and coefficient of friction **(b)** for uncoated and coated Si samples at 50% RH air (after [35.21])

affected as the test temperature was increased. This implies that at ambient humidity conditions, the BHPT film does not adsorb a large amount of water molecules. Moreover, this also implies that the BHPT surface has weak interactions with surface water molecules, such that friction force during sliding is not very much affected.

Fourier Transform Infrared Spectroscopy

Figure 35.48 is a summary of the FTIR spectra obtained for the different ionic liquid-coated samples, along with the uncoated Si substrate. The observed peaks are labeled with the chemical bonds that they correspond to. In the case of the Si substrate (with a contact angle of 40°) exposed to water molecules in the ambient, no peaks are observed. C–H stretching vibrations in the coated samples are observed at the 600–800 cm^{-1} range. For BHPET, the strong peak at $\approx$ 1060 cm^{-1} is for the C–O–C vibration, which is prominent due to the presence of the polyether chain in its cation. This peak overlaps with the C–O vibration, which is present in BHPT as the terminal primary alcohol (C−OH). In BMIM-PF_6, a peak appears in this range due to rocking vibrations of the methyl (CH_3) substituent. The peak at 1500–1600 cm^{-1} comes from the C=N vibrations, which is common to all three ionic liquids since they

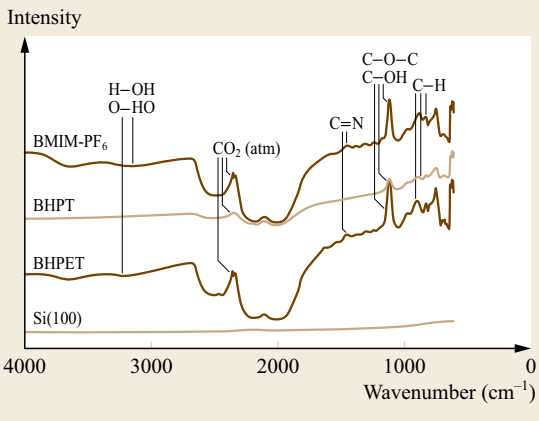


Fig. 35.48 FTIR spectra of uncoated and coated Si samples. The chemical bonds (or species) were listed above the spectra to indicate the possible bonding modes that correspond to the observed peaks (after [35.21])

are all based on the imidazolium cation. However, it is not observed in BHPT as the C=N vibrations may be weak. The wide peak at 3600–4000 cm^{-1} corresponds to hydrogen bonding, possibly due to water molecules adsorbed on the surface [35.21, 78, 79]. This is present in BMIM-PF_6 and in BHPET but not in BHPT. This

accounts for the much lower contact angle of BMIM-PF_6 and BHPET (39° and 59° respectively) compared to BHPT (81°). This implies that the surface of BHPT is more hydrophobic compared to either BMIM-PF_6 or BHPET, which is consistent with the observed low coefficient of friction for BHPT [35.21].

X-ray Photoelectron Spectroscopy

The XPS survey spectra obtained on the uncoated and coated Si samples are shown in Fig. 35.49a. In the uncoated sample, prominent peaks are observed at 99, 151 and 533 eV, corresponding to the binding energies of Si 2p, Si 2s and O 1s electrons respectively. Additional peaks are observed on the three coated samples at 285 and 689 eV, which correspond to C 1s and F 1s electrons respectively. Smaller peaks observed at approximately 1000 eV on all samples are due to Auger lines (KLL transitions) for oxygen [35.80].

The high-resolution, best-fit XPS spectra are shown in Figure 35.49b for the Si 2p, C 1s, O 1s and F 1s electrons. The peaks are labeled with the corresponding chemical bonds, which pertain to either the silicon substrate or groups found on the ionic liquid molecule. One noteworthy exception is the presence of peaks at approximately 292 eV, which confirms the presence of CF_2 on the surface. This indicates the immobilization of the BHPT and BHPET ionic liquids, which occurs by the reaction of the anion with the hydroxyl groups present on the silicon surface [35.21, 23, 62].

Relationship Between IL Structure and Adhesion, Friction and Wear Properties

Figure 35.50 is an interpretation of how the IL cations interact with the silicon substrate. For the monocationic BMIM-PF_6, only weak interactions between the imidazolium ring and the silicon surface is expected. For the dicationic ionic liquids, multiple cation attachment schemes are possible. In BHPT, the hydroxyl groups attached to the imidazolium cation at the ends of the chain provide a means for strong H-bonding interactions with active sites on the silicon surface. As shown in Figure 35.50, either one (case 1) or two (case 2) hydroxyl groups can create this bond. The second case is particularly desirable because if the two hydroxyl groups are bonded (i. e., not exposed to the surface), they are not available to interact with water molecules in the ambient, leading to a reduction in the adhesion and friction forces, as well as enhanced wear resistance [35.21].

In BHPET, these hydroxyl group attachment schemes are also applicable. However, the additional mechanism of intramolecular hydrogen bonding can also take place (case 3). This is not as desirable as

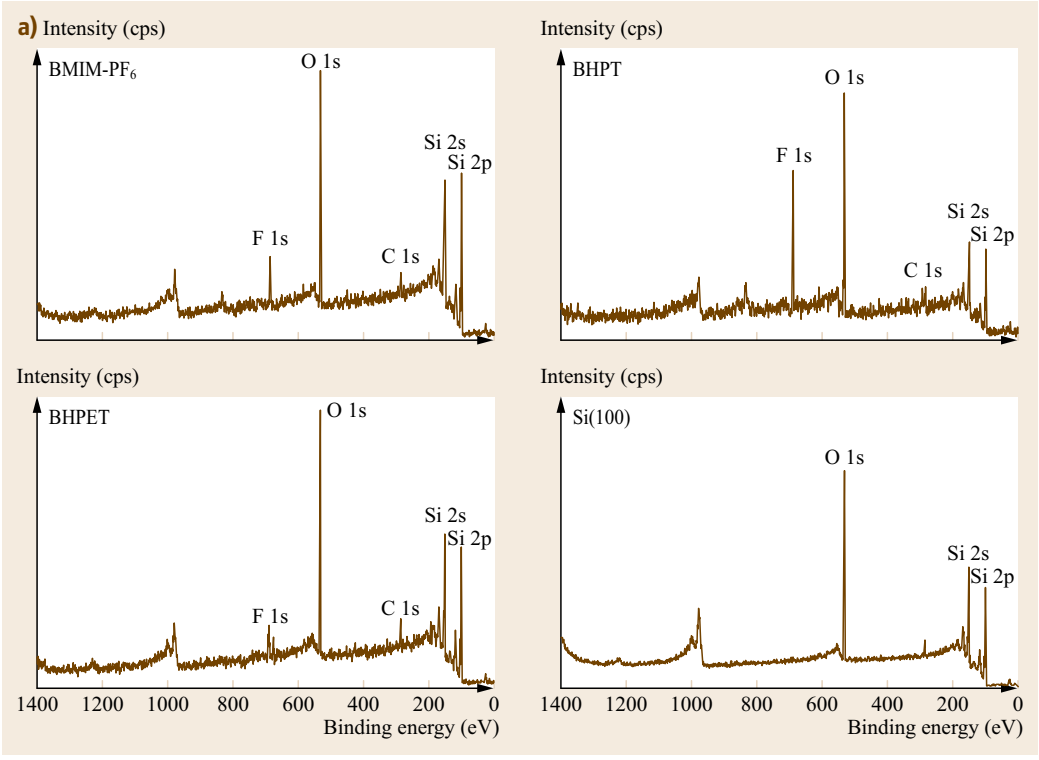


Fig. 35.49
(a) XPS spectra of uncoated and coated Si samples. Survey scan from 0–1400 eV provides the surface elemental composition. **(b)** High-resolution (deconvoluted) XPS spectra for Si 2p, C 1s, O 1s and F 1s reveals the different binding environments present on the surface (after [35.21])

the second case because it depletes the available chain ends with hydroxyl groups, which can bond to the silicon surface. The interaction of the lubricant film with the silicon substrate is weakened, and water molecules can displace the lubricant from the substrate. In addition, the polyether chain that links the two cations contains five oxygen atoms in each chain. These oxygen atoms can also form H-bonds with the water molecules in the ambient. This can account for the large difference in the friction properties between BHPET and BHPT. While the application of the BHPT film has lowered the coefficient of friction of the silicon surface, the BHPET film did not have a similar effect due to the molecular interactions described above. This interpretation of lubricant-substrate interaction is corroborated by the adhesion and friction data at varying humidities and temperatures, as well as the FTIR spectra. Both BMIM-PF_6 and BHPET have large peaks in their FTIR spectra that correspond to H-bonds of water molecules, which is consistent with the observed sensitivity of their adhesion and friction to the change in humidity and temperature. Meanwhile, BHPT does

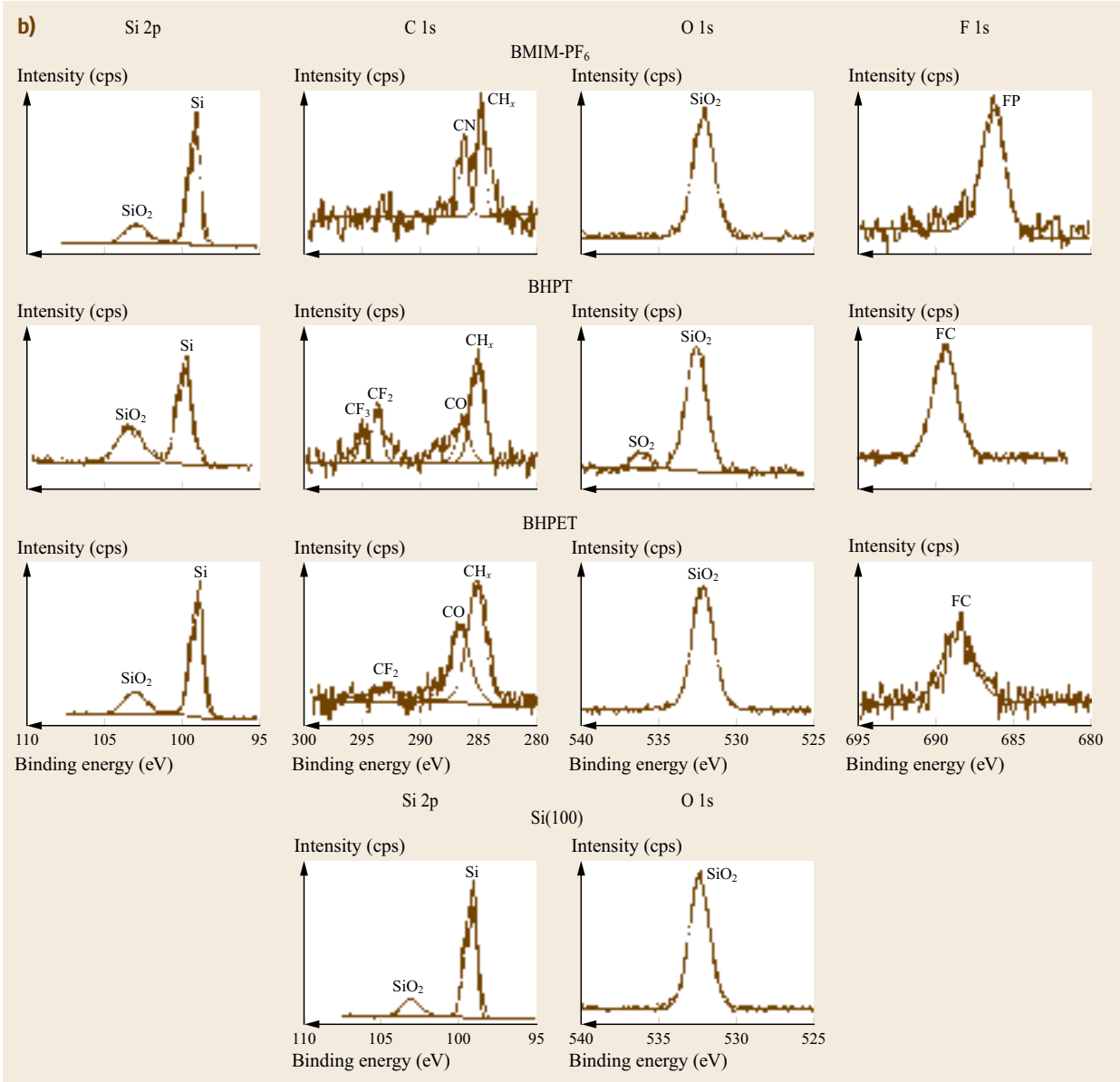


Fig. 35.49 (continued)

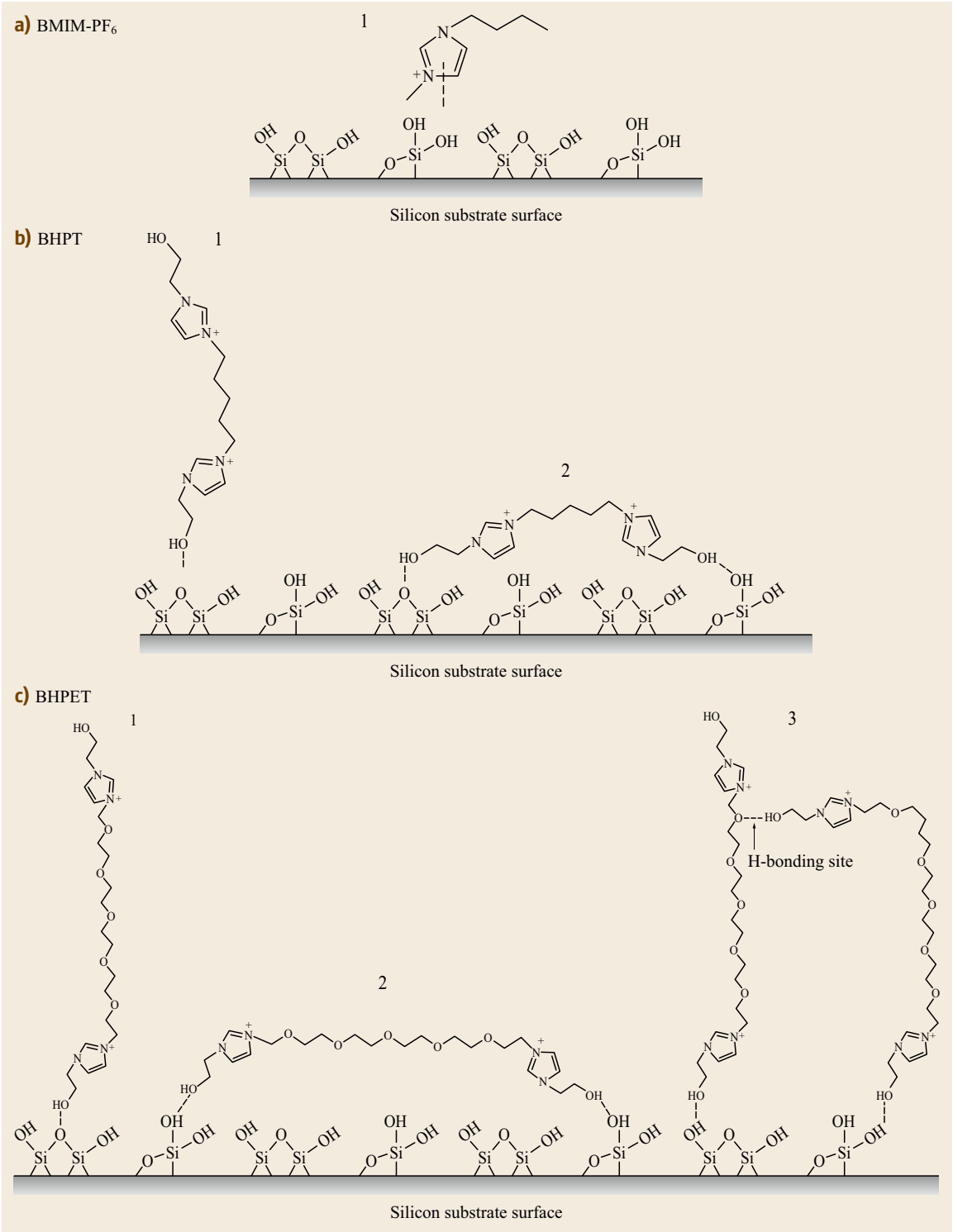


Fig. 35.50a,b Schematic for the attachment of the anions of **(a)** BMIM-PF_6, and **(b)** BHPT and BHPET to the silicon substrate (after [35.21])

not have the aforementioned peak, and its adhesion and friction properties appear to be less sensitive to water molecules compared to the other two ionic liquids [35.21].

Figure 35.51 is a schematic showing the interaction of the anions with the silicon substrate. As mentioned previously, XPS spectra indicate immobilization of the ionic liquid, which occurs through the reaction of the anion with the hydroxyl groups on the substrate surface. In BMIM-PF_6, the O attaches to P, while in BHPT and BHPET (which have the same anion), the O bonds to C. Since the reactions involve the creation of new covalent bonds, this implies that the anions of the ILs investigated are more strongly attached to the substrate compared to the cations, which are chemically adsorbed [35.21].

Based on the micro- and nanoscale friction and wear measurements, the ionic liquids show strong potential as lubricants for MEMS/NEMS because they have desirable thermal and electrical conductivities as well as desirable tribological properties.

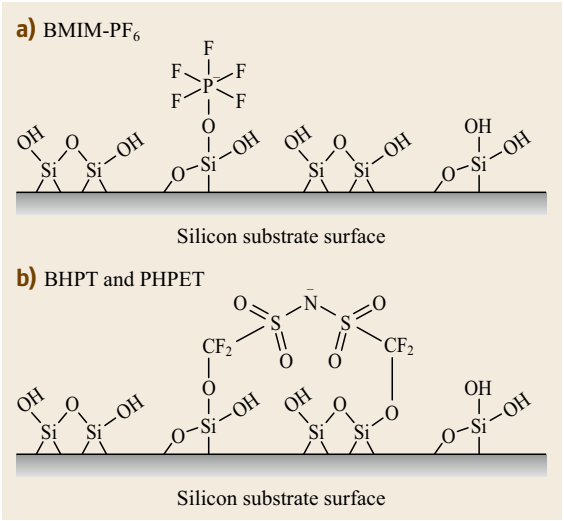


Fig. 35.51 Schematic for the attachment of the cations of **(a)** BMIM-PF_6, **(b)** BHPT and **(c)** BHPET to the silicon substrate (after [35.21])

35.4 Conclusion

Boundary films are formed by physisorption, chemisorption, and chemical reaction. A good boundary lubricant should have a high degree of interaction between its molecules and the solid surface. As a general rule, liquids are good lubricants when they are polar and thus able to grip solid surfaces (or be adsorbed). In this chapter, we presented an overview of various perfluoropolyethers (PFPEs) and ionic liquid films.

Nanodeformation studies have shown that fully bonded PFPE lubricants behave as soft polymer coatings, while the unbonded lubricants behave like liquid. AFM studies have shown that the physisorbed nonpolar molecules on a solid surface have an extended, flat conformation. The spreading property of PFPE is strongly dependent on the molecular end groups and the substrate chemistry.

Using a solid-like Z-DOL (fully bonded) film can reduce the friction and adhesion of Si(100), while using the liquid-like lubricant Z-15 showed a negative effect. Si(100) and a Z-15 film showed an apparent time effect. The friction and adhesion forces increased as a result of the growth of the meniscus up to an equilibrium time, after which they remained constant. The use of a Z-DOL (fully bonded) film can prevent rest time effects. During sliding at high velocity, the meniscus was broken and did not have enough time to rebuild, which led to a decrease of friction force and adhesive force of Z-15 and Z-DOL (fully bonded). The influence of relative humidity on friction and adhesion is dominated by the amount of adsorbed water molecules. Increasing humidity can either increase friction through increased adhesion by water menisci or reduce friction through an enhanced water-lubricating effect. Increasing temperature led to desorption of the water layer, a decrease in water surface tension, a decrease in viscosity, and easier orientation of the Z-DOL (fully bonded) molecules. These changes caused a decrease of friction force and adhesion at high temperature. During cycling tests, the molecular interaction between the attached Z-15 molecules to the tip and the Z-15 molecules on the film surface caused the initial increase in friction. Wear tests showed that Z-DOL (fully bonded) can improve the wear resistance of silicon. Partially bonded PFPE films appear to be more durable than fully bonded films. This suggests that partially/fully bonded films are good lubricants for micro/nanoscale devices operating in different environmental conditions.

The surface potential technique has been useful in detecting lubricant removal and the initiation of substrate wear. The increase in surface potential is attributed to the change in the work function of the silicon after wear and electrostatic charge buildup of debris in the lubricant. Coatings with a mobile lubricant fraction were better able to protect the silicon substrate from wear compared to the fully bonded coating. This enhanced protection is attributed to a lubricant replen-

ishment mechanism. The contact resistance technique provides complementary information to surface potential data in detecting exposure of the substrate after wear and is a promising method for studying conducting lubricants.

Degradation of novel PFPE lubricants – Z-DOL, Z-TETRAOL, A20H, and Z-DOL+A20H (30 vol.%) – was studied. The coefficient of friction of Z-TETRAOL was lower than Z-DOL and A20H in high vacuum. Fully bonded Z-TETRAOL exhibited higher durability than fully bonded Z-DOL on Si, and both were more durable than untreated Z-DOL. A20H showed very high durability in high vacuum. Environment influences the friction and durability of PFPE films. Generally, PFPE films are less durable in high vacuum than in other environments because of intimate contact between surfaces. In argon, the PFPE films showed low friction and high durability. Water molecules can act as a lubricant for PFPE films at a moderate humidity level, while at a high-humidity level they can penetrate the PFPE films and cause increased friction. The durability of fully bonded Z-TETRAOL and partially bonded A20H is higher than that of Z-DOL in tested environments. The mixture of Z-DOL+A20H (30 vol.%) showed low durability in high vacuum while it showed high durability in argon and air with various humidity levels.

The ionic liquids exhibited low adhesion, friction, and wear properties comparable to the PFPE lubricants. For example, ionic liquid BMIM-PF_6 is comparable to Z-TETRAOL. Based on the surface height, adhesion, and friction data, chemical bonding treatment facilitated attachment of the ionic liquid to the silicon substrate surface, leading to a more uniform coating and lower adhesion force and coefficient of friction. The partially bonded coatings have the lowest coefficients of friction and longest durability as they possess a desirable combination of lubricant bonded to the substrate, as well as a mobile fraction that facilitates sliding. In micro- and nanoscale experiments, the ionic liquid exhibited comparable durability with the lubricant Z-TETRAOL, which has high thermal stability and extremely low vapor pressure. The low postwear surface potential change observed on the silicon coated with ionic liquid is indicative of enhanced charge dissipation compared to the Z-TETRAOL-coated and uncoated surfaces, which are poor conductors. Contact resistance data is consistent with the surface potential data with regards to identifying the role of the mobile and immobile lubricant fractions in protecting the surface from wear.

The dicationic liquid BHPT exhibits superior nanoscale friction and wear resistance properties. This is attributed to the presence of a pentyl chain and hydroxyl groups on both chain ends, which facilitate molecular orientation, as well as bonding interactions with the substrate surface. The other dicationic liquid, BHPET, has less desirable adhesion, friction, and wear properties compared to either BHPT or BMIM-PF_6. Intermolecular hydrogen bonding in BHPET reduces the chain ordering on the substrate surface, which accounts for the observed higher adhesive force and coefficient of friction relative to the other ionic liquids investigated. From the nanoscale wear tests, surface potential, and contact resistance imaging, it was found that the films are in this order of decreasing durability: BHPT > BMIM-PF_6 > BHPET. The microscale coefficient of friction was higher than the nanoscale value due to the difference in the length scales and configurations of the two test methods.

Nanotribological experiments performed at various humidities and temperature conditions indicated that the adhesive force and coefficient of friction of the ionic liquid films, especially BMIM-PF_6 and BHPET, are highly sensitive to the amount of water molecules present on the surface. This was confirmed by FTIR spectroscopy, where the BMIM-PF_6 and BHPET films showed peaks corresponding to hydrogen bonding mainly resulting from water adsorption and partially from intermolecular interactions. The absence of this peak in BHPT indicated that a much smaller amount of water was adsorbed by this ionic liquid, which is consistent with its superior nanotribological properties. The anions are more strongly attached to the substrate compared to the cations. The anions could be covalently bonded, while the cations are chemically adsorbed. XPS analysis confirmed the immobilization of the anion to the silicon substrate.

In summary, PFPE and ionic liquids can be used for nanoscale boundary lubrication in demanding environments.

References

35.1 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

35.2 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)

35.3 B. Bhushan: Magnetic recording surfaces. In: *Characterization of Tribological Materials*, ed. by W.A. Glaeser (Butterworth-Heinemann, Boston 1993) pp. 116–133

35.4 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)
35.5 B. Bhushan: Macro- and microtribology of magnetic storage devices. In: *Modern Tribology Handbook, Vol. 2: Materials, Coatings, and Industrial Applications*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 1413–1513
35.6 V.J. Novotny, I. Hussla, J.M. Turlet, M.R. Philpott: Liquid polymer conformation on solid surfaces, J. Chem. Phys. **90**, 5861–5868 (1989)
35.7 V.J. Novotny: Migration of liquid polymers on solid surfaces, J. Chem. Phys. **92**, 3189–3196 (1990)
35.8 C.M. Mate, V.J. Novotny: Molecular conformation and disjoining pressures of polymeric liquid films, J. Chem. Phys. **94**, 8420–8427 (1991)
35.9 C.M. Mate: Application of disjoining and capillary pressure to liquid lubricant films in magnetic recording, J. Appl. Phys. **72**, 3084–3090 (1992)
35.10 G.G. Roberts: *Langmuir–Blodgett Films* (Plenum, New York 1990)
35.11 A. Ulman: *An Introduction to Ultrathin Organic Films* (Academic, Boston 1991)
35.12 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)
35.13 B. Bhushan, Z. Zhao: Macroscale and microscale tribological studies of molecularly thick boundary layers of perfluoropolyether lubricants for magnetic thin-film rigid disks, J. Inf. Storage Process. Syst. **1**, 1–21 (1999)
35.14 E. Hoque, J.A. DeRose, B. Bhushan, H.J. Mathieu: Self-assembled monolayers on aluminum and copper oxide surfaces: Surface and interface characteristics, nanotribological properties, and chemical stability. In: *Applied Scanning Probe Methods vol. IX – Characterization*, ed. by B. Bhushan, H. Fuchs, M. Tomitori (Springer, Heidelberg 2008) pp. 235–281
35.15 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer Academic, Dordrecht 1998)
35.16 B. Bhushan: Nanotribology and nanomechanics in nano/biotechnology, Philos. Trans. R. Soc. A **366**, 1499–1537 (2008)
35.17 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Cham 2016)
35.18 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374**, 607–616 (1995)
35.19 B. Bhushan: Nanotribology, nanomechanics and nanomaterials characterization, Philos. Trans. R. Soc. A **366**, 1351–1381 (2008)
35.20 M. Palacio, B. Bhushan: Ultrathin wear-resistant ionic liquid films for novel MEMS/NEMS applications, Adv. Mater. **20**, 1194–1198 (2008)
35.21 M. Palacio, B. Bhushan: Molecularly thick dicationic liquid films for nanolubrication, J. Vac. Sci. Technol. A **27**, 986–995 (2009)
35.22 B. Bhushan, K. Kwak: Role of lubricants, scanning velocity, and environment on adhesion, friction and wear of Pt-Ir coated probes for atomic force microscopy probe-based ferroelectric recording technology, J. Phys. Condes. Matter **20**, 325240 (2008)
35.23 B. Bhushan, K. Kwak, M. Palacio: Nanotribology and nanomechanics of AFM probe based data recording technology, J. Phys. Condens. Matter **20**, 365207 (2008)
35.24 M. Palacio, B. Bhushan: Nanotribological and nanomechanical properties of lubricated PZT thin films for ferroelectric data storage applications, J. Vac. Sci. Technol. A **26**, 768–776 (2008)
35.25 H. Liu, B. Bhushan: Nanotribological characterization of molecularly-thick lubricant films for applications to MEMS/NEMS by AFM, Ultramicroscopy **97**, 321–340 (2003)
35.26 J. Ruhe, G. Blackman, V.J. Novotny, T. Clarke, G.B. Street, S. Kuan: Thermal attachment of perfluorinated polymers to solid surfaces, J. Appl. Polym. Sci. **53**, 825–836 (1994)
35.27 R.A. Reich, P.A. Stewart, J. Bohaychick, J.A. Urbanski: Base oil properties of ionic liquids, Lub. Eng. **49**, 16–21 (2003)
35.28 V.N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14**, 2378–2391 (1996)
35.29 G.S. Blackman, C.M. Mate, M.R. Philpott: Interaction forces of a sharp tungsten tip with molecular films on silicon surface, Phys. Rev. Lett. **65**, 2270–2273 (1990)
35.30 G.S. Blackman, C.M. Mate, M.R. Philpott: Atomic force microscope studies of lubricant films on solid surfaces, Vacuum **41**, 1283–1286 (1990)
35.31 C.A. Kim, H.J. Choi, R.N. Kono, M.S. Jhon: Rheological characterization of perfluoropolyether lubricant, Polym. Prepr. **40**, 647–649 (1999)
35.32 M. Ruths, S. Granick: Rate-dependent adhesion between opposed perfluoropoly(alkylether) layers: Dependence on chain-end functionality and chain length, J. Phys. Chem. B **102**, 6056–6063 (1998)
35.33 X. Ma, J. Gui, K.J. Grannen, L.A. Smoliar, B. Marchon, M.S. Jhon, C.L. Bauer: Spreading of PFPE lubricants on carbon surfaces: Effect of hydrogen and nitrogen content, Tribol. Lett. **6**, 9–14 (1999)
35.34 U. Jonsson, B. Bhushan: Measurement of rheological properties of ultrathin lubricant films at very high shear rates and near-ambient pressure, J. Appl. Phys. **78**, 3107–3109 (1995)
35.35 C. Hahm, B. Bhushan: High shear rate viscosity measurement of perfluoropolyether lubricants for magnetic thin-film rigid disks, J. Appl. Phys. **81**, 5384–5386 (1997)
35.36 C.M. Mate: Atomic-force-microscope study of polymer lubricants on silicon surface, Phys. Rev. Lett. **68**, 3323–3326 (1992)
35.37 C.M. Mate: Nanotribology of lubricated and unlubricated carbon overcoats on magnetic disks studied by friction force microscopy, Surf. Coat. Technol. **62**, 373–379 (1993)
35.38 S.J. O'Shea, M.E. Welland, T. Rayment: Atomic force microscope study of boundary layer lubrication, Appl. Phys. Lett. **61**, 2240–2242 (1992)
35.39 S.J. O'Shea, M.E. Welland, J.B. Pethica: Atomic force microscopy of local compliance at solid-liquid interface, Chem. Phys. Lett. **223**, 336–340 (1994)

35.40 B. Bhushan, L. Yang, C. Gao, S. Suri, R.A. Miller, B. Marchon: Friction and wear studies of magnetic thin-film rigid disks with glass-ceramic, glass, and aluminum-magnesium substrates, Wear **190**, 44–59 (1995)

35.41 V.N. Koinkar, B. Bhushan: Micro/nanoscale studies of boundary layers of liquid lubricants for magnetic disks, J. Appl. Phys. **79**, 8071–8075 (1996)

35.42 B. Bhushan, S. Sundararajan: Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46**, 3793–3804 (1998)

35.43 B. Bhushan, C. Dandavate: Thin-film friction and adhesion studies using atomic force microscopy, J. Appl. Phys. **87**, 1201–1210 (2000)

35.44 M. Palacio, B. Bhushan: Surface potential and resistance measurements for detecting wear of chemically-bonded and unbonded molecularly-thick perfluoropolyether lubricant films using atomic force microscopy, J. Colloid Interf. Sci. **315**, 261–269 (2007)

35.45 M. Palacio, B. Bhushan: Nanotribological properties of novel lubricants for magnetic tapes, Ultramicroscopy **109**, 980–990 (2009)

35.46 S. Sundararajan, B. Bhushan: Static friction and surface roughness studies of surface micromachined electrostatic micromotors using an atomic force/friction force microscope, J. Vac. Sci. Tech. A **19**, 1777–1785 (2001)

35.47 B. Bhushan, J. Ruan: Atomic-scale friction measurements using friction force microscopy: Part II – application to magnetic media, ASME J. Tribol. **116**, 389–396 (1994)

35.48 T. Stifter, O. Marti, B. Bhushan: Theoretical investigation of the distance dependence of capillary and van der Waals forces in scanning probe microscopy, Phys. Rev. B **62**, 13667–13673 (2000)

35.49 J.N. Israelachvili: *Intermolecular and Surface Forces*, 2nd edn. (Academic, London 1992)

35.50 S.K. Chilamakuri, B. Bhushan: A comprehensive kinetic meniscus model for prediction of long-term static friction, J. Appl. Phys. **15**, 4649–4656 (1999)

35.51 H. Ishigaki, I. Kawaguchi, M. Iwasa, Y. Toibana: Friction and wear of hot pressed silicon nitride and other ceramics, ASME J. Tribol. **108**, 514–521 (1986)

35.52 T.E. Fischer: Tribochemistry, Annu. Rev. Mater. Sci. **18**, 303–323 (1988)

35.53 K. Mizuhara, S.M. Hsu: Tribochemical reaction of oxygen and water on silicon surfaces. In: *Wear Particles*, ed. by D. Dowson (Elsevier, New York 1992) pp. 323–328

35.54 S. Danyluk, M. McNallan, D.S. Park: Friction and wear of silicon nitride exposed to moisture at high temperatures. In: *Friction and Wear of Ceramics*, ed. by S. Jahanmir (Marcel Dekker, New York 1994) pp. 61–79

35.55 V.A. Muratov, T.E. Fischer: Tribochemical polishing, Annu. Rev. Mater. Sci. **30**, 27–51 (2000)

35.56 H. Yoshizawa, Y.L. Chen, J.N. Israelachvili: Fundamental mechanisms of interfacial friction I: Relationship between adhesion and friction, J. Phys. Chem. **97**, 4128–4140 (1993)

35.57 H. Yoshizawa, J.N. Israelachvili: Fundamental mechanisms of interfacial friction II: Stick slip friction of spherical and chain molecules, J. Phys. Chem. **97**, 11300–11313 (1993)

35.58 J. Ruhe, V. Novotny, T. Clarke, G.B. Street: Ultrathin perfluoropolyeher films-influence of anchoring and mobility of polymers on the tribological properties, ASME J. Tribol. **118**, 663–668 (1996)

35.59 K.C. Eapen, S.T. Patton, J.S. Zabinski: Lubrication of microelectromechanical systems (MEMS) using bound and mobile phase of fomblin Z-DOL, Tibol. Lett. **12**, 35–41 (2002)

35.60 D. DeVecchio, B. Bhushan: Use of a nanoscale kelvin probe for detecting wear precursors, Rev. Sci. Instrum. **69**, 3618–3624 (1998)

35.61 M. Palacio, B. Bhushan: Wear detection of candidate MEMS/NEMS lubricant films using atomic force microscopy-based surface potential measurements, Scripta Mater. **57**, 821–824 (2007)

35.62 B. Bhushan, M. Palacio, B. Kinzig: AFM-based nanotribological and electrical characterization of ultrathin wear-resistant ionic liquid films, J. Colloid Interf. Sci. **317**, 275–287 (2008)

35.63 Z. Tao, B. Bhushan: Bonding, degradation, and environmental effects on novel perfluoropolyether lubricants, Wear **259**, 1352–1361 (2005)

35.64 B. Bhushan, Z. Tao: Lubrication of advanced metal evaporated tape using novel perfluoropolyether lubricants, Microsyst. Technol. **12**, 579–587 (2006)

35.65 B. Bhushan, M. Cichomski, Z. Tao, N.T. Tran, T. Ethen, C. Merton, R.E. Jewett: Nanotribological characterization and lubricant degradation studies of metal-film magnetic tapes using novel lubricants, ASME J. Tribol. **129**, 621–627 (2007)

35.66 B. Bhushan, J. Ruan: Tribological performance of thin film amorphous carbon overcoats for magnetic recording disks in various environments, Surf. Coat. Technol. **68/69**, 644–650 (1994)

35.67 X. Zhao, B. Bhushan: Comparison studies on degradation mechanisms of perfluoropolyether lubricants and model lubricants, Tribol. Lett. **9**, 187–197 (2000)

35.68 X. Zhao, B. Bhushan: Studies on degradation mechanisms of lubricants for magnetic thin-film rigid disks, Proc. Inst. Mech. Eng., Part J **215**, 173–188 (2001)

35.69 X. Zhao, B. Bhushan, C. Kajdas: Lubrication studies of head-disk interfaces in a controlled environment, part 2: Degradation mechanisms of perfluoropolyether lubricants, Proc. Inst. Mech. Eng., Part J **214**, 547–559 (2000)

35.70 B.J. Kinzig, P. Sutor: Ionic liquids: Novel lubrication for air and space, Phase I Final Report for AFOSR/NL (Surfaces Research and Applications, Lenexa 2005)

35.71 H. Wang, Q. Lu, C. Ye, W. Liu, Z. Cui: Friction and wear behaviors of ionic liquid of alkylimidazolium hexafluorophosphates as lubricants for steel/steel contacts, Wear **256**, 44–48 (2004)

35.72 G.J. Kabo, A.V. Blokhin, Y.U. Paulechka, A.J. Kabo, M.P. Shymanovich, J.W. Magee: Thermodynamic properties of 1-butyl-3-methylimidazolium hexafluorophosphate in the condensed state, J. Chem.

Eng. Data **49**, 453–461 (2004)
35.73 C. Frez, G.J. Diebold, C.D. Tran, S. Yu: Determination of thermal diffusivities, thermal conductivities and sound speed of room-temperature ionic liquids by the transient grating technique, J. Chem. Eng. Data **51**, 1250–1255 (2006)
35.74 S. Carda-Broch, A. Berthod, D.W. Armstrong: Solvent properties of the 1-butyl-3-methylimidazolium hexafluorophosphate ionic liquid, Anal. Bioanal. Chem. **375**, 191–199 (2003)
35.75 J.L. Anderson, R. Ding, A. Ellern, D.W. Armstrong: Structure and properties of high stability germinal dicationic ionic liquids, J. Am. Chem. Soc. **127**, 593–604 (2005)
35.76 T. Payagala, J. Huang, Z.S. Breitbach, P.S. Sharma, D.W. Armstrong: Unsymmetrical dicationic ionic liquids: Manipulation of physicochemical properties using specific structural architectures, Chem. Mater. **19**, 5848–5850 (2007)
35.77 M.H. Valkenberg: C. deCastro, W.F. Holderich: Immobilization of ionic liquids on solid supports, Green Chem. **4**, 88–93 (2002)
35.78 R.T. Morrison, R.N. Boyd: *Organic Chemistry*, 6th edn. (Wiley, New York 1992)
35.79 L. Zhang, Q. Zhang, J. Li: Electrochemical behaviors and spectral studies of ionic liquid (1-butyl-3-methylimidazolium tetrafluoroborate) based on sol-gel electrode, J. Electroanal. Chem. **603**, 243–248 (2007)
35.80 L.C. Feldman, J.W. Mayer: *Fundamentals of Surface and Thin Film Analysis* (Prentice-Hall, Upper Saddle River 1986)

Biomime

Part H

Part H Biomimetics and Bioinspired Surfaces

36 **Plant Surfaces: Structures and Functions for Biomimetic Applications**
Wilhelm Barthlott, Bonn, Germany
Matthias Mail, Bonn, Germany
Bharat Bhushan, Columbus, USA
Kerstin Koch, Kleve, Germany

37 **Bioinspired Nanostructured Antibiofouling and Anti-inorganic Surfaces**
Bharat Bhushan, Columbus, USA

36. Plant Surfaces: Structures and Functions for Biomimetic Applications

Wilhelm Barthlott, Matthias Mail, Bharat Bhushan, Kerstin Koch

This chapter presents an overview of plant surface structures and their evolution, combines surface chemistry and architecture with their functions and refers to possible biomimetic applications.

Within some 3.5 billion years biological species evolved highly complex multifunctional surfaces for interacting with their environments, providing some 10 million living prototypes (i. e., estimated number of existing plants and animals) for engineers. The complexity of the hierarchical structures and their functionality in biological organisms surpasses all abiotic natural surfaces: even superhydrophobicity is restricted in nature to living organisms and was probably a key evolutionary step in the invasion of terrestrial habitats some 350–450 million years ago by plants and insects. Special attention should be paid to the fact that global environmental change implies a dramatic loss of species and with it many biological role models.

Plants, the dominating group of organisms on our planet, are sessile organisms with large multifunctional surfaces and thus exhibit particular intriguing features. Superhydrophilicity and superhydrophobicity are focal points in this chapter. We estimate that superhydrophobic plant leaves (e.g., grasses) comprise in total an area of around 250 million km^2, which is about 50% of the total surface of our planet.

A survey of structures and functions based on our own examinations of almost 20 000 species is provided; for further references we refer to [36.1]. A basic difference exists between aquatic nonvascular and land-living vascular plants; the latter exhibit a particularly intriguing surface chemistry and architecture. The diversity of features is described in detail according to their hierarchical structural order. The first underlying and essential feature is the polymer cuticle superimposed by epicuticular wax and the curvature of single cells up to complex multicellular structures. A descriptive terminology for this diversity is provided.

Simplified, the functions of plant surface characteristics may be grouped into six categories:

1. Mechanical properties
2. Influence on reflection and absorption of spectral radiation
3. Reduction of water loss or increase of water uptake, moisture harvesting
4. Adhesion and nonadhesion (lotus effect, insect trapping)
5. Drag and turbulence increase
6. Air retention underwater for drag reduction or gas exchange (Salvinia effect).

This list is far from complete.

A short overview of the history of bionics and the impressive spectrum of existing and anticipated biomimetic applications are provided. The major challenge for engineers and materials scientists, the durability of the fragile nanocoatings, is also discussed.

36.1 **500 Million Years of Evolution for Innovative Technologies** 1266

36.2 **Chemistry of Plant Surfaces** 1270
36.2.1 Chemical Composition of Wax 1271
36.2.2 Chemical Heterogeneities 1274
36.2.3 Crystallinity 1275

36.3 **Structuring of Plant Surfaces: Hierarchical Architecture Between Nano- and Macrostructures** 1279
36.3.1 The Cuticle 1279
36.3.2 Hierarchical Sculpturing 1279
36.3.3 First Sculptural Level 1280
36.3.4 Second Sculptural Level 1281
36.3.5 Third Sculptural Level 1283
36.3.6 Fourth Sculptural Level 1284

36.4 **Physical Basis of Surface Wetting** 1285
36.4.1 Static Wetting Processes 1285
36.4.2 Dynamic Interactions 1286

36.5 **Superhydrophilic and Superhydrophobic Plant Surfaces** 1286

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_36

36.6 **Functional Diversity of Plant Surfaces** 1289
36.6.1 Mechanical Properties 1289
36.6.2 Attachment 1289
36.6.3 Reflection, Absorption, and Transmission of Spectral Radiation 1289
36.6.4 Reduction of Water Loss 1291
36.6.5 Superhydrophilicity 1291
36.6.6 Superhydrophobicity 1291
36.6.7 Anti-adhesive *Slippery* Surfaces and Aquaplaning 1293
36.7 **Biomimetic Application** 1294
36.7.1 Mechanical Properties 1294
36.7.2 Attachment 1294
36.7.3 Reflection, Transmission, and Absorption of Spectral Radiation 1294
36.7.4 Reduction of Water Loss 1294
36.7.5 Superhydrophilicity 1294
36.7.6 Superhydrophobicity 1295
36.7.7 Other Applications 1296
36.8 **Living Prototypes: Evolution of Plant Surfaces and Biodiversity** 1297
36.8.1 Water Plants 1297
36.8.2 Land Plants 1297
36.9 **Conclusions** 1298
References 1298

36.1 500 Million Years of Evolution for Innovative Technologies

Surfaces define the boundaries for the well-structured world of solids, and it is surfaces that define their interactions. They play crucial roles in environmental interactions. This is of particular importance for sessile organisms with large functional surfaces: plants. Green plants cover the terrestrial biomes of our planet and – not surprisingly – show a stunning diversity of hierarchical surface structures, which has been revealed with the help of scanning electron microscopy techniques (SEM) first employed in the 1970s (survey in [36.2]). It is even possible to examine the hierarchical surface structures at the macroscopic scale, as illustrated in two



Fig. 36.1a,b Hierarchical surface sculpturing of plants on the macroscopic scale. **(a)** The Saguaro (*Carnegiea gigantea*) is the largest cactus; it can grow up to 21 m tall. The stems are ribbed – even in the full sun of a desert in Arizona large areas of the plant are shaded. At the same time the ribs and elastic cuticle allow a rapid increase of the volume after sporadic rainfalls: the stems expand. Loss of water is a major problem for desert plants: the Saguaro is encrusted in a wax layer, but due to ultraviolet (UV) exposure the surfaces age and become wettable. Saguaros can live 150 years or more. **(b)** In contrast, the Giant Arum (*Amorphophallus titanum*) lives in the deepest shadows of the humid rainforest understories in Sumatra. Its flower opens for only one to two days; it reaches a height of three meters and is the largest blossom in the plant kingdom. The giant pleated *petal* (spathe) weights less than one kilogram: the largest lightweight construction amongst plants, possibly even in any organism. The riblets serve as mechanical stabilizers: when the spathe opens its surface is hydrophobic to shed rain droplets. Very unusual wax crystals occur on the unpleasant-smelling central column, which heats periodically to almost 40 °C to generate a convection flow to attract insects

of the giants in the plant kingdom; the Saguaro cactus (Fig. 36.1a) and the Titan Arum (Fig. 36.1b). On the other hand, the details of structures like wax crystals on their surface (Fig. 36.4h and Fig. 36.7) are only revealed by scanning electron microscopes.

Pollen and spores exhibit particularly refined hierarchical structures; they are distinct from all other plant surfaces like leaves. The functional properties of pollen (e.g., the pollen of *Cucurbita pepo*, Fig. 36.2a) are associated with the attachment and detachment to the pollinating insect and the stigma of the flower, and possibly temperature control under insolation. Pollen are hydrophilic, and spores are occasionally superhydrophobic, even bacterial spores. Wind-dispersed miniature *dust seeds* (e.g., the seeds of *Aeginetia* and *Triphora*, Fig. 36.3) are dispersed by air and possess a surface roughness to increase their Reynolds number for long-distance dispersal. Pollen and spores are not discussed in this chapter: palynology is a very specialized field and a vast literature exists (e.g., [36.3, 4]). Internal *noncuticular* surfaces (Fig. 36.2) like pollen or conducting vessels are also not considered in the following sections; they are multifunctional and differ very much from *outer* cuticular surfaces of plants. Strangely enough biomimetics has little interest in this particular subject, although applications are plausible.

The diversity of plant surface structures arises from the variability of cell shapes, and hierarchically superimposed micro- and nanostructures of the cell surfaces (mainly wax crystals), and by the formation of multicellular structures (Fig. 36.4) [36.2, 5, 6]. Based on these cellular and subcellular units a nearly unlimited

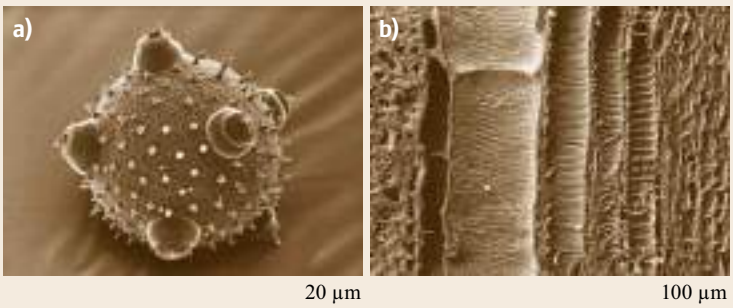


Fig. 36.2a,b Internal (*noncuticular*) functional surfaces are usually hydrophilic – two examples from a squash or pumpkin (*Cucurbita pepo*) are illustrated. **(a)** Pollen grain. Its surface functions are connected with attachment and detachment to the pollinating insect and the stigma of the pumpkin flower, and possibly temperature control under insolation. In wind-dispersed pollen these structures might also increase the Reynolds numbers. **(b)** A vessel-element of the same plant exhibiting complex spiral and perforated structures to transport water within the plant. The structural elements of internal surfaces fundamentally differ from the outer cuticular surfaces (compare e.g., Figs. 36.3, 36.7, 36.14, 36.15)

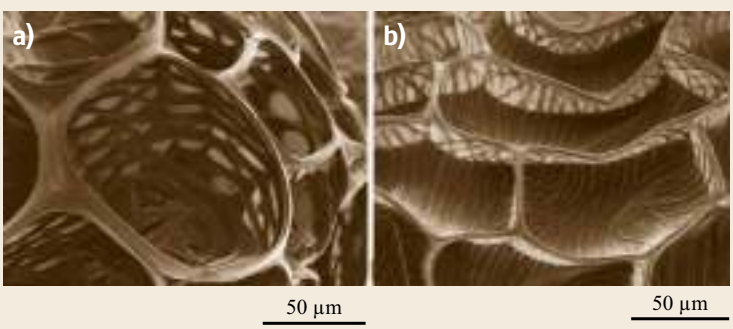
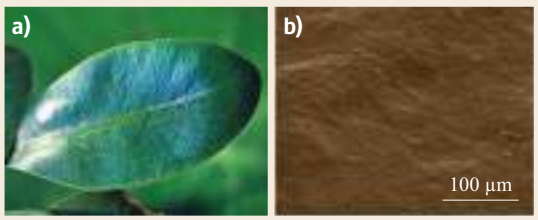
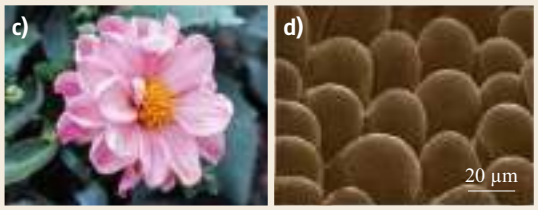


Fig. 36.3a,b SEM micrographs of two seed surfaces with concave cell sculpturing. The miniature seeds of both species **(a)** *Aeginetia indica* and **(b)** *Triphora triantophora* are optimized for seed dispersal by wind. They are hydrophobic and float for a short time in water (compare *Aeginetia* in Fig. 36.27): the concave sculpture of the nonliving cells can be interpreted as a shrinkage deformation during seed maturing and drying. The bands that form an inner network in **(a)** and the surface pattern in **(b)** are built by cellulose. All these features are lightweight constructions and generate high Reynolds numbers to prolong the floating time



Iridescent leaf of *Elaphoglossum wurdackii* with a smooth surface



Velvety leaves of a *Dahlia flower* with convex, structured cells



Hairy leaves of *Leucadendron argenteum* with straight hairs



Waxy leaves of *Eucalyptus macrocarpa* with wax threads

Fig. 36.4a–h Macroscopic optical appearance of plant surfaces and their surface microstructures. In **(a)** the leaves of *Elaphoglossum wurdackii* appear glossy because of a flat surface structure, shown in figure **(b)**. Their iridescence is caused by thin layers within the cuticle. In **(c)** the flower petals of *Dahlia* appear velvety due to the convex microstructure of the epidermis cells, shown in **(d)**. In **(e)** the silvery appearance of the *Leucadendron argenteum* leaves is caused by a dense layer of light-reflecting hairs **(f)**. In **(g)** the leaf and flower bud surfaces of *Eucalyptus macrocarpa* appear white or bluish, caused by a dense covering of thread-like wax crystals **(h)**

combination of structures leads to a high structural and functional diversity of the surfaces of the some 450 000 different species of land plants [36.1]. Superhydrophobicity is one of the most remarkable features of many plant surfaces; most families of higher plants [36.1, 7] include many species where the entirety or part (e.g., only lower side of the leaves) of the assimilating leaf surface is superhydrophobic. Grasses (Fig. 36.5a,b) with a few exceptions (e.g., maize), are superhydrophobic and are, with around 12 000 species, one of the largest plant families that dominates the largest ecosystems of our globe. As is the case in many plants, often only the young leaves are superhydrophobic (Fig. 36.5b), older ones may become wettable: superhydrophobicity is often an instable state in plant surfaces (compare also the Saguaro, Fig. 36.1a) – as is also seen in technical surfaces. Some grasses (like *Elymus arenarius* with a static contact angle of 161°) exhibit similar properties to lotus leaves [36.8]. One square meter of a meadow may exhibit a minimum of four square meters of leaf surfaces (compare e.g., [36.9]). According to UN Food and Agriculture Organization (FAO) assessments [36.10], grasslands (e.g., steppes, savannas, wheat and rice fields) cover some 52.5 million km^2. A rough calculation indicates that at least 250 million km^2 (with other plant families included, possibly much more) are superhydrophobic: this means more than half of the total surface area of our whole planet. In actuality, the dimensions are probably considerably higher and might equal the total surface of our planet.

There is a basic and obvious difference in surface structure and function between aquatic and terrestrial plants. In terrestrial (vascular) plants an epidermis is present, as the specialized outermost cell layer with a cuticle of the primary tissues of all leaves and several other organs plays an important role in environmental interactions and surface structuring. A simplified model presented in Fig. 36.6 shows a layered stratification of the outermost part of epidermis cells. Starting with the outside, one finds a highly functional thin outermost layer; the polymer cuticle with its superimposed waxes. This outermost layer covers nearly all aerial tissues of land-living plants as a continuous extracellular membrane, but is absent in roots. One of the most important attributes of the cuticle is its function as a transpiration barrier, which enables plants to overcome the physical and physiological problems connected to an ambient environment, such as desiccation.

The cuticle is basically a biopolymer made of a polyester called cutin, impregnated with integrated (intracuticular) waxes. Additionally, waxes on the cuticle surface (epicuticular waxes) play an important role in surface structuring at the subcellular scale. They occur in different morphologies, show a large variability in their chemistry and are able to self-assemble into three-dimensional crystals. Intracuticular waxes function as the main transport barrier to reduce the loss of water and small molecules such as ions from inside of the cell, and also for reducing the uptake of liquids and molecules from the outside. Epicuticular waxes form the boundary layer for many interactions with the plant's environment, like wettability or spectral

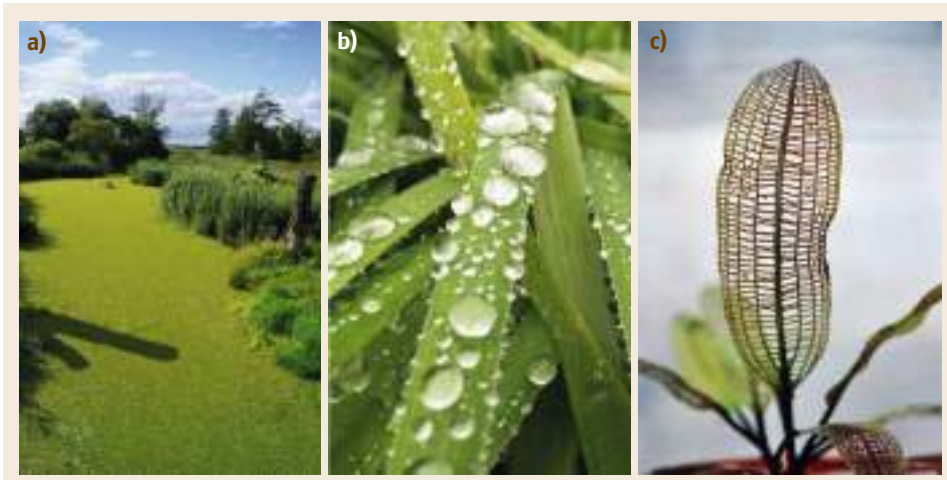


Fig. 36.5a–c Plant surfaces exposed to the air are very often superhydrophobic. Like the leaves **(a)** of the reed (*Phragmites autralis*) and the grassland in the background, or the surfaces of the floating fern (*Salvinia natans*) in the foreground-left in the Oder national park, Germany. Dew droplets in the early morning roll off of the superhydrophobic grass leaves **(b)**. Since grasslands alone form the largest terrestrial ecosystems (around 52.5 million km^2) of our planet we estimate there are at least about 250 million km^2 of superhydrophobic leaf surfaces, which equal about 50% of the Earth's total surface. But all plant roots are superhydrophilic – like the surfaces of water plants, illustrated in **(c)**: the Madagascar Laceleaf (*Aponogeton madagascariensis*) additionally exhibits a grid of a lattice-like network to reduce the flow resistance. Source: **(a)** kindly provided by Pierre Ibisch

reflection (Sect. 36.6). The next layer (Fig. 36.6) is the pectin layer. It connects the cuticle to the much thicker underlying cellulose wall, which is built by single cellulose fibrils. Pectin is not always formed as a layer, but in some species, especially during the early ontogeny of the cuticle, a layered structure has been shown by transmission electron microscopy (TEM). Additionally polysaccharides, not shown in this schematic, are integrated into the cellulose wall. The last layer shown is the plasma membrane, which separates the living, water-containing compartment cell from the outer nonliving part of the epidermis.

We focus on *superhydrophobic* and *superhydrophilic* surfaces, which are of particular importance for biomimetic applications (e.g., self-cleaning: lotus effect). *Superhydrophilicity* means that a droplet imposed on a surface *spreads* instantly and a contact angle cannot even be measured, e.g., in the leaves of *Ruellia* [36.11]. In contrast, on a *superhydrophobic* surface water remains as an almost globular droplet with a contact angle of more than 150°. The SEM micrographs presented were largely taken from our archive of almost 220 000 SEM micrographs at the University of Bonn, which has been built up as a result of over four decades of research on biological surfaces (compare [36.1]) by the first author and his collaborators.

Biomimetics and bionics (which we consider here as synonymous) are surmised to be modern scientific

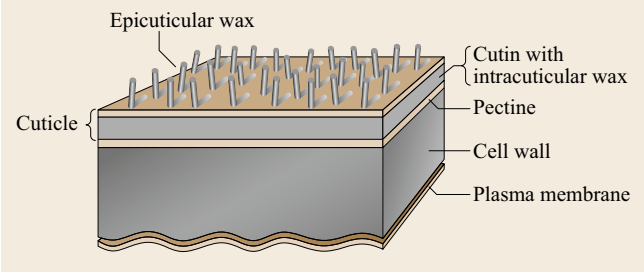


Fig. 36.6 A simplified model of the stratification of the outermost layers of a plant epidermal cell. The schematic shows the outermost wax layer in its most common form, as a composite of three-dimensional waxes with an underlying wax film. Below this layer is the cuticula, made of a cutin network and integrated waxes. The cuticula is connected with the underlying cellulose wall by a pectin layer. Below the cell wall, the plasma membrane is shown. This membrane separates the water-containing living part of the cells from the outermost nonliving outer cell wall and cuticle, as shown in the schematic

fields; despite the evidence that inspiration from living organisms is as old as mankind. The magnificent 17 000-year-old paleolithic paintings in the caves of Lasceaux are bio-inspired – like the Cadillac tail fins in the 1960s. Bio-inspiration in the sense of nonfunctional *biodecoration* is an inspiration for art and design into modern times [36.12, 13]. Early attempts to copy

mechanical functions were not particularly successful – Ovid's story of Daedalus and Icarus and Leonardo da Vinci's design of flying machines and other devices did not translate into technical success stories.

Historically, the dream of flying and the use of the strange phenomena *electricity* were the two fundamental forces for the foundation of what we call today *bionics* or *biomimetics*. The construction of an electric battery based on observations of the *Torpedo* fish (today we call it Electric Ray) by Allessandro Volta in 1800 was the first milestone [36.12] of bionics. And Icarus' dream was realized with the first well-documented, repeated, and successful flights by Otto Lilienthal from 1894 onwards; his design was based on his analysis of the flight of birds. The term *Biotechnik* (usually abbreviated in German as *Bionik*) for the new field was coined by *Raul Francé* in 1920 [36.14] and finally rediscovered under the influence of cybernetics under the names *bionics* [36.15] and *biomimetics* between 1960 and 1964; the misleading term *biomimicry* arose in 1982 (for a historical survey see [36.12]). Surfaces came surprisingly late into the focus of bionics. The Swiss engineer George de Mestral observed in 1941 the way that the burrs (*Arctium*) clung to his trousers and his dog, and, in 1958, he developed the bionic *hook and loop* fastener under the trade mark Velcro®. Starting with the discovery of hierarchically structured superhydrophobic lotus surfaces [36.2, 16, 17] and the drag-reducing shark skin [36.18, 19], biomimetic surface technologies (e.g., lotus-, shark-, gecko-, moth eye-, and Salvinia-effect) became a most important field [36.1, 7, 20, 21]. The publication of the *lotus effect* in 1997 [36.16] led to a change of paradigms in surface technologies and to an astonishing number of publications and the popularization of the term superhydrophobicity [36.1].

It is inevitable that the estimated 250 million km^2 of superhydrophobic surfaces were noticed in the early history of mankind. The remarkable abilities of self-cleaning lotus leaves are mentioned as early as the second century B.C. in the Sanskrit text of the Bhagavad Gita (e.g., book 5,10) and in Chinese poems. Even the name *lotus* (Greek λωτός) may refer to the Greek and Latin word elements for *washed*. In the first century A.D., Pliny the Elder introduces his remarkable *Naturalis Historia* with the modern statement *the world is a perfect globe* and later compares this phenomena with droplets: *everywhere one finds ... droplets forming ... perfect ... globes ... on dust or on hairy leaves ...*, thus anticipating hierarchical structuring as a precondition for superhydrophobicity. *J.W. von Goethe* (1820) [36.22] is more precise: *Raindrops remain on certain leaves as perfect spheres, which we attribute to an exsudated being ...* and continues in the same context: *The surfaces of a ripe plum ... appears blue*. This is an almost concise description of superhydrophobicity and the scattering and reflection of light caused by hierarchically structured epicuticular waxes, which in former times were called *exsudates*. Superhydrophobicity is so obvious that names of many plants refer to it: Pliny refers extensively to the superhydrophobic fern *Adiantum* (Greek: *not wetting*) with a contact angle larger than 160°. On the leaves of the well-known Lady's mantle (*Alchemilla*) dew droplets are preserved as perfect globes, which led to intense research by scientists today [36.23], but usually it is overlooked that the scientific name *Alchemilla* refers to the power attributed to droplets in medieval *Alchemia* (al-kimiya, alchemy). Even the vernacular names in most languages (*Lady´s mantle*, *Frauenmantel*, *manteau de notre Dame*) for *Alchemilla* refer to these protective superrepellent properties.

Biological role models provide an extraordinary diversity for innovative surface technologies, which are described for plants in the following sections primarily under the view of biologists.

36.2 Chemistry of Plant Surfaces

Here we consider only the surfaces of higher or vascular plants (Tracheophyta). Primary aquatic plants (from unicellular algae to seaweeds, see Sect. 36.8) lack a cuticle and have widely differing superhydrophilic surfaces. For biomimetic applications vascular plants are most important. In land or vascular plants waxes, from monomolecular layers to thick crusts or three-dimensional (3-D) crystals, form the boundary layer of the surface (Fig. 36.7). They are sometimes visible as a white or bluish coloration of leaves and fruits, as in wheat, cabbage, grapes, or plums. These colorations are caused by reflection of parts of the visible light spectrum by a dense coverage of 3-D wax structures. The fan palm *Copernicia prunifera*, the natural source of carnauba wax, has massive crusts of epicuticular wax, weighing several mg cm^{-2}. Carnauba wax is commercially used, e.g., for car and furniture polishes, medical products and candy. Even when there is not a bluish coloration visible, three-dimensional waxes are often present. In plants, three-dimensional waxes are responsible for several surface functions. Waxes are not only an essential part of the plant cuticle, but can also be found in fungi, lichens and animals. Waxes occur as filling material within the basic cutin network (intra-

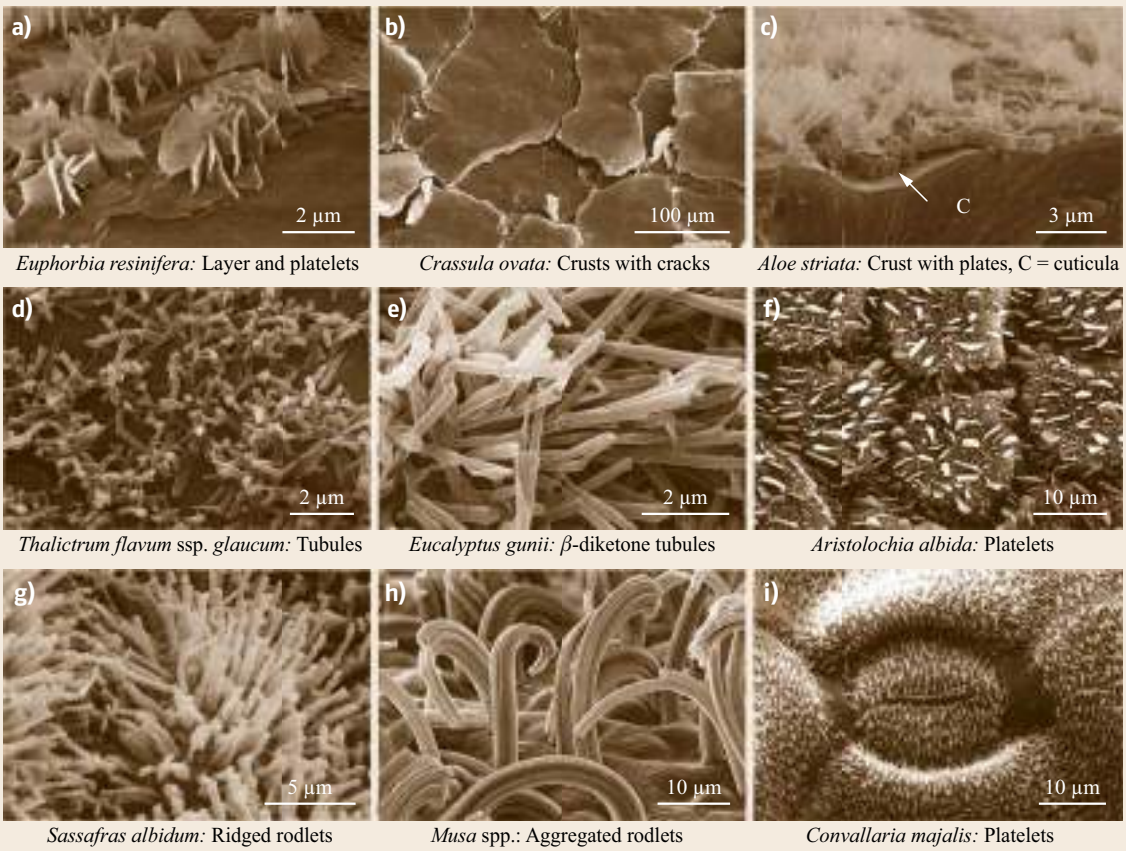


Fig. 36.7a–i SEM micrographs of epicuticular waxes. In **(a)** waxes on a leaf of *Euphorbia resinifera* have been partially removed to show the composite structure of a basal wax layer with three-dimensional wax platelets on it. In **(b)** a wax crust with fissures on a leaf of *Crassula ovata* is shown. A cross-section through the periclinal wall of *Aloe striata* in **(c)** shows the cuticle (indicated by C) and a wax layer (indicated by an *arrow*) with wax platelets on top. In **(d)** nonacosan-ol tubules on *Thalictrum flavum* spp. *glaucum* leaves and in **(e)** β-diketone wax tubules on *Eucalyptus gunnii* leaves are shown. In **(f)** wax platelets on *Aristolochis albida* leaf and in **(g)** transversely ridged rodlets on a leaf of *Sassafras albidum* are shown. In **(h)** longitudinally aggregated wax threads form large aggregated rodlets on the lower side of the leaves of *Musa* species (spp.). In *Convallaria majalis* leaves, shown in **(i)**, wax platelets are arranged in a pattern, similar to magnetic field lines, around the stomata. Thin wax films are not visible in SEM, but are present below and between the three-dimensional waxes shown here

cuticular), but are also found on top of the cuticle (epicuticular). The epicuticular waxes occur in widely differing morphologies, all of which are crystalline and thus self-assembling (Fig. 36.7) (survey in [36.24]). However, waxes of different plants, and also waxes of different parts of a plant, vary in their morphology and chemical composition – they are absent in roots. In general, plant waxes are mixtures of long-chain hydrocarbons and their derivatives, and in some species they also contain cyclic compounds. Because of the strong correlation between the wax crystal morphology and their chemical composition, some waxes, such as the nonacosan-10-ol tubules of the lotus leaves and many other plants, have been named after their main wax constitution [36.25].

36.2.1 Chemical Composition of Wax

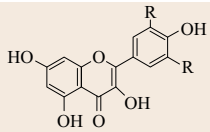
The term *wax* is used for a variety of biogenic products that contain fatty materials of various kinds. Well-known examples are beeswax, paraffin and carnauba wax from wax palms (*Copernicia prunifera*). Plant waxes are mixtures of aliphatic hydrocarbons and their derivatives, with carbon chain lengths between 20 and 40, and in the case of esters (two connected chains) about 60 atoms. Several reviews have addressed the

chemical composition of plant waxes [36.26–28]. The chemical composition of plant waxes is highly variable amongst different plant species, or within the organs of one species (e.g., upper or lower side of leaves) and even during organ development [36.29]. The main component classes are primary and secondary alcohols, ketones, fatty acids, and aldehydes. Alkanes are very common in plant waxes, but usually occur in low concentrations. Other compounds are more rarely found in plant waxes, but in those waxes where they occur, they may be the dominant compound. The most common wax compounds and their typical chain lengths are shown in Table 36.1. Examples of commonly found waxes and their major compounds are presented in Table 36.2. For some of those waxes, it has been shown that their dominant compounds crystallize in the same morphology as the complete wax mixture. Examples are the primary alcohols and the β-diketone waxes found on different parts of wheat plants [36.30]. However, an increasing number of publications report the discovery of new wax components and a long list of rare and uncommon ingredients, such as methyl-branched aliphatics [36.31]. Environmental factors, such as temperature or light intensity, influence the quantity of waxes and their chemical composition [36.32–34].

Many or even most plant *waxes* do not match the chemical definition of true waxes and they are usually complex mixtures of differing compounds. For example, triterpenoids are cyclic hydrocarbons, which occur in high concentrations in the epicuticular coatings of grapes (*Vitis vinifera*) [36.32]. Other plant waxes contain polymeric components such as polymerized aldehydes, which are only slightly soluble in chloroform [36.35, 36]. It should be noted that nearly all the existing data of the chemical composition of plant waxes is based on solvent-extracted waxes. These are mixtures of epicuticular and intracuticular waxes, which may be chemically different, as shown for the waxes of *Prunus laurocerasus* by *Jetter* and *Schäffer* [36.29] and by *Wen* et al. [36.37], for *Taxus baccata*. The development of more selective methods of wax sampling allows selective removal of the epicuticular waxes and their analysis separately from the intracuticular wax fractions [36.29, 38].

Epicuticular wax structures usually occur in the size range from 0.2 to 100 µm (Fig. 36.7); thus, the appropriate microscopic techniques for investigation of their morphology are SEM and low pressure- or environmental SEM. Several SEM investigations showed that most of the epicuticular waxes form three-dimensional structures, with great variations of their morphologies. Comprehensive overviews of the terminology and micromorphology of epicuticular waxes are given by *Barthlott* et al. [36.24], *Jeffree* [36.28], and in [36.40]. The comprehensive classification of *Barthlott* et al. [36.24], which we follow here, includes 23 different wax types. It is based on chemical and morphological features and also considers orientation

Table 36.1 The most common chemical compounds in plant waxes and their spectrum of chain length. Further examples of occurrence are given in Table 36.2

1. Aliphatic compounds **1.1 In waxes frequently existing, but mostly as minor compounds**		**Chain length**
Alkanes	CH_3–$(CH_2)n$–CH_3	Odd C_{19}–C_{37}
Primary alcohols[a]	CH_3–$(CH_2)n$–CH_2-OH	Even C_{12}–C_{36}
Esters	CH_3–$(CH2)n$–C0–0–$(CH_2)m$–CH_3	Even C_{30}–C_{60}
Fatty acids	CH_3–$(CH2)n$–COOH	Even C_{12}–C_{36}
Aldehydes	CH_3–$(CH2)n$–CHO	Even C_{14}–C_{34}
1.2 In waxes rarely existing, but if, than as major wax compounds		
Ketones e.g., palmitones	CH_3–$(CH_2)n$–CO–$(CH_2)m$–CH_3	Odd C_{25}–C_{33}
β-Diketones	CH_3–$(CH_2)n$–CO-CH_2-CO-$(CH_2)m$–CH_3	Odd C_{27}–C_{35}
Secondary alcohols (e.g., nonacosan-10-ol)	CH_3–$(CH_2)n$–CH_2OH–$(CH_2)m$–CH_3	Odd C_{21}–C_{33}
2. Cyclic compounds		
Flavonoids	e.g., Quercetin 	
Triterpene	e.g., β-Amyrin	

[a] Primary alcohols are common minor constitutions in waxes, but can occur as major compounds in the wax, e.g., of grasses, eucalypts, clover and other legumes, etc. (after [36.28])

Table 36.2 Common wax types in plant species and their major chemical compounds. With exception of the fruit surface of *Benincasa hispida*, data represent the waxes on the leaves of the species. All references for the chemical data are listed in [36.39] and examples of the wax types listed here are shown in Fig. 36.7

Wax type	Species	Dominating chemical compound(s)
Films	*Hedera helix*	Primary alcohols, aldehydes
Films	*Magnolia grandiflora*	Fatty acids C_{24}–C_{30}, primary alcohols C_{24}-C_{28}
Films	*Prunus laurocerasus*	Alkanes C_{29}, C_{31}
Crust	*Crassula ovata*	Aldehydes C_{30}, C_{32}, alkane C_{31}
Di-ketone tubules	*Eucalyptus globulus*	β-Diketones C_{33}
Di-ketone tubules	*Leymus arenarius*	β-Diketone C_{31}, hydroxy-β-diketone C_{31}
Nonacosanol tubules	*Ginkgo biloba*	Secondary alcohol C_{29}
Nonacosanol tubules	*Nelumbo nucifera*	Secondary alkane-diols C_{29}
Nonacosanol tubules	*Thalictrum flavum* ssp. *glaucum*	Secondary alcohol C_{29}
Nonacosanol tubules	*Tropaeolum majus*	Secondary alcohol C_{29}
Nonacosanol tubules	*Tulipa gesneriana*	Secondary alcohol C_{29}
Platelets	*Convallaria majalis*	Primary alcohol C_{26}, C_{28}, aldehydes
Platelets	*Euphorbia myrsinites*	Primary alcohol C_{26}, aldehydes
Platelets	*Galanthus nivalis*	Primary alcohol C_{26}
Platelets	*Iris germanica*	Primary alcohols C_{26}
Platelets	*Triticum aestivum*	Primary alcohol C_{28}
Transversely ridged rodlets	*Aristolochia tomentosa*	Ketones
Transversely ridged rodlets	*Gypsophila acutifolia*	Alkanes C_{31}
Transversely ridged rodlets	*Liriodendron chinense*	Ketones
Longitudinal ridged rodlets	*Benincasa hispida*	Triterpenol acetates

of single crystals on the surface and the orientation of the waxes to each other (pattern formation). In this classification, the wax morphologies include thin films and several three-dimensional structures such as crusts, platelets, filaments, rods and tubules, which have a hollow center. Morphological subtypes are, for example, entire and nonentire wax platelets. A further subclassification is based on the arrangement of the crystals, e.g., whether they are randomly distributed, in clusters, in parallel orientation, or in specific arrangements around stomata, as in the *Convallaria* type (Fig. 36.7h). The most common wax morphologies are introduced in the following section and shown in Table 36.2.

Probably all terrestrial plant surfaces are covered by thin (in the extreme monomolecular) wax films; the three-dimensional wax crystals appear on underlying wax films as shown in Fig. 36.7a for the waxes of *Euphorbia resinifera* and has been reported for several species [36.24, 41–44]. Wax films are often incorrectly referred to as *amorphous* [36.39]. On several plant surfaces wax films are limited to a few molecular layers that are hardly visible in SEM. By mechanical isolation of the epicuticular 3-D waxes, e.g., freezing in glycerol, the waxes can be removed from the cuticle, and transferred onto a smooth artificial substrate for microscopic investigations [36.38]. With this method, the remnants of the wax film can be detected, and the film thicknesses can be determined. Wax film formation has been investigated on a living plant surface by atomic force microscopy (AFM, Fig. 36.8) [36.43, 45]. Such investigations show that wax films are composed of several monomolecular layers, with thicknesses up to several hundred nanometers. In the following, these relatively thin wax films ($< 0.5\,\mu m$) are called two-dimensional (2-D) waxes, and the thicker wax layers ($0.5-1\,\mu m$)



Fig. 36.8 AFM experimental setup for long-term investigations of wax crystallization on a living plant surface. The tip of the leaf of a snowdrop (*Galanthus nivalis)* has been fixed on the specimen holder with a drop of two-compound glue. Existing waxes have been removed and the rebuilding (self-healing) of the wax was studied over several hours. Appropriate scan conditions for living plant surfaces are given in the text, and the method of wax removal is described in detail in [36.43]

and wax crusts (> 1 μm) are called three-dimensional (3-D) waxes. Wax crusts are often found in succulent plants, as on the leaves of *Crassula ovata*, shown in Fig. 36.7b. Such a multilayered assembly of waxes is detectable by a cross-section through the epidermis, as shown in Fig. 36.7c for *Aloe striata*.

Three-dimensional waxes occur in different morphologies. Most common are tubules, platelets, rodlets and longitudinally aggregated rodlets shown in Fig. 36.7d–i.

Wax tubules are hollow structures, which can be distinguished chemically and morphologically. The first type, called nonacosanol tubules, contains large amounts of asymmetrical secondary alcohols, predominantly nonacosan-10-ol and its homologues and to a certain degree also asymmetrical diols [36.25, 46, 47]. Nonacosan-10-ol is the most common *waxy* coating of all major vascular plant groups and was evolved with the conquest of land some 450 million years ago (a phylogenetic tree is provided by [36.1]). The nonacosanol tubules are usually 0.3−1.1 μm long and 0.1−0.2 μm wide. The second type of tubules contains high amounts of β-diketones, such as hentriacontan-14,16-dione [36.48]. This particular kind of wax tubule is characteristic for many grasses (Poaceae) but also occurs in various other groups [36.49]. Fig. 36.7e shows that the β-diketone tubules are two to five times longer than the nonacosanol tubules shown in Fig. 36.7d. Their length reaches from 2 to 5 μm, and diameters vary between 0.2 and 0.3 μm.

Platelets, as shown in Fig. 36.7f, are very common wax structures found in all major groups of plants. Following the terminology of *Barthlott* et al. [36.24], waxes are termed platelets when flat crystals are connected with their narrow side to the surface. Platelets can be further differentiated by their outline into, e.g., entire or undulated ones. Platelets vary considerably in shape, chemical composition and spatial pattern. For platelets, only limited information about the connection between morphology and chemical composition is available. In some species, wax platelets are dominated by high amounts of a single chemical compound, which can be primary alcohols, alkanes, aldehydes, esters, secondary alcohols or flavonoids [36.28]. In contrast to platelets, plates are polygonal crystalloids with distinct edges and are attached to the surface at varying angles.

The morphology of three-dimensional wax structures is not necessarily determined by the dominating chemical compound or compound class. One example of wax crystals determined by a minor component of a complex mixture is the transversely ridged rodlets, shown in Fig. 36.7g, which contain high amounts of hentriacontan-16-one (palmitone) [36.50]. Wax rodlets are massive sculptures that are irregular, polygonal, triangular or circular in their cross-sections. They have a distinct longitudinal axis, with a length : width ratio usually not exceeding 50 : 1. In addition, rodlets may have a variable diameter along the length of their axis. More complex structures are the longitudinally ridged rodlets, such as those found on banana leaves (*Musa* species), shown in Fig. 36.7h. These waxes consist exclusively of aliphatic compounds, with high amounts of wax esters and less of hydrocarbons, aldehydes, primary alcohols, and fatty acids. The origin of these wax aggregates is still not clear, and so far all attempts to recrystallize these wax types have failed. As a consequence of that, it is assumed that their origin is connected to structural properties of the underlying plant cuticle.

Brassica oleracea is known to have a very complex wax crystal morphology: several cultivars form several different wax types, and several different wax morphologies can occur on the same cell surface [36.32]. Why the different three-dimensional wax morphologies coexist on the surface of a single cell is unknown, as is whether these different morphologies are built up by phase separation of different compounds or if they are formed by the same compound.

The last example in Fig. 36.7i represents plant surfaces on which waxes are arranged in a specific pattern. Examples are parallel rows of longitudinally aligned platelets, with the orientation extending over several cells (e.g., in *Convallaria majalis*, Fig. 36.7i), or rosettes, in which the arrangements of platelets is more or less in radially assembled clusters. In particular, the parallel orientation of platelets on the leaves of several plant species leads to the question of how the orientation is controlled by the plant. It is assumed that the cutin network functions as a template for the growth of the three-dimensional wax crystals, but there is still a lack of information about the molecular structure of the cuticle, so this question is still unanswered.

Certain surface wax morphologies and their orientation patterns are characteristic for certain groups of plants; thus, patterns and the morphology of plant waxes have been used in plant systematics. *Barthlott* et al. [36.49] provide an overview of the existence of the most important wax types in plants, based on SEM analysis of at least 13 000 species, representing all major groups of vascular plants.

36.2.2 Chemical Heterogeneities

Surfaces of a particular plant species may exhibit chemical heterogeneities in the classical sense; the best example are the superhydrophilic pinning anchor cells on top of each superhydrophobic trichome of *Salvinia molesta* (Fig. 36.28): In a broader sense, all organism

have chemically heterogeneous surfaces: root surfaces differ dramatically from leaf surfaces. And within one leaf the upper side differs from the underside. In leaves of *Quercus robur*, contact angles range from 30° to 130° depending on the part of the leaf where wettability was determined [36.8].

The aquatic watermilfoil *Myriophyllum brasiliense* is – like all submersed water plants – superhydrophilic, so a contact angle of the leaves cannot be determined. However, as soon a flowering shoot approaches the water level wax crystals are generated and the new leaves outside of the water exhibit a contact angle of 162° like in a lotus leaf [36.8].

36.2.3 Crystallinity

All aliphatic plant surface waxes have a crystalline order. The classical definition for crystals implies a periodic structure in three dimensions, but with the increasing importance of liquid crystals and the detection of quasicrystals it has become necessary to extend the definition, so that certain less periodic and helical structures, as found for some waxes, were included [36.51].

The crystal structure of the epicuticular waxes can be examined by electron diffraction (ED), nuclear mass resonance (NMR) spectroscopy and x-ray powder diffraction (XRD). Electron diffraction with the transmission electron microscope provides the structure information of single wax crystals of less than 1 µm size, as shown in Fig. 36.9a,b, for a single wax platelet. However, even with a low-intensity imaging system, the crystal structure is rapidly destroyed by the electron beam intensity. Therefore, XRD is useful for determining the crystal symmetry, as well as providing information about different types of disorder. Very thin mono- or bimolecular layers of waxes, as shown in Fig. 36.9c,d, are of course not periodic in three dimensions, but form two-dimensional crystals at the molecular level. As mentioned before, in addition the planar wax structures, such as films and platelets, many natural plant waxes develop irregular three-dimensional morphologies, or structures such as threads and tubules with a large extension in one direction. These morphologically different waxes were found to occur in three different crystal structures. The majority of waxes exhibit an orthorhombic structure, which is most common for pure aliphatic compounds. Tubules containing mainly secondary alcohols show diffraction reflections of a triclinic phase, with a relatively large disorder, and β-diketone tubules show a hexagonal structure [36.39].

Self-assembly of waxes is an inherent result of the crystalline nature. That different wax morphologies on plant surfaces originate by self-assembly of the wax molecules has been shown by the recrystallization of waxes, which were isolated from plant surfaces [36.30, 43, 47, 52–54]. In these studies, most waxes recrystallized in their original morphology, as found on the plant surfaces.

Self-assembly processes resulting in nano- and microstructures are found in nature, as well as in engineering. They are the basis for highly efficient ways of structuring surfaces down to the molecular level. Self-assembly is a general process of structuring in which atoms, molecules, particles or other building units interact and self-organize to form well-defined structures. The processes of self-assembly in molecular systems are determined by five characteristics: the components, interactions, reversibility, environment and mass transport with agitation [36.55]. The most important driving forces are weak and noncovalent in-

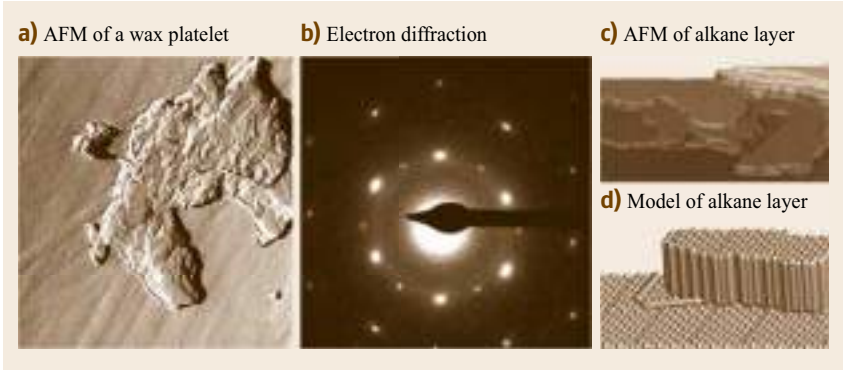


Fig. 36.9a–d The layered and crystalline structure of alkane waxes is demonstrated by an AFM map of a single wax platelet **(a)** and the corresponding electron diffraction pattern, shown in **(b)**. In **(a)** the steps visible on the crystal surface are caused by a perpendicular orientation of the molecules. Such steps can be monomolecular e.g., for alkanes, or in some waxes bilayers are formed by polar molecules of primary alcohols. The AFM map of recrystallized alkanes **(c)** and the model shown in **(d)** demonstrate the layered orthorhombic wax structure

termolecular interactions, such as Van der Waals and Coulomb interactions, hydrophobic interactions and hydrogen bonds. During self-assembly, their interactions start from a less-ordered state, e.g., dissolved waxes in a solution, to a final more-ordered state, i. e., a crystal [36.56, 57]. Environmental factors such as temperature, solvent and substrate might influence the self-assembly process, and in the case of waxes, their morphology.

The most suitable microscopy technique for studying the self-assembly process of waxes under environmental conditions is atomic force microscopy (AFM) because it combines sufficient resolving power to image nanostructures with the ability to work at STP (standard temperature and pressure) with living plant material (Fig. 36.8). Self-assembly of waxes has been studied directly on plant surfaces, as well as the recrystallization of waxes and single wax compounds on artificial surfaces. However, AFM is not suitable for all plant surfaces. Within a leaf surface large structures such as hairs with dimensions of several tens of micrometers can emerge out of the epidermis and pose a barrier against the surface scanning probe. Additionally, high-aspect-ratio structures caused by cell surface structures might cause artifacts in the resulting images. Species with smooth or slightly convex cell-surface sculptures are most appropriate for AFM investigations. The process of wax regeneration occurs over several hours; thus the loss of water from inside the plant has to be minimized to reduce the specimen drift by material shrinking during investigation. This precondition limits the range of specimens for AFMs with a small specimen chamber, because the sizes and shapes of the leaves must allow them to be mounted in the AFM without cutting them. An experimental setup where the complete plant is placed close to the AFM and a leaf is fixed on the AFM specimen holder is shown in Fig. 36.8. The leaf was fixed at its lower side to the specimen holder with a drop of a two-compound glue, and waxes on the upper leaf side were removed by embedding them into a drop of water-soluble glue. After hardening, the glue and the embedded waxes were removed from the leaf surface and the process of wax regeneration was studied. Temperature increase in long-term investigations, caused by the laser beam on top of the cantilever, induces expansion of the water in the leaf, resulting in a drift of the specimen. To minimize this, reflective cantilevers must be used, and the laser beam intensity should be reduced by integrating an attenuation filter above the cantilever [36.43]. However, the waxes themselves are fragile; thus appropriate scan conditions at scan sizes of 3–20 µm are tapping mode and scan rates of 0.7–2 Hz, encompassing 256 lines per image and a set point near the upper limit to minimize the interaction between tip and sample. Figure 36.10 shows the regeneration of a wax film on a leaf of snowdrop (*Galanthus nivalis*) by formation of a multilayered wax film and the growth of three-dimensional wax platelets. This and further investigations show that the growth of the three-dimensional wax crystals occurs by apical ac-

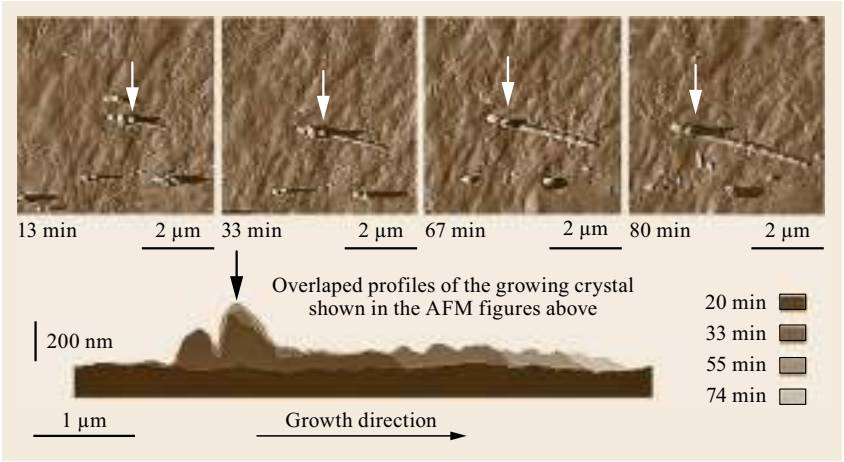


Fig. 36.10 AFM maps and a series of profile lines, taken from repeated scans during the crystal growth on a leaf of snowdrop (*Galanthus nivalis*). The first AFM map represents the wax regeneration within 13 min; the last map was taken after 80 min after wax removal. The *white arrows* mark the same position of the crystal as the *black arrow* marks in the profile figure. In the profile lines below the maps, outlines of the growing crystal have been overlapped to demonstrate that the extension is occurring at the distal end of the growing crystal and that at this time the growth in height is limited to a few nanometers. Outlines have been taken from four AFM scans: 20, 33, 55 and 74 min after the wax regeneration process began. The experimental setup is shown in Fig. 36.8

cumulation of new wax molecules on only one side of the crystal. The regeneration of the wax film results in a multilayered crystalline coverage on the plant cuticle. The time needed to regenerate waxes shows large variations depending on the species, with some species never regenerating the wax that was removed. In these plants, wax synthesis seems to be inactive when leaves are mature [36.58].

Alternatively, self-assembly of plant waxes can be studied by recrystallization of the waxes on artificial substrates (Fig. 36.11). Based on those studies, the formation of wax tubules and platelets has been described in detail. Wax platelets, characteristic for wheat leaves (*Triticum aestivum*), are constructed from the primary alcohol octacosan-1-ol [36.30]. Crystallization of the wax mixture isolated from the leaves and of pure octacosan-1-ol on different artificial substrates showed a substrate-dependent growth. On a nonpolar, crystalline substrate (highly ordered pyrolytic graphite, HOPG) platelets grow with a vertical orientation to the substrates, whereas on a polar surface, such as mica, crystals grow horizontally to the substrate surface. On amorphous polar glass only amorphous wax layers grow. This substrate dependence demonstrates epitaxial control of crystal growth depending on the orientation and order of the first layers of molecules adhering on the substrate surface. Octacosan-1-ol forms ordered bilayer structures on the substrate. In these, the first layers of molecules lie flat on nonpolar substrates, but stand upright (perpendicular) on crystalline polar surfaces. The grown platelet morphology results from an anisotropic crystal growth, caused by a faster parallel assembly of the molecules at the length side of already-existing molecules than at the ends of the molecules [36.59]. AFM micrographs in Fig. 36.12 and schematics of the molecule orientation demonstrate the differences of growth on polar and nonpolar substrates for octacosan-1-ol molecules. In both cases, flat crystals with different orientation grow. Crystals grown horizontal to the substrate surface are called plates (Fig. 36.12a,b) and those grown perpendicular to the substrate surface are termed platelets (Fig. 36.12c,d). The substrates on which the crystals grow influence the crystal morphology and their orientation. This fact can be used to create different kinds of nano- and micropatterns on technical surfaces [36.30, 54, 60, 61]. In summary, substrates can have a direct influence on the self-assembly processes of wax crystals, and can function as a template on the molecular level. In this case, the substrate organizes the assembly of the molecules in a specific spatial arrangement [36.62, 63]. Such a template effect was reported for wax platelets formed by primary alcohols [36.30]. On HOPG substrate, the spatial pattern of the re-

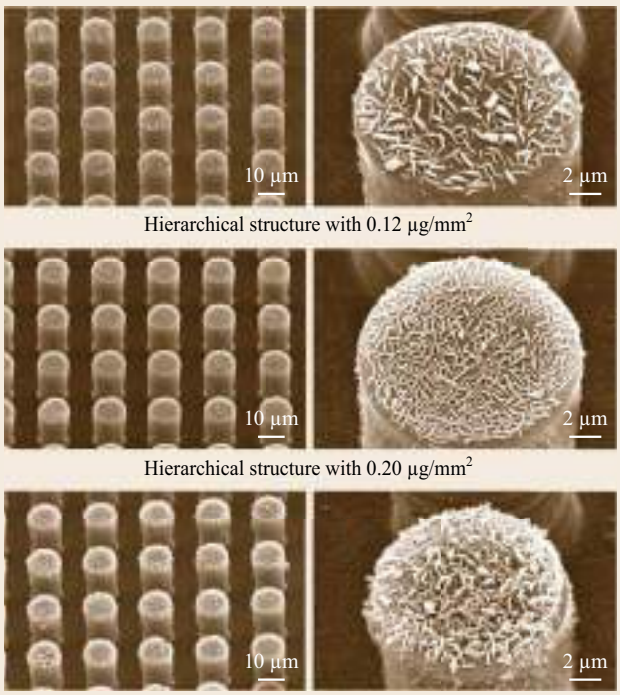


Fig. 36.11 Biomimetic superhydrophobic hierarchical structured technical surfaces. The silicon replica with pillars of 14 µm in diameter and 30 µm in height with 23 µm pitch is covered by self-assembled alkanes (hexatriacontane). From *top to bottom* an increase in crystal density is shown. Highest water repellence and lowest hysteresis has been found for the structures given in the *middle line*, where 20 µg cm^{-2} hexatriacontane was applied on the surfaces. These surfaces have been used for detailed study of wetting and adhesion (after [36.61])

assembled wax platelets strictly followed the hexagonal symmetry of the crystalline substrate. However, the cutin matrix of the cuticle, which acts as a substrate in plant surfaces, is assumed to be amorphous, and an epitaxial growth on an amorphous substrate seems paradoxical.

An example of wax crystals composed of more than one compound are the transversely ridged rodlets. These waxes can be recrystallized from the total wax mixture, but not from individual compounds such as alkanes or palmitones. For these waxes, it is assumed that their morphology is also formed by a self-assembly-based crystallization process, but the presence of minor amounts of other compounds is required as an additive for crystal growth [36.50].

The origin of wax tubules, shown schematically in Fig. 36.13a and in SEM micrographs in Fig. 36.13b, has been debated for a long time. Several observa-

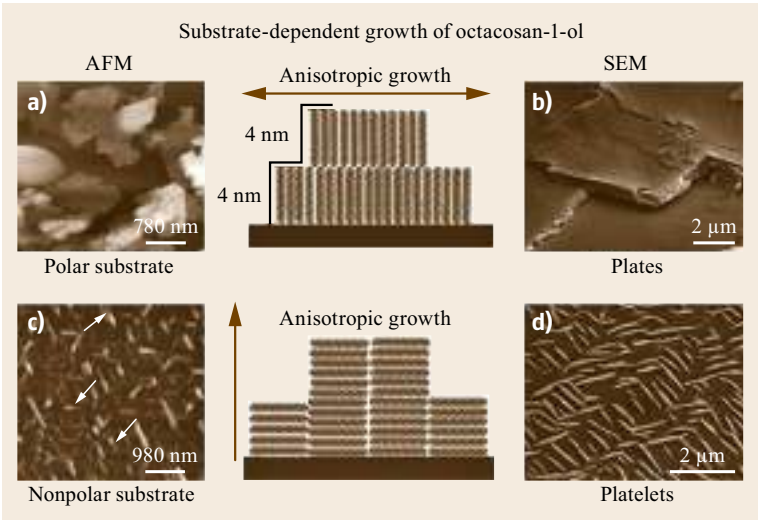


Fig. 36.12a–d AFM maps and schematics of the molecular orientation demonstrate the differences of growth on polar and nonpolar substrates for octacosan-1-ol molecules. AFM figures **(a,c)** show growing crystals, whereas the SEM figures **(b,d)** show the final crystal morphology. On both substrates flat crystals with different orientations were grown. Crystals grown parallel to the substrate surface, as shown in **(a,b)**, are called plates; crystals grown perpendicular to the substrate surface are called platelets **(c,d)**. The principle of anisotropic crystal growth is shown schematically for both preferred growth directions and described in the text

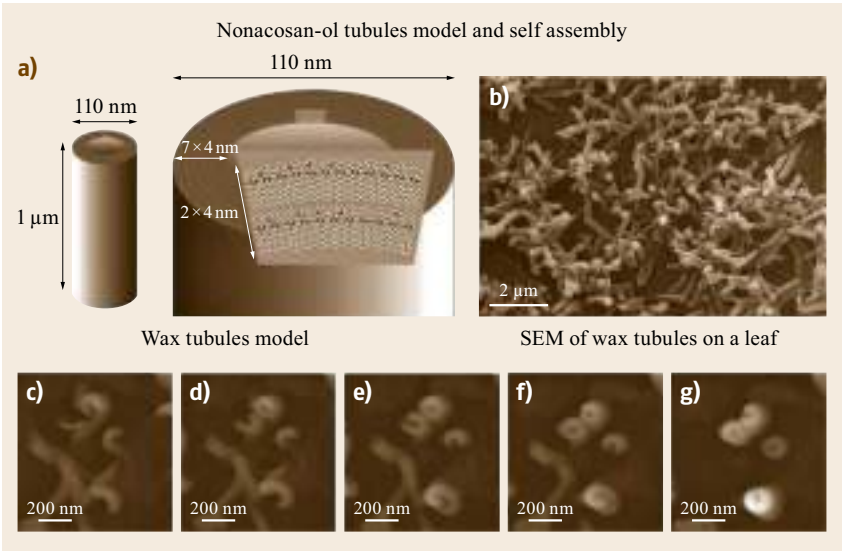


Fig. 36.13a–g A model and SEM micrograph of the molecular order of nonacosan-10-ol tubules and AFM analysis of their self-assembly. Based on SEM characterization, chemical analysis, single-compound crystallization and crystallographic data, a model of the nonacosanol tubules has been developed **(a)**. Original nonacosan-10-ol tubules are shown in the SEM micrograph **(b)** for *Thalictrum flavum* ssp. *glaucum* leaves. Consecutive AFM figures of tubule formation (nonacosan-10-ol wax from *Tropaeolum majus*) were made after applying a wax solution on HOPG. After 65 min the waxes mainly formed curved rodlets, which were arranged horizontal to the substrate **(c)**. The same area of the HOPG substrate shows that waxes start to form circles **(d–f)** and after 223 min the rodlets initially observed were dissolved and short tubules were formed **(g)**

tions, such as spiral lines on the surfaces of some nonacosanol tubules [36.53], led to the assumption that tubules arise from a twisting or folding of a platelet-like precursor form. Recrystallization experiments with nonacosan-10-ol waxes showed that these tubules grow perpendicular to the substrate surface when recrystallized on HOPG. This vertical orientation of the tubules allows a detailed study of the growth process by AFM and showed that the building of nonacosan-10-ol tubules from lotus (*Nelumbo nucifera*) and nasturtium (*Tropaeolum majus*) leaves is based on a continuous growth of a small circular precursor structure by supplementation of the wax on top of it [36.54]. The AFM micrographs shown in Fig. 36.13c–g, are con-

secutive AFM images of growing tubules, made during the tubule formation process. The terminal ends of growing tubules are asymmetric in height. This asymmetry seems to be caused by an accumulation of new wax molecules at edges found at the terminal end of the tubules and indicates a helical growth mechanism for the tubules. The pure nonacosan-10-ol alcohol, the dominating compound of wax tubules, can crystallize in different forms [36.42, 47, 53]. Here, *Jetter* and *Riederer* [36.47] show that a range of alkanediols, present in the waxes of many secondary alcohol tube-forming species, also have tube-forming capability.

Chemical analysis of the leaf waxes of lotus and nasturtium *(Tropaeolum majus)* show that waxes of both species are composed of a mixture of aliphatic compounds, with nonacosan-10-ol (a secondary alcohol) and nonacosandiols (an C_{29} alkane with two alcohol groups) as their main components [36.54]. These compounds have been separated from the rest of the wax compounds and used for recrystallization experiments. It could be shown with mixtures of nonacosan-10-ol and nonacosandiols components that a minimum amount of two percent of nonacosandiols supports tubule formation [36.25].

Analysis of wax chemistry, crystalline order, and their self-assembly has led to a better understanding of the molecular architecture of three-dimensional waxes [36.64]. Based on these data, a model of nonacosan-10-ol tubule structure has been developed, as shown in Fig. 36.13a. Here it is assumed that the lateral oxygen atoms at the side of the straight molecules hinder the formation of the normal, densely packed, orthorhombic structure and require additional space, causing a local disorder between the molecules and cause a spiral growth, leading to the tubule form.

Glands or glandular trichomes may produce very particular substances, and they can be found on approximately 30% of all vascular plants [36.65]. Multicellular glands include salt glands, nectaries, or the adhesive-secreting glands of some carnivorous plants [36.66]. Secretion and accumulation of toxic compounds at the plant surface allows direct contact with insects, pathogens and herbivores, and might therefore be an effective defense strategy [36.66]. The exudates of glands are, for example, terpenoids, nicotine, alkaloids or flavonoids. The exudates of some ferns and angiosperms, in particular several members of the *Primulaceae*, are composed of flavonoids [36.67, 68]. These flavonoid exudates or *farina* are morphologically similar to waxes, but are chemically distinct from plant waxes. Other glandular trichomes, such as the glands of the carnivorous plants of the genera *Drosera* (sundew) and *Pinguicula* (butterworts) secrete adhesives and enzymes to trap and digest small insects like mosquitoes and fruit flies.

Chemical heterogeneities are implied by the presence of other glands. The definition of this phenomenon depends on the scale: all biological surfaces show chemical heterogeneities, the most common case are leaves with a hydrophilic upper side and a superhydrophobic lower side. On a much smaller scale the trichomes of certain Salvinia species are most remarkable for hydrophilic islands within a superhydrophobic surface; the so-called Salvinia paradox (Fig. 36.28) [36.1, 69].

36.3 Structuring of Plant Surfaces: Hierarchical Architecture Between Nano- and Macrostructures

36.3.1 The Cuticle

Primary water plants (from unicellular algae to giant seaweeds) lack a cuticle; this particular polymer layer is restricted to higher plants (Sect. 36.8).

The cuticle covers leaves, flowers, stems, fruits and seeds and serves as a protective continuous layer covering the primary organs of all vascular plants and mosses. But in roots and secondary structures (e.g., bark) a cuticle is not present. The cuticle is a hydrophobic composite material, composed of a polymer called cutin and integrated and superimposed lipids called *waxes* (Sect. 36.2). The cuticle network is formed by cutin, a polyester-like biopolymer composed of hydroxyl and hydroxyepoxy fatty acids, and sometimes also by cutan, which is built by polymethylene chains. Nonlipid compounds of the cuticle are cellulose, pectin, phenols and proteins. Large differences in the chemical composition and microstructure of the cuticle have been found by comparing different species and different developmental stages. Chemical composition, microstructure and biosynthesis of the cuticle have been reviewed by several authors [36.28, 70–77].

36.3.2 Hierarchical Sculpturing

The cuticle and molecular wax films are the foundation of the surface. Additional levels of *hierarchical sculpturing* (or less-precise *structuring*) are formed by wax crystals (Fig. 36.7), the form of single cells (Fig. 36.18), composed sculptures like simple or multicellular hairs (Fig. 36.14), the form and curvature of whole organs

likes leaves (Fig. 36.27b,c) up to the gross morphological levels visible from larger distances like the riblets of the Titan arum flower (*Amorphophallus*, Fig. 36.1). The concept of hierarchical sculpturing of plant surfaces and a coherent terminology were introduced by [36.2] and used in many publications (e.g., [36.5, 78]). Based on the comprehensive survey by [36.1] we have revised this concept starting with a *flat* surface as first level, followed by a sequence of four or more superimposed *micro-architectural* hierarchical levels. The term *sculpture* (*architecture*) or sculptural seems more appropriate for three-dimensional features (discussed previously in [36.2]). The terms *structure* or *structural* also include chemistry and chemical heterogeneities, thus we prefer to use the term *sculpture* for the micro-architectural elements within the following four levels.

36.3.3 First Sculptural Level

This level comprises flat surfaces defined by their hydrophilic or hydrophobic chemistry on the scale of the resolution of a scanning electron microscope (SEM).

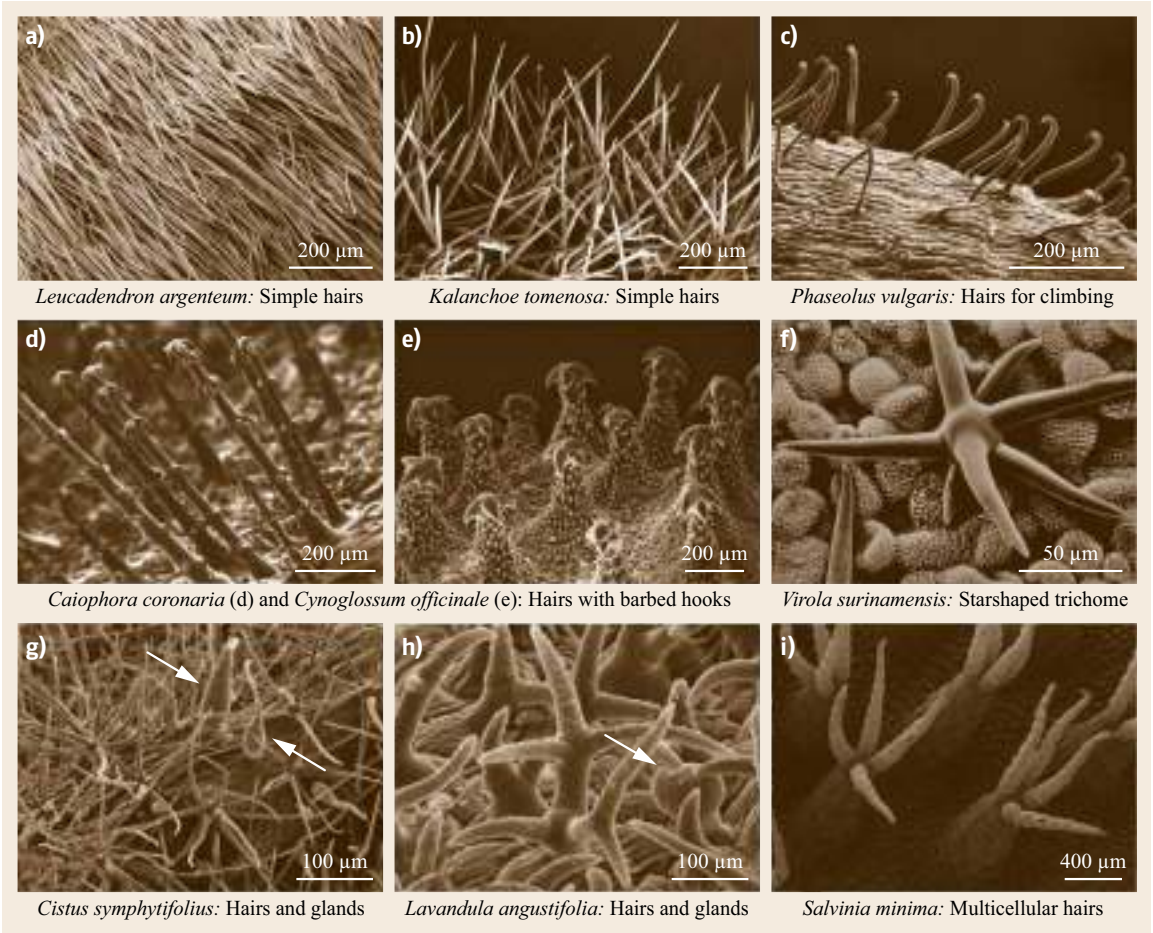


Fig. 36.14a–i SEM micrographs of hairs and glands on plant surfaces. A dense layer of straight, unbranched hairs almost orientated parallel to the leaf surface on *Leucadendron argenteum* **(a)**. The unbranched hairs of *Kalanchoe tomentosa* shown in **(b)** are orientated upright. The shoot surface of a bean shoot *Phaseolus vulgaris* with terminal hooks to facilitate climbing is shown in **(c)**. Single hairs on a leaf of *Caiophora coronaria* **(d)** and those on the seed surface of *Cynoglossum officinale* are characterized by terminal and lateral barbed hooks **(e)**. The star-like trichome on the leaf of *Virola surinamenis* **(f)** has a flat surface in contrast to the convex epidermal cells with epicuticular wax crystalloids. Simple branched star-like hairs and two morphological different glands (*arrows*) on the leaf of *Cistus symphytifolius* **(g)**, and multiple ramified hairs and short stalked glands (*arrow*) on a leaf of *Lavandula angustifolia* are shown in **(h)**. The four trichomes of *Salvinia minima* originate from a common base **(i)**, and in contrast to *S. molesta* (Fig. 36.28) the tips are free and do not show the eggbeater shape

Flat is a relative category that depends on the scale. Here, we limit the definition of flat to surfaces that feature structures of usually less than 10 nm in height. Flat surfaces are rarely found in plants and animals, but are seen for example in the leaves of rubber figs (*Ficus elastic*).

36.3.4 Second Sculptural Level

The second level comprises cell surfaces covered with structures between 50 nm and 20 µm. Structures at this level on plants are usually formed by epicuticular wax crystals (e.g., Fig. 36.7), which may exhibit a large spectrum of shapes like rodlets, platelets or tubules. They may exceed 200 µm in height (e.g., *Strelitzia*, *Copernicia*, *Benincasa*).

The main elements of the second sculptural level in plants are (a) *epicuticular wax crystals* (the diversity of waxes was discussed in the preceding sections on the chemistry of surfaces), (b) *cuticular folds* (Fig. 36.15), and (c) *subcuticular inclusions* (Fig. 36.16). Functionally, a superimposition of cuticular folds and wax crystals is not necessary: structuring on a specific level

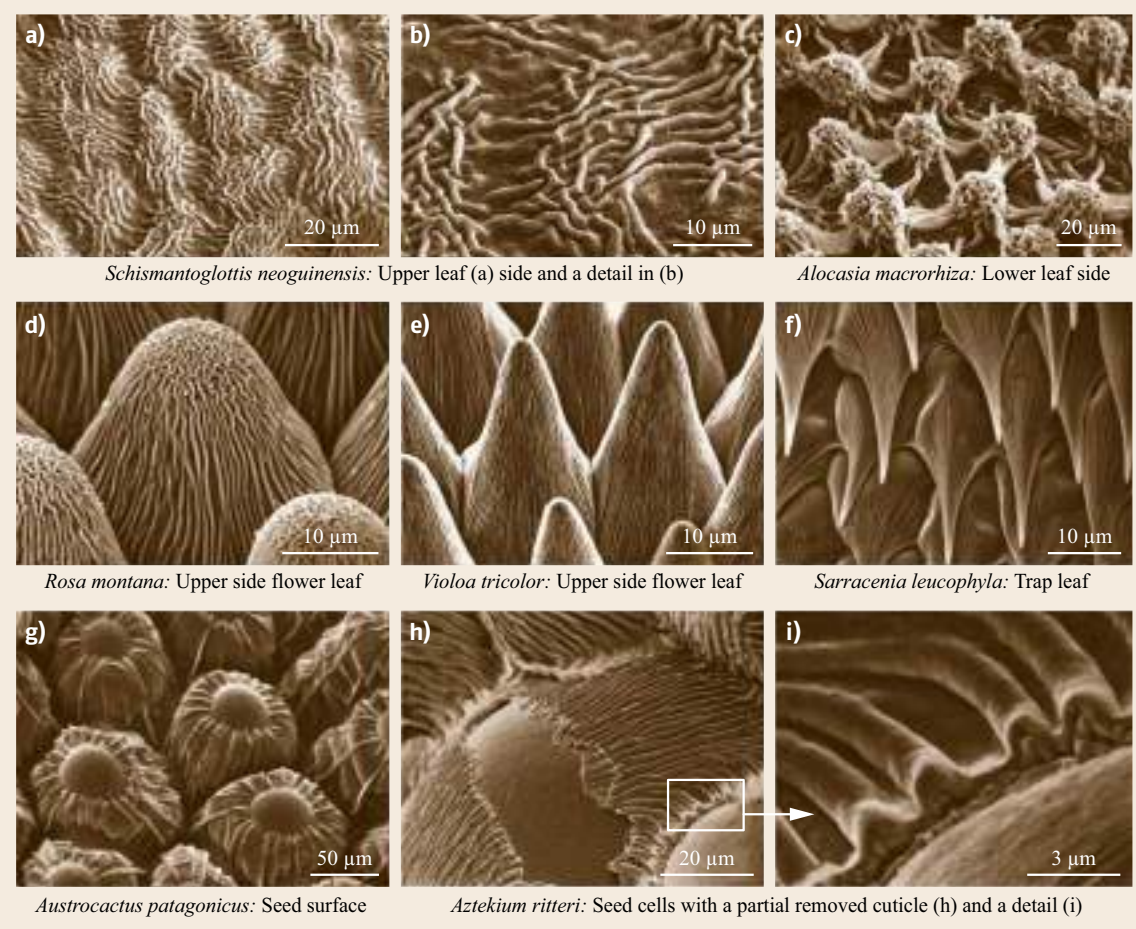


Fig. 36.15a–i SEM micrographs of cell surfaces with cuticular folding. In **(a,b)** the irregular cuticular folding on a leaf of *Schismatoglottis neoguinensis* is restricted to the central field of the cells. In **(c)** *Alocasia macrorhiza*, the cells are flat (tabular), with nodes like exposed central fields of the cells. The cells of a flower petal of *Rosa montana* with a rippled folded cuticle in the central field of the cells and parallel folding, running to the anticlinal walls of the cells, is shown in **(d)**. In **(e)** conical cells of the flower petals of *Viola tricolor* with parallel folding are shown. The cells of the inner side of a tube-like leaf of the carnivorous plant *Sarracenia leucophylla* is shown in **(f)**. These cells have a conical hair-papilla in the downward direction with a parallel cuticle folding, with larger distance at the base and denser arrangement at the cell tip. In **(g–i)** seed surfaces are shown. In *Austrocactus patagonicus* **(g)** the central field of the cells is unstructured, whereas a rough folding exists in the anticlinal fields. In *Aztekium ritteri* **(h,i)** a part of the cuticle has been removed to show the cuticle folding

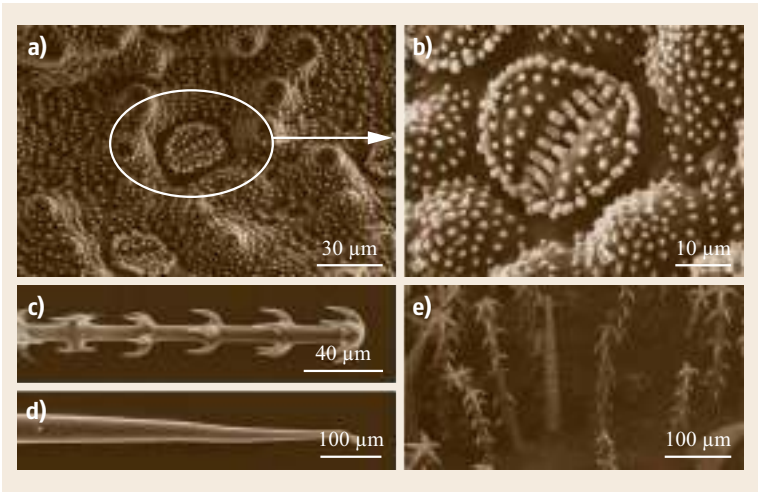


Fig. 36.16a–e Cell-surface structuring by subcuticular silicon dioxide and apatite insertions. SEM micrographs of the horsetail (*Equisetum arvense*) are shown in (a,b). (b) provides a detail of (a) and shows that the stomata and their surrounding cells have a micropattern of small enhanced spots, formed by subcuticular inserts of silicon oxide crystals. In (c–e) complex glochidiat hooked (e) and stinging (d) hairs are shown on the leaf surface in the flower nettle *Loasa*. The trichomes of Loasaceae are unique and mineralized by apatite. (c,e) from [36.79]

is performed by one group of elements, which seems to be a basic law [36.2]. The data gathered by investigating thousands of plant surfaces constitutes the rule that, for example, wax crystals and cuticular folds exclude each other. These frequently found morphological modifications of the outermost cell walls are known to influence the second hierarchical level. They are schematically shown in Fig. 36.17. In the first case, shown in Fig. 36.17a, the surface structure is induced by concavities of the cell wall, which lead to coves and folding of the surface. The second kind of structuring originates from subcuticular inserts of mineral crystals, such as silicon oxides (Fig. 36.17b). The third kind of surface structuring results from the folding of the cuticle itself (Fig. 36.17c). Additionally, on many plants,

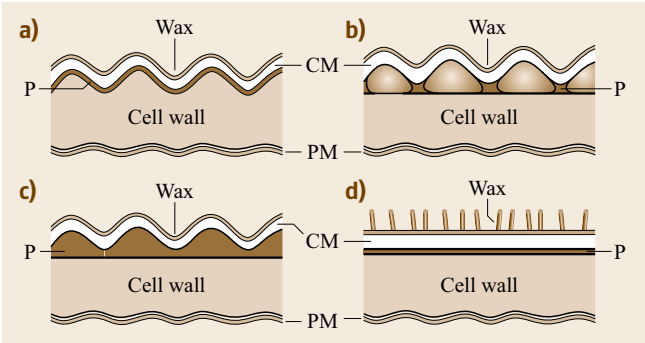


Fig. 36.17a–d Schematic of cross-sections through plant epidermis cells show different sources leading to microstructuring of cell surfaces. In (a) the surface profile is induced by coves of the underlying cell wall, in (b) by insertion of subcuticular minerals, in (c) by folding of the cuticle and in (d) by waxes, which are located on top of the cuticle (epicuticular wax). Wax = epicuticular waxes, CM = cuticular membrane, P = pectin, PM = plasma membrane (modified after [36.2])

waxes on top of the cuticle lead to surface structuring as shown in Fig. 36.17d. Waxes and their structural diversity have already been introduced; thus cuticular folds and subcuticular inserts are introduced in the next section.

Cuticular foldings have been described for nearly all epidermal surfaces of plants, but are frequently found in the leaves of flowers (petals), and on seed surfaces. They occur as folding or tubercular (verrucate) patterns, which originate due to the cuticle itself by an overproduction of cutin [36.80]. The pattern of cuticular folds can be categorized according to the thickness (width) of the folds, distances between the folds, and by their orientation [36.2]. Additionally, the pattern of folding within a single cell can be different in the central (inner area) and anticlinal field (outer area) of a cell. Fig. 36.15 shows different patterns of cuticle folding. On the leaves of *Schismatoglottis neoguinensis* (Fig. 36.15a,b) the folding is orderless and covers the central and anticlinal field of the cells. On the lower leaf side (adaxial) of *Allocasia macrorhiza*, shown in Fig. 36.15c, the cuticle forms node-like folding in the central part of each cell. The flower petals of *Rosa montana*, shown in Fig. 36.15d, have convex cells with a small central field with a rippled folded cuticle and parallel folds in the anticlinal field. The papilla cells of the flower petals of *Viola tricolor*, shown in Fig. 36.15e, have a parallel folding from the center to the anticlines of the cells. The cells inside the trap of the carnivorous plant *Sarracenia leucophylla*, shown in Fig. 36.15f, are hair-papillae, with a conical shape curved in a downward direction. On these, a parallel cuticle folding exists with larger distance at the base and a denser arrangement at the cell tip. The seed surface of *Austrocactus patagonicus*, shown in Fig. 36.15g, has cupular-formed cells with unstructured central fields and broad parallel folds

in the anticlinal fields. A high magnification SEM micrograph of the seed surface of *Aztecium ritteri*, shown in Fig. 36.15h,i, shows a partially removed cuticle and demonstrates that the origin of surface folding is caused by the cuticle itself.

Some microstructures on epidermis cells arise from subcuticular inserts of mineral crystals, as indicated in Fig. 36.17b. These subcuticular inserts can be solid crystals of silicon dioxide, as shown in Fig. 36.16a,b for tin plant or horsetail (*Equisetum*) plants. Calcium oxalate crystals are also frequently found in plants, and verification of silicon or calcium presence can be made simply by energy dispersive x-ray (EDX) analysis, included in an SEM. Silicon (Si) is a bioactive element associated with beneficial effects on mechanical and physiological properties of plants. It is a common element found in plants and occurs as monosilic acid or in the polymerized form as phytoliths ($SiO_2 \times nH_2O$) [36.81]. In plants, silica tends to crystallize in the form of silica in cell walls, cell lumina, at intercellular spaces and in the subcuticular layer [36.82]. Recently [36.79] investigated calcium apatite, a material that plays a crucial role in animals, in the complex trichomes of the family Loasaceae (Fig. 36.16c–e). Calcium oxalate crystals have been reported for more than 250 plant families [36.83], where they are deposed within the living tissue.

36.3.5 Third Sculptural Level

The third level contains unicellular (multicellular in certain hair types in plants) structures usually caused by particular shapes of the outer cell wall that may vary from convex to papillose cells and ultimately to hairs, which may be unicellular or multicellular (for a terminology see [36.2]). Dimensions range from about 2 µm to several centimeters, i. e., in trichomes (hairs). Structures of the second level may be superimposed onto structures of the third level (e.g., Fig. 36.22). To understand this level often a thorough microscopic analysis is essential, and the description and terminology for this diversity is complex.

The Outlines of Cells

The description of plant micro- and nanostructures requires the use of some basic uniform terms, for example to describe the outline of a single epidermis cell. Several variations are known and introduced in detail by *Barthlott* and *Ehler* [36.2], *Barthlott* [36.78] and *Koch* et al. [36.84, 85]. In the following, a brief introduction is given.

The boundary walls between two adjacent epidermal cells are called anticlinal walls, whereas the outer wall forming the cell surface is called the periclinal wall. The primary sculpture of a single cell encompasses the outline, including the shape and relief of the anticlines and curvature of the outer periclinal wall. There are two basic forms of cells, the tetragonal and polygonal form, both of which can have a uniform length of their sides or be elongated. Additionally, the course of the anticlinal walls can be straight or uneven. It is assumed that the outline of anticlines has an influence on the mechanical stability of the epidermis tissue, but experimental evidence for this hypothesis is not available. The cell sculptures or curvature of the outer epidermis wall (periclinal wall) can be tabular (flat), convex (arced to the outside) or concave (arced to the inside), and have a large influence on the surface roughness at the micrometer scale. Additionally, only the central area of a cell can form a convex outgrowth and form a papilla or hair-like structure. The convex cell type is the most common cell type of epidermal surfaces, often found on flower leaves, stems and leaves [36.86]. These cell morphologies originate by expansion of the outer side (periclinal wall) of the epidermis cells. They can be divided into several subtypes, depending on the outline of the epidermis cells and their aspect ratio (width to height), which determines their designation. In Fig. 36.18 a schematic of different convex cell outlines and their designations is given. The terminology is based on the cell outline and aspect ratios ($\beta =$ width/height) of the cells and includes: convex ($\beta \geq 3/1$), hemispherical ($\beta \approx 2/1$), cupola ($\beta < 3/2$), conical ($\beta > 3/2$), papilla ($\beta < 3/2$ and $> 1/2$), hair-papilla ($\beta < 1/3$ and $> 1/6$), and hairs ($\beta < 1/7$). In these, hairs are built by the outgrowth of a single surface cell. Hairs are often named trichome (Greek: trichoma).

The leaf surfaces of *Leucadendron argenteum* and *Kalanchoe tomentosa*, shown in the SEM micrographs in Fig. 36.14a,b, are two representative surfaces with hairs. Hairs can decrease, but also increase the loss of water and influence the wettability of the surfaces [36.87]. The wide spectrum of functions of plant hairs has been reviewed by *Wagner* et al. [36.66], and more recently by *Martin* and *Glover* [36.86]. With respect to their functions, it is important to notice that hairs can be glandular or nonglandular (nonsecreting), dead or living, and hairs can also be built up by several cells (multicellular), which are introduced later. Unicellular trichomes can be found on the aerial surfaces of most flowering plants (angiosperms), some conifers (gymnosperms) and on some mosses (bryophytes) [36.66]. Many plants of dry habitats show a dense cover of dead, air-filled hairs to reflect the visible light, which makes the surfaces appear white. The structures of hairs are often more complex; thus, the definition based on the aspect ratio fits well only

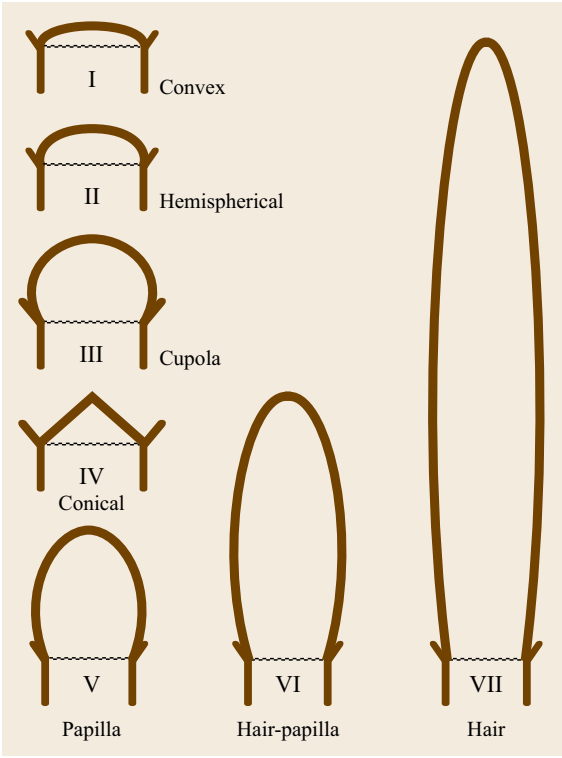


Fig. 36.18 A schematic of different convex cell outlines and their aspect ratios (β = width/height) of the convex cell types and their terminology: convex ($\beta > 3/1$), hemispherical ($\beta \approx 2/1$), cupola ($\beta < 3/2$), conical ($\beta > 3/2$), papilla ($\beta < 3/2$ and $> 1/2$), hair-papilla ($\beta < 1/3$ and $> 1/6$), and hair ($\beta < 1/7$)

for simple, undivided hairs. On the shoots of common beans (*Phaseolus vulgaris*), hairs form hooks to get better adhesion for climbing (Fig. 36.14c) and in *Caiophora coronaria* (Fig. 36.14d) and *Cynoglossum officinale* (Fig. 36.14e) the hairs have lateral barbed hooks. The stellate hairs of *Virola surinamenis* differ from the other epidermal cells, covered by a dense layer of wax crystals, by having completely smooth surfaces (Fig. 36.14f). Further trichomes are the simple or double-branched hairs and secretion glands on the leaves of *Cistus symphytifolius* and *Lavandula angustifolia* as shown in Fig. 36.14g,h. These complex hair structures require a more differentiated description than the aspect ratio used for simple hairs [36.88, 89]. The sizes and morphologies of trichomes are often species specific, making some trichomes useful as morphological features in plant systematics [36.89]. Deformation induced by water loss of dead desiccated cells can leads to concave cell morphologies and other complex modifications (Fig. 36.3). This is characteristic for seed coats and can result in most complex hierarchical sculpturing formed in cacti [36.90] or orchids [36.91].

36.3.6 Fourth Sculptural Level

The final level is made up of multicellular structures, caused by specific arrangement patterns of several epidermal cells. There is a wide variety of possibilities for this group of structuring. Multicellular hairs are common in all groups of vascular plants, apart from conifers.

Particularly interesting forms occur in the floating ferns of the genus *Salvinia*. Within this genus, morphologically different kinds of water-repellent (superhydrophobic) hairs exist [36.92], which in some species (*S. auriculata, S. molesta*) show functionally important chemical heterogeneities (Sect. 36.2) by utilizing hydrophilic anchor cells at the tip of the trichomes to stabilize the air–water interfaces [36.1, 69]. Four different hair types have been described for the genus *Salvinia* [36.92]. Based on these morphological types, the genus Salvinia is divided into four groups, each with several species. The Cucullata-type is characterized by solitary and slightly bent trichomes and occurs in *S. cucullata* and *S. hastate*. The Oblongifolia-type forms groups of two trichomes, which bend in the same direction and sit on an emergence. This type occurs in *S. oblongifolia*. The Natans-type, shown in Fig. 36.14i, has four trichome branches, each elevated on a large multicellular base and in total has a height of up to 1300 µm. The heights of the trichome groups decrease towards the leaf margins. This type occurs in *S. natans* and *S. minima*. In the Molesta-type (Fig. 36.23c) four trichome branches are grouped together, connected with each other by their terminal cells and sitting on a large emergence. The heights of these trichomes reach up to 2200 µm in *S. molesta*, but also decrease towards the leaf margins. This trichome-type is characteristic for, e.g., *S. molesta* and *S. biloba*. In all these species, the epidermis is covered with small three-dimensional waxes in the form of transversely ridged rodlets. The development of these complex structures has been studied in *S. biloba*, by *Barthlott* et al. [36.92]. In an early stage of leaf development, the hair formation starts with a grouping of four cells. During the ontogeny of the leaf, four branches develop from these initial cells and form a crown-like structure, in which the single branches are connected with each other. Later, the base grows by cell division and cell expansion to develop a large base below the crown structure.

36.4 Physical Basis of Surface Wetting

Wetting is the fundamental process of liquid interaction at solid–gaseous interfaces. It describes how a liquid comes into contact with a solid surface. The basics of surface wetting are summarized here; extensive literature on the topic exists [36.93–101].

36.4.1 Static Wetting Processes

A droplet on a solid surface wets the surface depending on the chemical properties of the phases, as well as on the surface structure. A parameter to quantify this wettability is the contact angle (CA). A high contact angle describes surfaces on which a water droplet forms a spherical shape and the contact area between the liquid and the solid is low. Contact angle measurement is the main method for the characterization of the hydrophobicity or hydrophilicity of surfaces. The wetting behavior of solid surfaces can be divided into four classes, defined by their CA, as well as CA-hysteresis for the case of superhydrophobic surfaces (Fig. 36.19). On superhydrophilic surfaces, where a fluid will spread and cover a larger area of the surface, the CA ranges from 0° to < 10°. Surfaces with a CA between > 10° and < 90° are termed hydrophilic surfaces. Unwettable surfaces have high contact angles, meaning the liquid on the surface forms a semispherical or spherical droplet. Surfaces on which the CA is > 90° and < 150° are hydrophobic surfaces. A superhydrophobic surface is defined as one that has a static CA of > 150°, and if those superhydrophobic surfaces have a low hysteresis or a low tilting angle (TA) of less than 10° they are superhydrophobic and can have self-cleaning properties. This definition of superhydrophobic surfaces has been used in many reviews [36.98, 102–106] and represents the most-used classification of surface wettability. The basis for studying the wettability of smooth surfaces is given by Young's equation [36.107]. The CA of a liquid on a surface depends on the surface tension (molecular forces) of the specific liquid, solid surface and the surrounding gas. Thus, wetting depends on the ratio between the energy necessary for the enlargement of the surface and the gain of energy due to adsorption (Fig. 36.19) [36.93, 108, 109]. The basis for studying equilibrium wetting on rough surfaces was established many years ago by *Wenzel* [36.110] and *Cassie* and *Baxter* [36.111] and [36.112]. The Wenzel equation expresses a general amplification of the wettability induced by roughness and applies to a CA where droplets are in equilibrium, but not to advancing and receding angles of a droplet on a rough solid surface that give rise to contact-angle hysteresis (CAH). Hysteresis is responsible for the sticking of liquids on a surface, and is the difference of an advancing and receding angle of a moving droplet ($CAH = CA_{adv} - CA_{rec}$, Fig. 36.19). If a droplet moves over a solid surface, the CA at the front of the droplet (advancing CA) is greater than that at the back of the droplet (receding CA). However, if the droplet rolls with little resistance, the contact angle hysteresis is small. Advancing and receding contact angles can also be determined when additional liquid is added to a sessile drop and the contact line advances; if liquid is removed from the drop, the CA decreases to a receding value before the contact retreats.

On water-repellent surfaces, a droplet applied starts to roll off the surface when it is tilted to a specific angle. This tilt angle (TA) is simply defined as the tilting angle of a surface on which an applied drop of water starts to move. Low TA (< 10°) is characteristic for superhydrophobic and self-cleaning surfaces. If the droplet is in Cassie–Baxter stage [36.111], with air trapped between the surface and the applied water droplet, the real contact between the droplet and the surface is very small

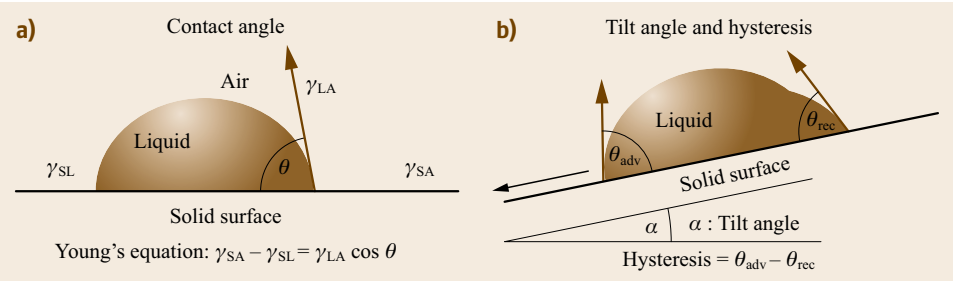


Fig. 36.19a,b The schematic shows the wetting of a solid surface. **(a)** γ_{LA}, γ_{LS} and γ_{SA} are the interfacial tensions at the boundaries between liquid (L), solid (S), and air (A), which determine the CA of an applied water droplet and is described by Young's equation. **(b)** The hysteresis of a water droplet on a tilted surface represents the adhesion of the liquid on the surface and can be determined by measuring the tilting angle or the advanced and receding angle of a water droplet

compared to wettable surfaces, on which an applied drop of water tends to spread, and contact angle is low. In intermediate wetting stages [36.113] high contact angles correspond with an increased contact between the surface and the water droplet applied. This phenomenon has also been described as the petal effect [36.109, 114] because of its occurrence in the flower leaves (petals) of some roses.

36.4.2 Dynamic Interactions

Dynamic interactions include bouncing, splashing and spreading of droplets. The impact behavior of falling drops has been subject to studies for several liquids, such as water, ethanol, emulsions etc. and various types of structured and unstructured, solid and liquid surfaces [36.115, 116]. In addition to the physiochemical surface properties and structures, other factors determine the impact behavior of falling drops: their velocity, size, surface tension and viscosity [36.117–119]. Falling drops can impact without any breakup or impact with a splash or splashing, in which the impacting drop releases smaller droplets flying away from the point of impact. Roughly two different types of splashing on incompliant surfaces are described, the *prompt splash* and the *corona splash*. The *corona splash* consists of relatively large droplets and is the outcome of the breaking up of fingers developing at a flattening drop's rim. In *prompt splash* very small and fast droplets are generated when the advancing lamella of a spreading drop is disturbed by rough surface structures higher than a specific fraction of the lamella, causing the lamella to rupture locally and release small splash droplets [36.120].

Dynamic interactions between water droplets and the surfaces are well researched in the different types of splash phenomena described e.g., in [36.115, 120–124]. More recently the contact times between droplets and surfaces came into focus: the time span a drop impacting and bouncing on a hydrophobic surface is in contact with the surface. The contact time as well as the shape of the rebound droplet strongly depends on the structure and the chemistry of the surface. A reduction of this contact time is of advantage in several applications, e.g., anti-icing, self-cleaning, or spray cooling [36.125]. The contact time might be important in biological surfaces.

Dynamic wetting processes and their control may play an underestimated role in the interaction between plant surfaces and rain, the maintenance of air layers underwater, or e.g., in pesticide applications.

36.5 Superhydrophilic and Superhydrophobic Plant Surfaces

The wettability of plant surfaces plays a crucial role in their interaction with the environment. A physical constraint of water plants is that they must be hydrophilic, whereas in land plants the whole range from superhydrophilic to superhydrophobic occurs (Fig. 36.20).

Hydrophilic and superhydrophilic surfaces (contact angle – if measurable – between 0° and $< 90°$) are known from all aquatic plants and many surfaces of land plants that usually have a papillate cell morphology and cuticular folds, but also from leaves with flat, tabular cells. Surfaces covered by wax may become hydrophilic with the erosion of theses layers (e.g., *Carnegiea*, Fig. 36.1a), a very common process in many plants (see below). Modified wax layers may even become superhydrophilic like in the Atacama desert cactus *Copiapoa cinera* adapted to fog harvesting. Hydrophilic behavior in flower petals of the daisy family (Asteraceae) and their polymer replica was analyzed in detail by *Koch* et al. [36.126] and provides information for biomimetic applications (Sect. 36.7).

Superhydrophobicity was discovered in terrestrial plants, and is one of the main and most obvious characteristics of most vascular nonaquatic plants [36.1]. Superhydrophobic surfaces cover hundreds of millions of square kilometers on our planet's surface (Sect. 36.1). One precondition is always a hydrophobic chemistry of the surface and hierarchical sculpturing on two or more levels [36.1, 2, 8, 20] – these are the only two essentials for superhydrophobicity in organisms. In plants it is generated – with only a few exceptions – by 3-D wax crystalloids of the morphological and chemical diversity previously described and illustrated (Fig. 36.7). Superhydrophobicity in this cases is caused merely by a sculpturing on the second level. For reasons of stability under dynamic conditions and to minimize mechanical damage an additional third hierarchical level (usually convex cell surfaces) is essential – the best example are the lotus leaves (Fig. 36.21) with contact angles $> 160°$ and a tilting angle of $< 4°$ [36.16] (survey of recent literature in [36.1]). The various functions of superhydrophobicity are discussed in Sect. 36.6. However, there is a phenomenon that *hairy* surfaces with cellular trichomes only covered by nonstructured molecular films also commonly cause this effect. The classic examples are the superhydrophobic leaves of the Lady's mantle (*Alchemilla)* analyzed in detail by *Otten* and *Herminghaus* [36.23]. This is so obvious and often observed, that the names

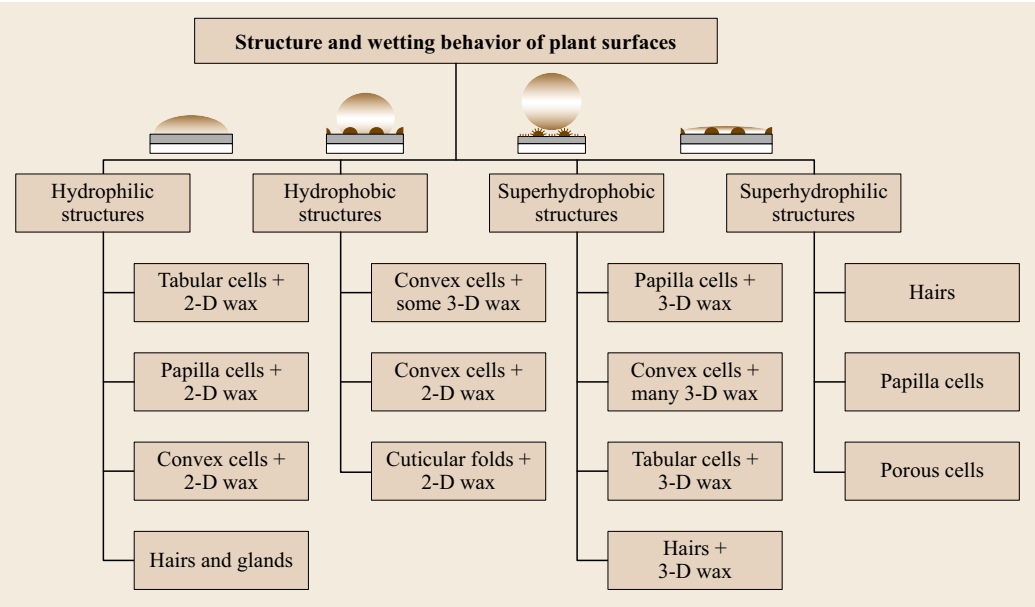
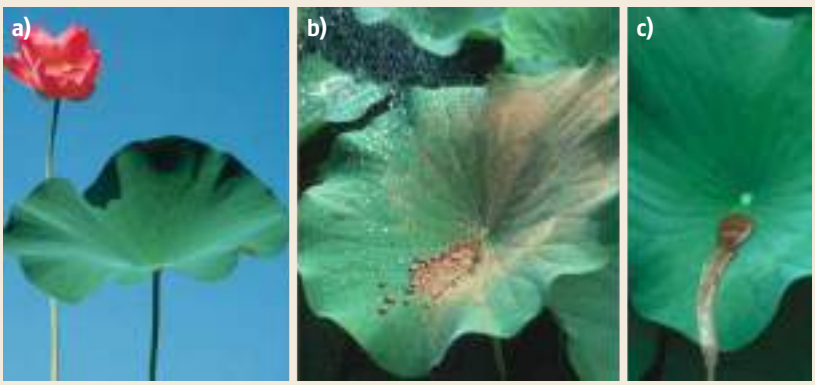
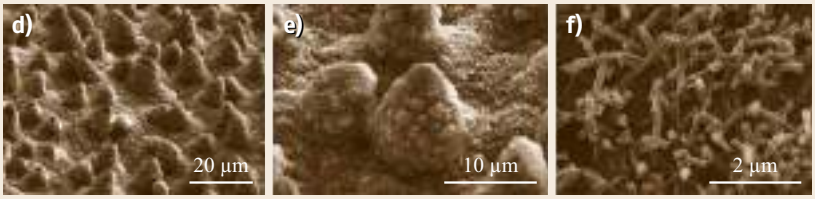


Fig. 36.20 Four groups of plant surface wettability and the possible surface structures and structure combinations. The drawings used for the four groups are correlated to specific contact angles. Both the hydrophobic and superhydrophobic surfaces can be built by convex cells with three-dimensional waxes on it, but only a dense layer of wax crystals results in superhydrophobicity



The lotus plant (*Nelumbo nucifera*) (a): Removement of dirt particles by water (b, c)



SEM micrographes of the lotus leaf (d) show the papillose cells (e) and wax tabules (f) on it

Fig. 36.21a–f Superhydrophobic and self-cleaning surface of lotus (*Nelumbo nucifera*). A flowering plant of lotus **(a)**, a lotus leaf contaminated with clay **(b)** and removal of the adhering particles by water **(c)**. The SEM micrographs **(d–f)** show the lotus leaf surface in different magnifications: **(d)** shows the papillate epidermis, **(e)** single-cell papilla and **(f)** the epicuticular nonacosan-10-ol tubules on the cell surface

Lady's mantle as well as *Alchemilla* refer to the superhydrophobic properties. The *Alchemilla-type* of superhydrophobicity also occurs in the feathers of many birds.

A list of extremely superhydrophobic plant surfaces was provided by *Neinhuis* and *Barthlott* [36.8]. Amongst the extremes – all measured with the same equipment under the same conditions – are plants

from most different relationships. Examples are: Indian cress (160° *Tropaeolum*), grasses (161° *Elymus*), Ginkgo trees (161° *Gingko*), lotus (162° *Nelumbo*, Fig. 36.22c), Californian poppy (162° *Eschscholzia*), Aroids (164° *Colocasia*, Fig. 36.22a), and *Euphorbia myrsinites* (162°, Fig. 36.22b). Extremely high contact angles caused by parallel oriented wax platelets are exhibited by some monocotyledons of the family *Alstroemeriacea* (*Peruvian lillies*); the extreme was measured in a climbing *Bomarea* reaching a static contact angle of 169° [36.127].

Superhydrophobicity is usually not very persistent in plant surfaces. Leaves, with a limited life span, do usually not last longer than one year. Persistent leathery leaves (like in many Mediterranean and tropical climates) are usually hydrophilic – but they may start as hydrophobic in their earliest developmental stages. These dynamics can be measured in the persistent leaves of *Welwitschia mirabilis*, a single leaf growing over two centuries from its base. They start superhydrophobic exhibiting tubular nonacosan crystals, which eroded after the first year causing the leaf to become wettable [36.1]. Plums are covered by a mechanically delicate whitish wax cover – touching it with a finger is enough to destroy the structure and thus the *glaucous* color. An extreme example is provided by the chemically complex brilliant white waxy coating of the succulent desert plants of the genus *Duddleya* (e.g., *D. brittonii*): it only takes a rain drop to destroy the unstable coating [36.68]. *Duddleya* is not self-healing, in contrast to other plants, were a wax cover destroyed may regenerate within hours.

Water plants are superhydrophilic, but when they rise above the water level they exhibit their old evo-

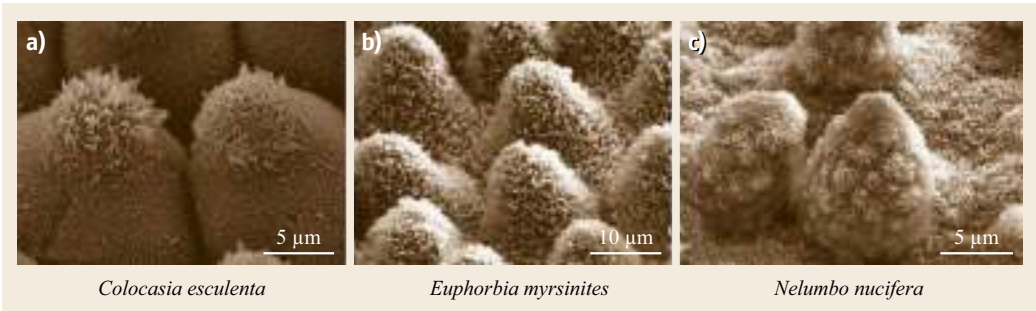


Fig. 36.22a–c Extremely superhydrophobic leaf surfaces of **(a)** *Colocasia esculenta*, (contact angle 164°) *Euphorbia myrsinites* (contact angle 162°) and lotus *Nelumbo nucifera* (contact angle 162°), data from [36.8]. Leaves of all three species are characterized by convex **(a,b)** to papillose **(c)** cells, covered by three-dimensional wax crystals

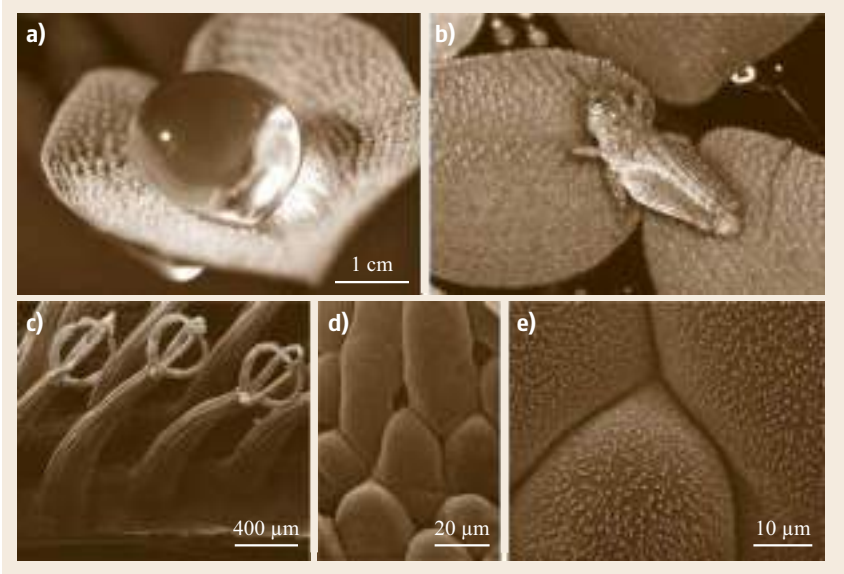


Fig. 36.23a–e Air-retaining surfaces of the floating fern *Salvinia*. In **(a)** a water droplet sits on the upper leaf side of *Salvinia biloba* and in **(b)** a mimicking grasshopper *Paulinia acuminata* feeding on the leaf even has a wax surface and is superhydrophobic. The SEM micrographs **(c–e)** of the leaf surface of *Salvinia* show the multicellular hairs **(c)**, and in higher magnification **(d)** the epidermis cells, and the wax rodlets **(e)** on the epidermis cells

lutionary potential: e.g., the flowering stalks of the watermilfoil *Myriophyllum* become superhydrophobic like the unique flowers of *Nymphoides* [36.1]. Floating species like *Salvinia* or *Pistia* are usually superhydrophobic. A striking example is illustrated in Fig. 36.23b: the grasshopper *Paulinia acuminate* feeds exclusively on *Salvinia* and optically mimics its color and surface (camouflage) – but it also is superhydrophobic based on wax crystals, the same adaption to its semiaquatic habitat as its host plant [36.1, 128].

Only few surfaces like lotus leaves have a very stable superhydrophobicity. As indicated in the introduction, superhydrophobicity usually disappears in aging biological surfaces and they become wettable (compare Fig. 36.1a).

36.6 Functional Diversity of Plant Surfaces

The surface of plants is the critical interface for interaction with the environment and fulfills many different functions (Fig. 36.24). Aspects and literature have been summarized by [36.1, 72, 73, 78, 129–131]. The surfaces are usually multifunctional and each attempt to categorize them is thus unrewarding [36.85, 132]. We simplify this in the following seven categories.

36.6.1 Mechanical Properties

The cuticle itself is a highly sophisticated chemically stable layer that serves as an elastic mechanical protective structure [36.76, 129]. The shape of a ripe tomato is maintained to a high degree by its thin cuticle. Hierarchical surface structures fulfill various mechanical tasks; they can reduce the ability of insects to walk (e.g., fish-hooked hairs on bean leaves against aphids, trapping surfaces in insectivorous plants or flowers of *Arum*) or increase the ability of insects to walk, as is the case in flowers, and even orient the visitors with cuticular folds [36.133]. Mechanochemical defense trichomes are common, such as in nettles (Urtica) or in Loasaceae (Fig. 36.14d)

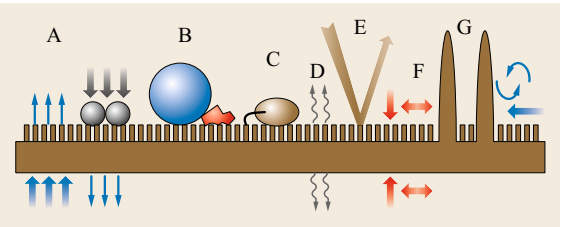


Fig. 36.24 Schematic of the most important functions of the plant boundary layer on a hydrophobic microstructured surface: A) transport barrier – limitation of uncontrolled water loss/leaching from interior and foliar uptake, B) surface wettability, C) anti-adhesive, self-cleaning properties (lotus effect) – reduction of contamination, pathogen attack and reduction of attachment/locomotion of insects, D) signaling – cues for host-pathogens/insect recognition and epidermal cell development, E) change of optical properties, F) mechanical properties – resistance to mechanical stress and maintenance of physiological integrity, G) reduction of surface temperature by increasing turbulent air flow over the boundary air layer (modified after [36.77])

36.6.2 Attachment

Attachment mechanisms are most common in zoochorous fruits like burrs, where complicated hooked spines and hairs have evolved in many groups (Asteraceae, Krameriaceae, Pedaliaceae) [36.134]. A particularly refined attachment mechanism with extractable cellulose threads occurs in the seeds of the orchid *Chiloschista* [36.91]. Attachment plays a role in the interaction between pollen and its disperser and the stigma, where it becomes deposed. Climbing plants like ivy (*Hedera*) or some highly specialized water plants growing on rocks in currents (e.g., *Podostemonaceae*) exhibit attachment devices by glue-like adhesives, which are not really surface phenomena.

36.6.3 Reflection, Absorption, and Transmission of Spectral Radiation

These phenomena are of crucial importance for light harvesting and temperature control under insolation. Colors as well as light reflection play an important role in the interaction between flowers and their pollinators.

Light Management

Absorption of light for photosynthesis is the precondition for plant life. Leaves need to collect electromagnetic radiation through photosynthesis while flowers need to harvest radiation in order to intensify their coloration to be more attractive for pollinators. In these processes the architecture of the plant surfaces plays an important role. While self-cleaning surfaces often combine convex- or concave-shaped epidermal cells

with water-repellent 3-D waxes, light-harvesting leaf surfaces often possess only convexly shaped epidermal cells. Specifically plants under low light conditions reduce the loss of light due to specular surface reflection by increasing the transmittance of energy via multiple reflections between the surface structures [36.135]. An extreme example is the *luminescent moss Schistostega* living in caves with batteries of spherical cells collecting the faintest light like lenses [36.136]. However, convexly shaped cells combined with a cuticular folding on top are well known for petal surfaces, especially in angiosperms [36.2]. These cuticular folds were thought to reduce the surface reflection [36.137], act as a specific optical signal for pollinators [36.2] and cause iridescence generated through diffraction gratings [36.138].

Coloration Signals

The intriguing interaction between flowers and their pollinators is the reason for the evolution of most refined coloration signals (e.g., [36.139]). The spectrum comprised virtually all grades of reflections (from black to white) and all visible colors, and even metallic-mirror-like surfaces have been evolved in the orchid *Ophrys speculum*, using thin layer iridescences like in our example (Fig. 36.4a). To intensify their color signal some flowers use refined surface structures. The surface architecture of the flowers of the so called *Johnny-jump-up*, *Viola tricolor*, is an excellent model for this color intensification process. *Viola* possess extreme papillated epidermal cells (aspect ratio about 2.93) covered with a fine cuticular folding (width about 0.26 μm). This surface topography acts as a light trap for incident light by reducing the specular reflection on the surface. Via this surface design more light passes through the cell wall. Furthermore, the extremely steep cell walls (papilla tip angle about 26°) cause an optimized scattering of the light within the petal (scattering angle of about 170°). This allows the absorption of more light by the pigments, which further results in the intensification of the color signal. Such a surface is highly interesting for biomimetic applications, for example in solar panel development. Flowers also use optical signals in the ultraviolet range, not visible to the human eye, but visible to visiting bees: *Bidens ferulifolia* in Fig. 36.25 provides an impressive example. Large databases on ultraviolet reflection of flowers are published [36.140–142]. Usually it is assumed that the reflection is determined only by chemical compounds (flavonoids) within the living cells (vacuoles), but we have recently shown that UV patterns are also supported by the surface structure. Another phenomenon caused by surface structures is iridescence of some leaves like in the fern *Elaphoglossum* (Fig. 36.4a). This phenomenon seems to occur mostly under low light intensities of tropical rainforest understories, but is also known in some flowering plants, e.g., the petals of some tulips [36.138, 143]. Camouflage (like in the grasshopper *Paulinia* Fig. 36.23b) plays an important role in desert plants (mainly South African Aizoaceae like the *living stones Lithops*): not only characteristics such as color, but also the microstructure of the surrounding environment are mimicked with high precision by their surfaces.

Temperature Control Under Insolation

The plant surface can only tolerate temperatures of up to 45 °C for living tissue. As a car top can heat up to more the 60 °C on a hot summer day, avoiding such temperatures is a crucial challenge for living tissue under solar radiation. Transpiration cooling is an obvious adaption, but inevitably connected with loss of water. Obviously, the highly reflective wax coatings or an indumentum of hairs decreases the heating [36.144–147]. An important aspect of temperature control – apart from reflections – may be the increase of turbulences by mechanical turbulence aids (e.g., the diversity of leaf mar-

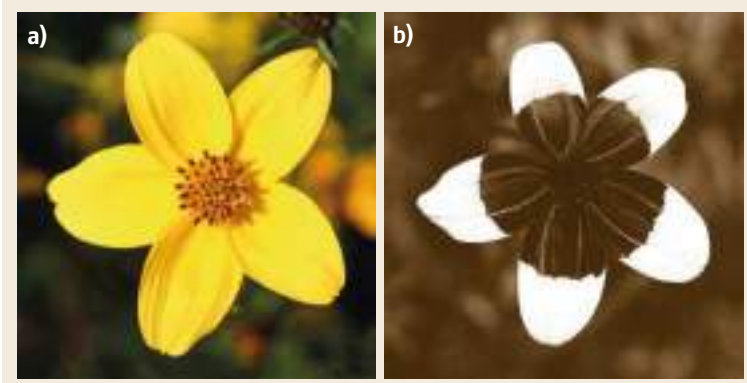


Fig. 36.25a,b Optical properties. Flower of *Bidens ferulaefolium* in the visible light for human eyes **(a)**. In **(b)** an ultraviolet image of the same flower is shown. Reflection and absorption of radiation are primarily influenced by chemical composition, but surprisingly also by the hierarchical surface architecture

gins) under dynamic (wind) conditions, which enhance the temperature exchange between the cooler air and the heated isolated leaf surfaces [36.148–151]. It is not by chance that succulents of the arid hot regions exhibit a high complexity of hierarchical surfaces sculpturing.

36.6.4 Reduction of Water Loss

Many of the land plant surface structures must be observed in context of reduction of water loss – a well-researched field. The cuticle, wax layers, and an indumentum of dense hairs play crucial roles in preventing loss of water; in some desert plants like *Sarcocaulon* the waxy layer might be more than one centimeter in thickness. This function is rather irrelevant for biomimetic applications: nonorganic materials like metals and synthetic polymers provide better solutions. *Uptake of water* is connected with superhydrophilicity (Sect. 36.6.5).

36.6.5 Superhydrophilicity

Superhydrophilicity is a physical constraint in water plants, but in land plants it is mainly connected with the uptake of water (roots, ephytic and fog collecting plants). Mosses without conductive tissue are superhydrophilic for an *ectohydric* water uptake; the most refined structures are found in *Rhacocarpaceae* (Fig. 36.26a,b) [36.152, 153]. All epiphytes (plants growing on other plants without contact to the soil) depend on rain, fog, and dew and they are all hydrophilic and most are even superhydrophilic [36.1]. The best-known examples are the hair-covered leaves of bromeliads (Fig. 36.26c). But there are many more plants with fog and dew collecting abilities in certain semiarid regions like the Atacama and Namib deserts [36.154–156], which are of particular interest for biomimetic application (fog collectors) (Fig. 36.26d). A remarkable case is the leaves of understory plants in tropical rainforests, like the superhydrophilic *Ruellia devosiana* with a strong convex sculpturing of its epidermis cells. Glands on the leaf surface release saponins, causing a thin hydrophilic coating, supporting a rapid spreading of water droplets [36.11].

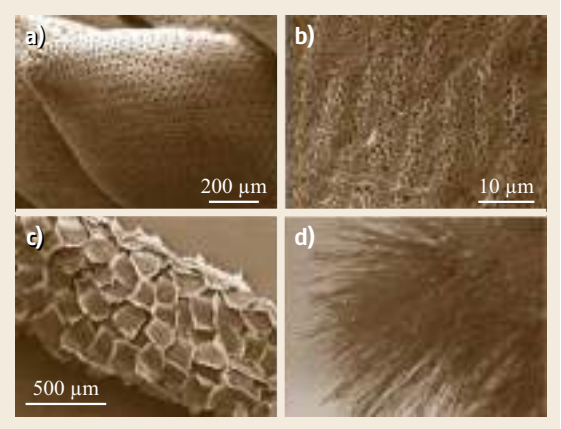


Fig. 36.26a–d In (a) the water absorbing pores of a *Sphagnum squarrosum* moss and in (b) the water adsorbing porous surface of the moss *Rhacocarpus purpurescens* are shown. In (c) the epiphytic growing Spanish moss (*Tillandsia usneoides*) is shown. (d) Water droplets in the needles of the Canary pine *(Pinus canariensis)* harvesting fog out of the trade wind clouds

Some superhydrophilic water plants have optimized surfaces and morphologies to resist water flows. By exhibiting grid-like leaves the Madagascar laceleaf (Fig. 36.5c) reduces its flow resistance. The superhydrophilic wave-swept giant seaweeds (Phaeophyta) are adapted by their morphology to reduce drag [36.157]. Their leathery thallus is covered by a mucilage, and recent work indicates that the superhydrophilicity of the mucilaginous surface acts as a drag-reducing agent like the mucilage of fish.

36.6.6 Superhydrophobicity

Superhydrophobicity is one of the most obvious characteristics of many land plant surfaces and may serve different functions. In particular the structures for air retention underwater are highly complex (e.g., *Salvinia*, Fig. 36.23), and they are often combined with a superimposed compartmentalization [36.1] of the whole organ surfaces from the microscopic to the macroscopic level (Fig. 36.27).

Superhydrophobicity serves many different functions, including:

1. *Unwettability per se* to avoid wetting for mechanical reasons: thin laminar leaves become too heavy when wetted. This is probably an important determination leading to the adaptation of superhydrophobicity in larger leaves in rainforests (e.g., *Cibotium*, [36.1]) – an extreme examples is the Titan Arum *Amorphophallus* (Fig. 36.1b).
2. *Reduction of adhesion* for insects – usually connected with superhydrophobicity – is caused by epicuticular wax crystals in insect-catching plants (e.g., *Sarracenia*, *Nepenthes*) or flowers that trap their pollinators (*Ceropegia, Aristolochia*) or to avoid unwanted nectar thieves (*Fritillaria, Lapageria*). Superhydrophobicity itself is not the primary function in these surfaces.

3. *Reduction of adhesion to avoid contamination* (*lotus effect* Fig. 36.21) enables plants to be cleaned from any kind of contaminating particle by raindrops, it is probably the most important function of superhydrophobicity in plants [36.1, 16, 159, 160]. Biologically, such surfaces in plants and animals should be primarily seen as a defense mechanism against fungal spores and the colonization of other microorganisms.
4. *Air layers for buoyancy*. Nonpersistent air layers for temporary buoyancy in water play an often underestimated role in very small seeds (Fig. 36.27a) and spores that may fall or are dispersed into water.

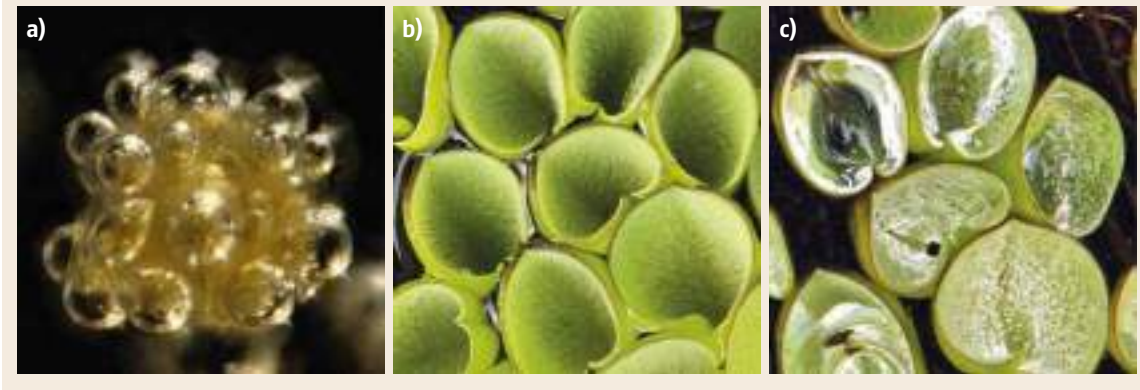


Fig. 36.27a–c Compartmentalization of surfaces generating small functional units for air retention **(a)** in microscopic dimensions in the seed of *Aeginetia indica* illustrated in Fig. 36.3 for temporary floating. A particularly refined compartmentalization system **(b,c)** is exhibited in *Salvina cucullata* in a macroscopic dimension: the superhydrophobic leaves (diameter around 1.5 cm) form hood-like compartments **(b)**, and when submersed in water **(c)** each leaf holds a very stable large air bubble

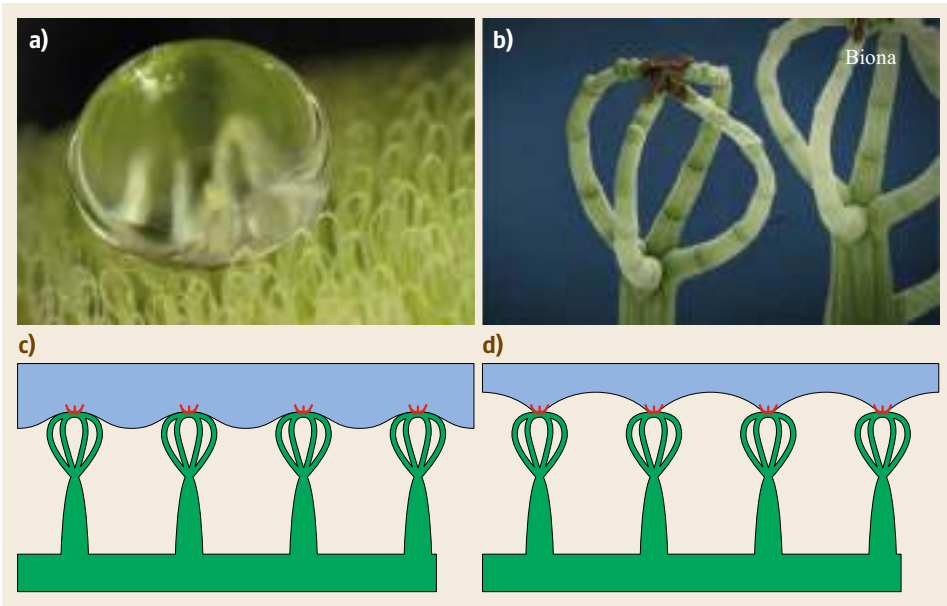


Fig. 36.28a–d The Salvinia paradox of chemical heterogeneities. **(a)** Water droplet on the eggbeater-shaped trichomes of *Salvinia molesta*. The deformations at the contact points of the droplet with the tips of the trichomes indicate an adhesion of the water; the droplet does not roll off an inclined surface. **(b)** Colored SEM image of the wax-covered superhydrophobic trichomes (green coloration) and the terminal hydrophilic four anchor cells for pinning the air–water interface. **(c,d)** Schematic of the air–water interface under dynamic hydrostatic conditions. **(b)** from [36.158]

5. *Air layers for gas exchange (respiration)* play mainly a role in animals (survey in [36.161, 162]), but also in plants like *Salvinia* that become temporarily inundated underwater.
6. *Air layers for fluid drag reduction* (survey in [36.1]). We could show that persistent air layers reduce the friction. The mechanism of this drag reduction is fairly simple: the air layer serves as a slip agent. On solid surfaces, the velocity of the water directly at the surface is zero due to the friction between the water and the surface molecules. If an air layer is mounted between the water and the solid surface, then the water streams over the air layer. The viscosity of air compared to water is 55 times lower, and because of this the air layer serves as a slip agent and the drag is reduced. Air layers as slip agents have evolved in many insects and play probably no role in plants. However, *Salvinia* provides the best example for the maintenance of permanent air layers underwater (Salvinia effect), showing the four essential criteria for air retention: hydrophobic chemistry, hair-like structures, undercuts and elasticity of the structures. Also the fifth, nonessential criteria, hydrophilic chemical heterogeneities of anchor cells within its superhydrophobic surface (Salvinia paradox, Fig. 36.28) could be found in some *Salvinia* species [36.69, 163–170].

36.6.7 Anti-adhesive *Slippery* Surfaces and Aquaplaning

Many plants have evolved special structured surfaces that hinder the attachment of animals, especially insects, to protect themselves against herbivores [36.171]. Most insects possess two different types of attachment structures: claws and adhesive pads [36.172, 173]. Whereas the former are used to cling to rough surfaces, the latter enable them to stick to perfectly smooth substrates. One strategy to reduce the attachment of insects is the secretion of epicuticular waxes that assemble into three-dimensional microstructures. The other strategy is the development of a slippery surface by inducing aquaplaning.

Aquaplaning

Several *Nepenthes* species do not possess a waxy layer, but are nevertheless fully functional insect traps. It was found that *Nepenthes* evolved another capture mechanism, which is based on special surface properties of the pitcher rim (peristome) (Fig. 36.29a,b). The peristome is characterized by a regular microstructure with radial ridges of smooth overlapping epidermal cells, which form a series of steps towards the pitcher inside. The peristome ridges (Fig. 36.29c) mostly extend into tooth-like structures at the inner edge, in between which large glands (extrafloral nectaries) are situated. Secretion of nectar by the glands attracts small insects, but also leads to a hydrophilic coverage of the surface. The plant surface microstructure combined with hydrophilic surface chemistry renders the pitcher rim completely wettable. Water droplets spread rapidly and form homogeneous thin films, which make the peristome extremely slippery for insects. When the peristome is wet, the fluid films prevent the insects' tarsal adhesive pads from making close contact with the surface, similar to the aquaplaning of a car tire on a wet road. In addition, the anisotropic microstructure of the peristome surface allows interlocking of claws only while the insect is running towards the pitcher inside, but not on the way out [36.174]. Under natural conditions the slippery water films are caused by rain, condensation and nectar secretion. In contrast to this, dry peristomes are not slippery for insects. This weather-dependent variation of peristome slipperiness leads to an intermittent and unpredictable activation of *Nepenthes* pitcher traps, which might make the evolution of specific avoidance behaviors more difficult [36.175].

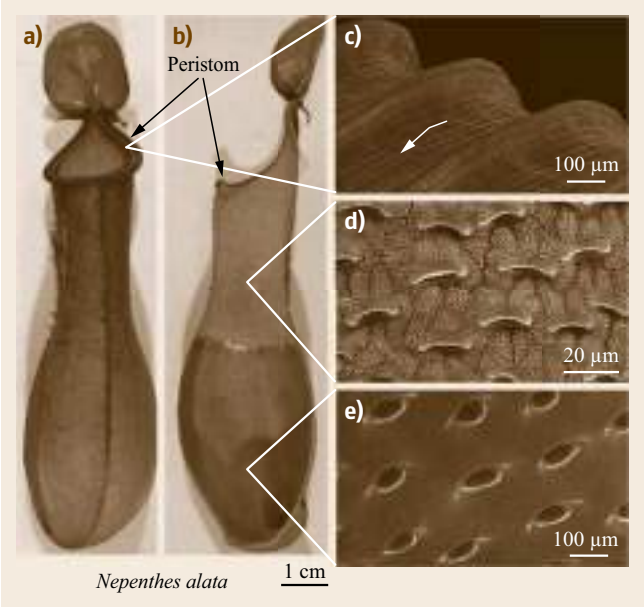


Fig. 36.29a–e Pitcher traps of the carnivorous plant *Nepenthes alata*. In **(a)** the complete pitchers trap of *N. alata* and in **(b)** a longitudinal cut through the trap are shown. **(c)** shows parallel ridges of the hydrophilic peristome. The *arrow* indicates the direction toward the inside of the pitcher (courtesy Holger Bohn). In **(d)** the waxy and slippery surface inside the trap with inactive stomata is shown. In **(e)** glands located in the digestive zone at the lower part of the trap are shown

36.7 Biomimetic Application

Bionics or biomimetics describes processes in which structures and concepts evolved by living organisms are taken and implemented into technologies (surveys in [36.1, 12, 20]). Bionics or biomimetics is an old field (historical survey in [36.12]), but functional surfaces came surprisingly very late to bionic applications. The first example was the hook-and-loop fastener by the Swiss engineer Georges de Mestral in the 1950s, popularly known as Velco®. It is based on plant surfaces (burrs). The drag reducing riblets of the shark skin were analyzed in the 1980s [36.18] and together with the self-cleaning lotus surfaces [36.2, 16] ushered in a new era of biomimetic applications – including the swimming competition in Olympic games (Sect. 36.7.6).

Surfaces in engineering are thin boundary layers, and they are exposed to the harsh physiochemical and mechanical influences of the environment. Thin layers, and in particular the biomimetic interesting hierarchical structures, have a restricted stability and durability. Nanocoatings play a most important and increasing economic role (compare [36.176]). We live – in the literal sense of Mark Twain's novel – in *the gilded age* of surface applications. The bulk of a technical solid is usually stable, but every homemaker knows the difference between the durability of a silver spoon in contrast to a silvered utensil. A medieval iron pan could still be used in the kitchen over the next centuries – but a polymer coated anti-adhesive pan lasts only a few years. The extremely long lifespan of spectacle glasses is today limited by the sophisticated and convenient anti-reflex coatings. In colloquial language, there is an obvious difference between *solid* and *superficial*. We need coating technologies – but durability and persistence is a major technical challenge for engineers and material scientists in a world of decreasing natural resources that must be used sustainably.

This section is organized using the same scheme as the previous section on plant functionality.

36.7.1 Mechanical Properties

Mechanical properties play a very important role in biomimetic composite materials, but are nearly negligible for surfaces. Technical materials available today (polymers, metals, etc.) have often superior mechanical properties.

36.7.2 Attachment

Attachment mechanisms of plants were the first biomimetic surface application and have been around since 1958 (see above) with the Velcro® hook-and-loop fasteners. Plants, in contrast to animals (e.g., geckos), are only a limited source of inspirations for attachment mechanisms.

36.7.3 Reflection, Transmission, and Absorption of Spectral Radiation

These phenomena are of great importance for biomimetics. The applications concentrate on light harvesting by bioinspired surfaces of solar panels. An appropriate model for such surfaces are the petals of *Viola tricolor* and its related species (Sect. 36.6.3). They show a highly reduced reflection of light and an optimized scattering within the petals. Additionally, they are superhydrophobic, showing a contact angle of 169° [36.177]. Temperature control under insolation (e.g., car tops, roofs) is another field of interest. Coloration technologies have always been bioinspired (like indigo and purple as ancient dyes used for millennia). Flowers exhibit sophisticated structures to intensify colors by hierarchical structuring. However, most inspirations are provided by animal coloration, like the iridescence of peacock feathers or butterfly (*Morpho*) wings. Butterflies also provide ultrablack colorations [36.178] like the Gaboon viper [36.179]. It is not by chance that the darkest technical black (Vantablack) is produced by carbon nanotubes.

36.7.4 Reduction of Water Loss

Reduction of water loss is essential for terrestrial plants, but like the many mechanical properties (Sect. 36.7.1) of minor importance for biomimetic applications. Technical materials like aluminum foil prevent loss better than a cuticle.

36.7.5 Superhydrophilicity

Superhydrophilicity has recently become relevant for technical applications, which are most diverse. Biomimetic fog-collecting meshes and other devices will probably play an increasing role in certain arid regions (e.g., Chile, Namibia) and it was shown that a biomimetic superhydrophilic hierarchical structuring increases the collection efficiency [36.155, 156, 180, 181]. *Koch* et al. [36.126] have shown that the hydrophilic *Gazania* petal structures provide a model for the design of microfluidic devices for small-volume liquid transport by capillary forces, which is beneficial in both fog harvesting and microfluidic devices.

Evidently, superhydrophilic surfaces are wettable, but the water film evaporates very fast and leads to the

fast drying *no-drop* glasses (e.g., Alltop®) and a sophisticated façade paint (StoColor Dryonic®). Drag reduction by mucilaginous surfaces seen in fish also play a technical role, and we have seen evidence in seaweed surfaces that plants have also evolved a similar technology.

The form and dynamics of bouncing droplets [36.95, 182] is influenced by surface chemistry and structure: superhydrophilic surfaces may prevent splashing, an important effect for biomedical applications (hygiene) or agriculture [36.183]. In plants, this phenomenon was analyzed in *Calathea zebrina*, a hydrophilic tropical understory plant, and its corresponding polymer replica (Fig. 36.30), by [36.124]. The antisplashing effect is promising for biomimetic applications.

36.7.6 Superhydrophobicity

The most obvious feature of many plant surfaces, superhydrophobicity has had an increasing and most diverse role in biomimetics as early as the 1990s (survey in [36.1]). Superhydrophobicity does not exist in abiotic nature and is one of the most important biomimetic applications for different aspects. Superhydrophobicity is generated in many plants by nonacosan-10-ol tubules with a diameter of 110 nm (Fig. 36.13). It could be shown that carbon nanotubes can also be used to generate superhydrophobic, even air-retaining surfaces [36.184].

Pure Water Repellency

Pure water repellency (e.g., umbrellas to textile building) is usually applied by sprays providing a sculpturing by nanoparticles on e.g., textiles (which are intrinsically hierarchically structured). It is obvious that dry conditions further reduce corrosion, and anticorrosion coatings are the focus of technical applications [36.185]. This also applies for anti-icing properties. A whole range of products and techniques are available, mostly coatings or sprays, and the product names often indicate their function (e.g., *NeverWet*). Dry surfaces may allow new dimensions in textile-based architecture owing to weight reduction during rainfall. Anti-ice and antifrost performance is increasing in importance [36.186, 187].

Self-Cleaning Lotus Effect

Self-cleaning surfaces are by far the most important applications, [36.1] and a vast literature exists beginning in 1997 [36.20, 54, 188–190]. It is also one of the most important applications in façade paints, with the added advantage that they remain functional for several decades.

The lotus effect is evolved in plants predominantly as a defense mechanisms against pathogens: the attachment e.g., of fungal spores is dramatically reduced [36.191] and no water films are available as a precondition for colonization or the formation of biofilms. This has an unexpected application: in commercial agriculture all plants are treated with pesticides and surfactants have to be added. It was shown that the application of the solution containing only the surfactant (without the active ingredient) increases the chances of infection significantly [36.191–196]. This phenomenon will hopefully lead to the development of a new generation of less aggressive, bioinspired surfactants for pesticide application.

Superhydrophobic Plants Surfaces

Superhydrophobic plants surfaces, like those of *Salvinia*, have a striking capacity to collect and adsorb spilled oil from water [36.197], a basis for biomimetic materials. Hydrophobic sand coating can be used to control deep drainage in tailings [36.198].

Air-Retaining Salvinia Effect

Air-retaining surfaces are a recent development, and only prototypes exist [36.1, 199, 200]. Fluid drag reduction of up to 30% in a hydrodynamic water channel has been measured [36.201]. The main application is seen

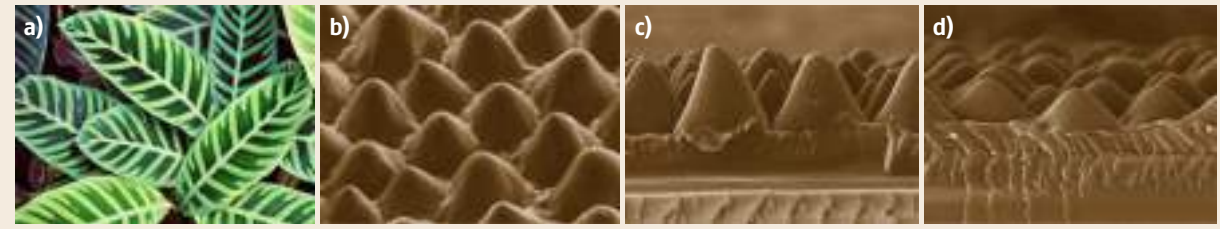


Fig. 36.30 **(a)** *Calathea zebrina* leaves with their light and dark green pattern. **(b)** SEM micrograph of the leaf surface (tilt angle 45°). SEM of the replicated dark leaf areas **(c)** and bright leaf areas **(d)** in side view (90° tilt angle). The visible color differences in the leaf are mirrored in the different shapes of the cell papillae of the light and dark surfaces areas

in ship hulls: container ships transport about 80% of the global goods and a reduction of the fuel (oil) consumption of 125 million t and of 395 million t of CO_2 is estimated. A novel technology using air-retaining grids has recently been introduced and can be combined with the existing refined microbubble technologies, where microbubbles adhere to the grid surface and the air layer under the grid can be regenerated. It is obvious that a permanently intact air layer prevents biofouling. Oil pollution is an unfortunate and unavoidable problem in our oceans. One possible solution is to create a hydrophobic barrier at the waterline of the ship to separate the ship hull from the water interface and avoid oil from creeping into the structures and suppressing the air.

Biomimetic drag-reducing boat surfaces based on shark skin riblets have been applied in America's Cup contests in 1987 and 2010; the Speedo Fastskin® swimming suits introduced in 2000 were adopted by almost all gold medal winners in the Olympic Games in Beijing 2008 (Fig. 36.31). But riblet structures reduce the friction only about 3%; the more complex Salvinia effect technologies can achieve about 30%. Swimwear covering the body almost totally was banned in 2009; however there is some evidence that in the Olympic Games in Rio 2016 a refined technology of bionic swimming suits combining superhydrophobic Salvinia-effect areas with shark riblet areas may have been applied. Underwater persistent air layers will probably fulfill many other functions like gas exchange (*plastron* in biology) or even refined sensory functions like in the backswimmer *Notonecta* [36.202].

36.7.7 Other Applications

As indicated in the preceding section, there are additional functions. Like the *aquaplaning* effect of the surfaces of the carnivorous *Nepenthes* plants that produce water films and become slippery to trap insects [36.174, 203]. In particular slippery liquid-infused surfaces (SLIPS) are of focal interest (e.g., [36.204, 205]).



Fig. 36.31a,b Some Olympic records today are also a triumph of bionics. Mark Spitz **(a)** wore his famous prebionic swimsuit in the 1972 Olympics in Munich. In contrast, biomimetic swimsuits covering almost the entire body were popular after 2000; the suits were based on the biomimetic drag-reducing shark riblet technology. At the World Swimming Championship in July 2009 in Rome almost all world records were broken using full-body biomimetic polyurethane swimsuits. Subsequently the length of swimwear was limited by the World Swimming Federation FINA in December 2009. But almost full-body swimwear is back today, as e.g., the swimming suit of Nađa Higl **(b)**: somewhat smaller than in 2009, but still covering large parts of the swimmers. Today it seems to be a high-tech combination of chemically heterogenic surfaces: shark riblets (drag reduction about 3%) and superhydrophobic Salvinia-effect (drag reduction about 30%) areas. For divers, drag reduction plays no role – the divers in Rio de Janeiro 2016 were still dressed like Mark Spitz in 1972. Sources: **(a)** kindly provided by the National Portrait Gallery, Smithsonian Institution. **(b)** from Wikimedia Commons

36.8 Living Prototypes: Evolution of Plant Surfaces and Biodiversity

The often-used phrase *inspired by nature* for biomimetics is misleading: a nuclear reactor is inspired by natural nuclear fission or the nuclear fusion in our sun. Thus a nuclear reactor is inspired by nature – but not bionic. Biomimetics is exclusively based on living organisms and the evolution of some 10 million living prototypes (survey in [36.1]). Superhydrophobic surfaces occur exclusively in living organisms, which evolved hierarchical structured superhydrophobic surfaces based on a very limited selection of molecules. There is no evidence for the occurrence of superhydrophobic surfaces in inorganic *nature* (man-made technical products are excluded). Surfaces are the crucial interface between an organism and its environment – 3.5 billion years of mutation and selection have created a stunning diversity in an estimated 10 million different species of plants and animals (survey in [36.1]). The two basic phases with contrasting physical constraints are life in water and life outside of water.

36.8.1 Water Plants

Nonvascular primary water plants evolved in water (originally the oceans) a couple of billion years ago and have no true conductive tissue. The groups are phylogenetically very different nonrelated clades; they comprise microscopic unicellular algae up to seaweeds that might reach a height of more than 45 m in *Macrocystis*. The common feature is their wettability (superhydrophilicity) and usually the absence of hierarchical structuring: flat and mucilaginous surfaces are characteristic. In certain groups of unicellular algae (e.g., Coccolithophora and Diatoms) complex surface structures occur, but usually embedded in mucilaginous or plasmatic covers; they are probably not the environmental interface.

Vascular secondary water plants are plants of terrestrial origin, which recolonized aquatic habitats in freshwater (e.g., *Elodea, Potamogeton*) or even the oceans like seagrasses (e.g., *Posidonia, Zostera*). They exhibit the characteristic superhydrophilic flat surfaces, often becoming mucilaginous by biofilms, as primary water plants. However, they retain the ability of their ancestors to produce superhydrophobic surfaces when leaves or flowering shoots emerge from the water (e.g., in *Myriophyllum*). The ancestral nonvascular *Bryophyta* (mosses, hornworts, and liverworts) depend on the uptake of water through leaves and are superhydrophilic, occasionally exhibiting refined absorption structures like *Rhacocarpus* [36.153]. However, some (e.g., *Funaria hygrometrica*) have the ability to produce superhydrophobic wax covers in their spore capsules (Fig. 36.32b); the same is true for propagation structures of lichens (Fig. 36.32a) [36.206] and the aerial hyphae of fungi (Fig. 36.32c) or even the capillitium of slime molds (Fig. 36.32d) [36.1].

36.8.2 Land Plants

True land plants are the vascular or higher plants, a well-defined phylogenetic unit. They comprise the ferns, clubmosses and horsetails, gymnosperms (including conifers) and the flowering plants (angiosperms). Higher plants evolved with the conquest of land in the late Ordovician or Silurian, 430–500 million years ago and evolved mechanically stabilizing lignose vascular tissues and a protective polymer layer, the cuticle (Sect. 36.3) on their stem and leaf surfaces; as a consequence of this rather impermeability, surface pores (stomata) evolved to enable and control gas exchange. It was shown by a phylogenetic analysis [36.1] that superhydrophobicity caused by epicuticular wax crystals evolved simultaneously with the conquest of land – a possibly overlooked evolutionary key invention for plant life outside of water and probably also for

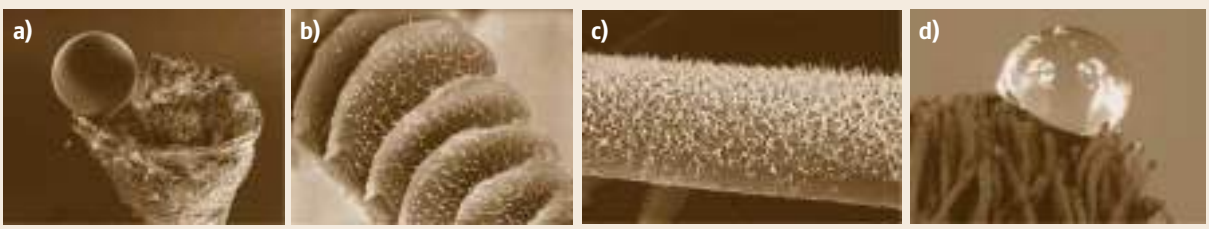


Fig. 36.32a–d Superhydrophobicity in lichens mosses, fungi, and slime molds. **(a)** Fruiting body of the lichen *Cladonia chlorophaea* with water droplet. **(b)** Peristome of the moss *Funaria hygrometrica*, covered by wax rodlets. **(c)** The aerial hyphae of the gray mold (*Botrytis*) are covered by wax-like crystals. **(d)** The fruiting body (capillitium) of the slime mold *Stemonitis* possesses a superhydrophobic surface caused by a granular layer of unknown chemical composition

insects. One of the most common waxy substances is the fatty secondary alcohol nonacosan-10-ol, which is responsible for the superhydrophobicity of many ferns, conifers or even lotus leaves as long ago as 250 million years [36.1].

Some 450 000 different species of plants are part of the astonishing biodiversity of our planet, which, including animals, comprises some 10 million species – but we know of less than 20% of them. Plants have evolved most intriguing functional surfaces over millions of years, like in the lotus (*Nelumbo*) or the floating ferns (*Salvinia*). Evolution is a slow process spanning billions of years. Mutation and selection (*trial and error*) exploited all constructional possibilities within this time and with the limited materials. But millions of years of research and development for technical engineering today is not a possibility: materials science has a goal of fabricating a particular product within a limited time, using experimental trial-and-error approaches, calculation and modeling. We lose a high amount of biodiversity in our changing world and it has been brought to our attention [36.12, 207] that this also means the loss of biological role models, the *living prototypes* for engineers.

36.9 Conclusions

The diversity of plant surface structures is a result of several billions of years of evolutionary processes. Plants evolved a stunningly high diversity of surfaces and functionality for their interaction with the environment – the self-cleaning properties of the lotus is only one example. Millions of years of mutation and selection, trial and error, mean free information for engineers and materials scientists.

Bionics is an old field of research and development starting around 1800 – but surfaces played a surprisingly late role for biomimetic applications, the only exception is the hook and loop fasteners (Velcro®) in the 1950s based on burrs. The publication of the lotus effect in 1997 [36.16] created awareness by engineers and materials scientists, and terms like *superhydrophobicity* came into use in the last two decades and opened a new era in surface technologies (survey in [36.1]). Surfaces play an increasing role, and the global market for nanocoatings is estimated to reach 14.2 billion US dollars by 2019 [36.176].

Biological surfaces have provided a remarkable number of innovations in the last three decades. Surface technologies have been largely influenced by research on biological interfaces and came rather late into focus in technical innovations. All data indicate we are only in the beginning of a new era of biologically inspired surface technologies.

Understanding biological surfaces is crucial – we have shown that we are still in the beginning of this process. But we are also in the beginning of a dramatic loss of biodiversity in the Anthropocene. Some 10 million different species (possible biological role models) exist – we lose a high amount of biodiversity in our changing world and it has been brought to our attention [36.12, 207] that this also means the loss of biological role models, the *living prototypes* for engineers. Bionics is another intrinsic value to the diversity of life that should be treasured.

Acknowledgments. This survey is based on over three decades of research on surfaces by a large working group with ever-changing members: first and foremost, we acknowledge all the students who had successful theses in this field. Their names can be found in the papers quoted in the references. We acknowledge the discussions with colleagues and friends in particular we would like to thank Horst Bleckmann (Bonn), Alfred Leder, Martin Brede (both Rostock), Thomas Schimmel, Stefan Walheim, Markus Moosmann, Torsten Scherer (all KIT Karlsruhe), Walter Erdelen (former UNESCO Paris), Anna-Julia Schulte (Euskirchen), Daud Rafiqpoor, Birte Böhnlein, Peter Häger (all Bonn), Gerhard Gottsberger (Ulm), Stanislav Gorb (Kiel), Georg Noga (Bonn), and Maximilian Weigend (Bonn). Technical assistance was provided over the years by Hans-Jürgen Ensikat, the late Wolfgang Roden, Alexandra Runge, Bernd Haeseling, and Danica Christensen. Our work in bionics and biodiversity was supported by the Deutsche Bundesstiftung Umwelt DBU, the German Research Council DFG, the Federal Ministry for Science and Education BMBF, and the Academy of Science and Literature in Mainz.

References

36.1 W. Barthlott, M. Mail, C. Neinhuis: Superhydrophobic hierarchically structured surfaces in biology: Evolution, structural principles and biomimetic applications, Philos. Trans. R. Soc. A **374**(2073), 20160191 (2016) doi:10.1098/rsta.2016.0191

36.2 W. Barthlott, N. Ehler: Raster-Elektronenmikroskopie der Epidermis-Oberflächen von Spermatophyten, Trop. Subtrop. Pflanzenwelt **19**, 1–105 (1977)

36.3 S. Blackmore, K. Ferguson (Eds.): *Pollen and Spores: Form and Function*, Linn. Soc. Symp. Series, Vol. 12 (Academic Press, London 1986)

36.4 S.N. Agashe, E. Caulton: *Pollen and Spores – Application with Special Emphasis on Aerobiology and Allergy* (Taylor & Francis, Enfield 2009)

36.5 K. Koch, B. Bhushan, W. Barthlott: Multifunctional plant surfaces and smart materials. In: *Springer Handbook of Nanotechnology*, 3rd edn., ed. by B. Bhushan (Springer, Berlin, Heidelberg 2010) pp. 1399–1436

36.6 B. Bhushan: *Springer Handbook of Nanotechnology*, 3rd edn. (Springer, Berlin, Heidelberg 2010)

36.7 W. Barthlott, M. Mail, B. Bhushan, K. Koch: Plant surfaces: Structures and functions for biomimetic innovations, Nano-Micro Lett. (2017) doi:10.1007/s40820-016-0125-1

36.8 C. Neinhuis, W. Barthlott: Characterization and distribution of water-repellent, self-cleaning plant surfaces, Ann. Bot. **79**, 667–677 (1997) doi:10.1006/anbo.1997.0400

36.9 S. Porembski, B. Martens-Aly, W. Barthlott: Surface/volume – rations of plants with special consideration of succulents, Beitr. Biol. Pflanz. **66**, 189–209 (1992)

36.10 J.M. Suttie, S.G. Reynolds, C. Batello: *Grasslands of the World*, Plant Production and Protection Series, Vol. 34 (Food and Agricultural Organisations of the UN (FAO), Rome 2005)

36.11 K. Koch, I.C. Blecher, G. König, S. Kehraus, W. Barthlott: The superhydrophilic and superoleophilic leaf surface *of Ruellia devosiana* (Acanthaceae): A biological model for spreading of water and oil on surfaces, Funct. Plant Biol. **36**, 339–350 (2009)

36.12 W. Barthlott, D. Rafiqpoor, W. Erdelen: Bionics and biodiversity – Bio-inspired technical innovation for a sustainable future. In: *Biomimetic Research for Architecture and Building Construction – Biological Design and Integrative Structures*, ed. by J. Knippers, K. Nickel, T. Speck (Springer, Berlin, Heidelberg 2016)

36.13 R.L. Ripley, B. Bhushan: Bioarchitecture: bioinspired art and architecture – A perspective, Philos. Trans. R. Soc. A **374**, 20160192 (2016) doi:10.1098/rsta.2016.0192

36.14 R.H. Francé: *Die Pflanze als Erfinder* (Franckh'sche Verlagshandlung, Stuttgart 1920), Engl. edn. *Plants as Inventors* (Simpkin and Marshall, London 1926)

36.15 Living prototypes – The key to new technology. In: *Proc. Bionic Symp.*, ed. by J.C. Robinette (Wright Air Development Division, Dayton 1960)

36.16 W. Barthlott, C. Neinhuis: Purity of the sacred lotus, or escape from contamination in biological surfaces, Planta **202**, 1–8 (1997) doi:10.1007/s004250050096

36.17 H.C. von Baeyer: The Lotus effect, The Sciences **40**, 12–15 (2000)

36.18 D.W. Bechert, A. Dinkelacker, W.-E. Reif: On the fluid dynamic of the shark skin, Bull. Am. Phys. Soc. **28** (1983)

36.19 D.W. Bechert, M. Bruse, W. Hage, R. Meyer: Fluid mechanics of biological surfaces and their technological application, Naturwissenschaften **87**, 157–171 (2000)

36.20 B. Bhushan: *Biomimetics – Bioinspired Hierarchical-Structured Surfaces for Green Science and Technology*, 2nd edn. (Springer, Berlin, Heidelberg 2016)

36.21 P. Forbes: *The Gecko's Foot* (Fourth Estate, London 2005)

36.22 J.W. von Goethe: Verstäubung, Verdunstung, Vertropfung. In: *Die Schriften zur Naturwissenschaft*, Morphologische Hefte, Vol. 9, ed. by D. Kuhn (H. Böhlaus, Weimar 1954) pp. 210–221

36.23 A. Otten, S. Herminghaus: How plants keep dry: A physicist's point of view, Langmuir **20**, 2405–2408 (2004) doi:10.1021/la034961d

36.24 W. Barthlott, C. Neinhuis, D. Cutler, F. Ditsch, I. Meusel, I. Theisen, H. Wilhelmi: Classification and terminology of plant epicuticular waxes, Bot. J. Linn. Soc. **126**, 237–260 (1998) doi:10.1111/j.1095-8339.1998.tb02529.x

36.25 A. Dommisse, J. Wirtz, K. Koch, K. Wandelt, W. Barthlott, T. Kolter: Synthesis of Snonacosan-10-ol, the main component of plant surface tubular wax crystals, Eur. J. Org. Chem. **2007**, 3508–3511 (2007) doi:10.1002/ejoc.200700262

36.26 M. Riederer, C. Markstädter: Cuticular waxes: A critical assessment of current knowledge. In: *Plant Cuticles an Integrated Functional Approach*, ed. by G. Kerstiens (University Scientific, Oxford 1996) pp. 189–200

36.27 L. Kunst, A.L. Samuels: Biosynthesis and secretion of plant cuticular wax, Prog. Lipid Res. **42**, 51–80 (2003)

36.28 C.E. Jeffree: The fine structure of the plant cuticle. In: *Biology of the Plant Cuticle*, ed. by M. Riederer, C. Müller (Blackwell, Oxford 2006) pp. 11–125

36.29 R. Jetter, S. Schäffer: Chemical composition of the *Prunus laurocerasus* leaf surface. Dynamic changes of the epicuticular wax film during leaf development, Plant Phys. **126**, 1725–1737 (2001)

36.30 K. Koch, W. Barthlott, S. Koch, A. Hommes, K. Wandelt, W. Mamdouh, S. De-Feyter, P. Broekmann: Structural analysis of wheat wax (*Triticum aestivum*, c.v. 'Naturastar' L.): From the molecular level to three dimensional crystals, Planta **223**, 258–270 (2005) doi:10.1007/s00425-005-0081-3

36.31 R. Jetter, L. Kunst, A.L. Samuels: Composition of plant cuticular waxes. In: *Biology of the Plant Cuticle*, Ann. Plant Rev., Vol. 23, ed. by M. Riederer, C. Müller (Blackwell, Oxford 2006) pp. 145–175

36.32 E.A. Baker: Chemistry and morphology of plant epicuticular waxes. In: *The Plant Cuticle*, ed. by D.F. Cutler, K.L. Alvin, C.E. Price (Academic, London 1982) pp. 139–165

36.33 T. Shepherd, D.W. Griffiths: The effects of stress on plant cuticular waxes, New Phytol. **171**, 469–499 (2006)

36.34 K. Koch, K.D. Hartmann, L. Schreiber, W. Barthlott, C. Neinhuis: Influence of air humidity on epicuticular wax chemical composition, morphology and wettability of leaf surfaces, Env. Exp. Bot. **56**, 1–9 (2006)

36.35 C. Markstädter, W. Federle, R. Jetter, M. Riederer, B. Hölldobler: Chemical composition of the slippery epicuticular wax blooms on *Macaranga* Thouars. (Euphorbiaceae) ant-plants, Chemoecology **10**, 33–40 (2000)

36.36 M. Riedel, A. Eichner, R. Jetter: Slippery surfaces of carnivorous plants: Composition of epicuticular wax crystals in *Nepenthes alata* Blanco pitchers, Planta **218**, 87–97 (2003)

36.37 M. Wen, C. Buschhaus, R. Jetter: Nanotubules on plant surfaces: Chemical composition of epicuticular wax crystals on needles of *Taxus baccata* L, Phytochemistry **67**, 1808–1817 (2007)

36.38 H. Ensikat, C. Neinhuis, W. Barthlott: Direct access to plant epicuticular wax crystals by a new mechanical isolation method, Int. J. Plant Sci. **161**, 143–148 (2000)

36.39 H.J. Ensikat, B. Boese, W. Mader, W. Barthlott, K. Koch: Crystallinity of plant epicuticular waxes: electron and X-ray diffraction studies, Chem. Phys. Lipids **144**, 45–59 (2006) doi:10.1016/j.chemphyslip.2006.06.016

36.40 D. Frölich, W. Barthlott: Die Mikromorphologie der epicuticularen Wachse und das System der Monocotylen, Trop. Subtrop. Pflanzenwelt **63**, 1–135 (1988)

36.41 N.D. Hallam, B.E. Juniper: The anatomy of the leaf surface. In: *The Ecology of Leaf Surface Microorganisms*, ed. by T.F. Preece, C.H. Dickinson (Academic, London 1971) pp. 3–37

36.42 C.E. Jeffree: The cuticle, epicuticular waxes and trichomes of plants, with reference to their structure, functions and evolution. In: *Insects and the Plant Surface*, ed. by B.E. Juniper, R. Southwood (Edward Arnold, London 1986) pp. 23–63

36.43 K. Koch, C. Neinhuis, H.J. Ensikat, W. Barthlott: Self assembly of epicuticular waxes on living plant surfaces imaged by atomic force microscopy (AFM), J. Exp. Bot. **55**, 711–718 (2004) doi:10.1093/jxb/erh077

36.44 K. Koch, H.J. Ensikat: The hydrophobic coatings of plant surfaces: Epicuticular wax crystals and their morphologies, crystallinity and molecular self-assembly, Micron **39**, 759–772 (2008)

36.45 K. Koch, A. Dommisse, C. Neinhuis, W. Barthlott: Self-assembly of epicuticular waxes on living plant surfaces by atomic force microscopy. In: *Scanning Tunneling Microscopy/Spectroscopy and Related Techniques*, ed. by P.M. Koenraad, M. Kemerink (American Institute of Physics, Melville 2003) pp. 457–460

36.46 P.J. Holloway, C.E. Jeffree, E.A. Baker: Structural determination of secondary alcohols from plant epicuticular waxes, Phytochemistry **15**, 1768–1770 (1976)

36.47 R. Jetter, M. Riederer: *In vitro* reconstitution of epicuticular wax crystals: Formation of tubular aggregates by long chain secondary alkanediols, Bot. Acta **108**, 111–120 (1995) doi:10.1111/j.1438-8677.1995.tb00840.x

36.48 I. Meusel, C. Neinhuis, C. Markstadter, W. Barthlott: Chemical composition and recrystallization of epicuticular waxes: Coiled rodlets and tubules, Plant Biol. **2**, 462–470 (2000) doi:10.1055/s-2000-5961

36.49 W. Barthlott, I. Theisen, T. Borsch, C. Neinhuis: Epicuticular waxes and vascular plant systematics: Integrating micromorphological and chemical data. In: *Deep Morphology: Toward a Renaissance of Morphology in Plant Systematics*, ed. by T.F. Stuessy, V. Mayer, E. Hörandl (Reg. Veg. Gantner, Ruggell 2003) pp. 189–206

36.50 I. Meusel, C. Neinhuis, C. Markstadter, W. Barthlott: Ultrastructure, chemical composition, and recrystallization of epicuticular waxes: Transversely ridged rodlets, Can. J. Bot. **77**, 706–720 (1999) doi:10.1139/cjb-77-5-706

36.51 International Union of Crystallography: Report of the Executive Committee for 1991, Acta Crystalogr. A **48**, 922–946 (1992)

36.52 C.E. Jeffree, E.A. Baker, P.J. Holloway: Ultrastructure and recrystallization of plant epicuticular waxes, New Physiol. **75**, 539–549 (1975)

36.53 R. Jetter, M. Riederer: Epicuticular crystals of nonacosan-10-ol: in vitro reconstitution and factors influencing crystal habits, Planta **195**, 257–270 (1994) doi:10.1007/BF00199686

36.54 K. Koch, A. Dommisse, W. Barthlott: Chemistry and crystal growth of plant wax tubules of Lotus (*Nelumbo nucifera*) and Nasturtium (*Tropaeolum majus*) leaves on technical substrates, Cryst. Growth Des. **6**, 2571–2578 (2006)

36.55 G.M. Whitesides, M. Boncheva: Beyond molecules: Self-assembly of mesoscopic and macroscopic components, Proc. Natl. Acad. Sci. USA **99**(8), 4769–4774 (2002)

36.56 J. Zhang, W. Zhong-Lin, J. Liu, C. Shaowei, G. Liu: *Self Assembled Nanostructures* (Kluwer, New York 2003)

36.57 N. Boden, P.J.B. Edwards, K.W. Jolley: Self-assembly and self-organization in micellar liquid crystals. In: *Structure and Dynamics of Strongly Interacting Colloids and Supermolecular Aggregates in Solutions*, ed. by S.H. Chen, J.S. Huang, P. Tartaglia (Kluwer, Dordrecht 1992)

36.58 C. Neinhuis, K. Koch, W. Barthlott: Movement and regeneration of epicuticular waxes through plant cuticles, Planta **213**, 427–434 (2001) doi:10.1007/s004250100530

36.59 D. Dorset: Development of lamellar structures in natural waxes – An electron diffraction investigation, J. Phys. D – Appl. Phys. **32**, 1276–1280 (1999)

36.60 B. Bhushan, K. Koch, Y.C. Jung: Biomimetic hierarchical structure for self-cleaning, Appl. Phys. Lett. **93**, 093101 (2008)

36.61 B. Bhushan, K. Koch, Y.C. Jung: Nanostructures for superhydrophobicity and low adhesion, Soft Matter **4**, 1799–1804 (2008)

36.62 S. De Feyter, F.C. De Schryver: Self-assembly at the liquid/solid interface: STM reveals, J. Phys. Chem. B **109**, 4290–4302 (2005)

36.63 F.C. Meldrum, S. Ludwigs: Template-directed control of crystal morphologies, Macromol. Biosci. **7**, 152–162 (2007)

36.64 K. Koch, A. Dommisse, A. Niemietz, W. Barthlott, K. Wandelt: Nanostructure of epicuticular plant waxes: Self-assembly of wax tubules, Surf. Sci. **603**, 1961–1968 (2009) doi:10.1016/j.susc.2009.03.019

36.65 A. Fahn: Structure and function of secretory cells, Adv. Bot. Res. **31**, 37–75 (2000)

36.66 G.J. Wagner, E. Wang, R.W. Shephers: New approaches for studying and exploiting an old protuberance, the plant trichome, Ann. Bot. **93**, 3–11 (2004)

36.67 E. Wollenweber: The distribution and chemical constituents of the farinose exudates in gymnogrammoid ferns, Am. Fern J. **68**, 13–28 (1978)

36.68 W. Barthlott, E. Wollenweber: Zur Feinstruktur, Chemie und taxonomischen Signifikanz epicuticularer Wachse und ähnlicher Sekrete, Trop. Subtrop. Pflanzenwelt **32**, 7–67 (1981)

36.69 W. Barthlott, T. Schimmel, S. Wiersch, K. Koch, M. Brede, M. Barczewski, S. Walheim, A. Weis, A. Kaltenmaier, A. Leder, H.F. Bohn: The *Salvinia* paradox: Superhydrophobic surfaces with hydrophilic pins for air retention under water, Adv. Mater. **22**, 2325–2328 (2010) doi:10.1002/adma.200904411

36.70 P.J. Holloway: Plant cuticles: Physicochemical characteristics and biosynthesis. In: *Air Pollution and the Leaf Cuticle*, ed. by K.E. Percy, J.N. Cape, R. Jagels, C.J. Simpson (Springer, Berlin, Heidelberg 1994) pp. 1–13

36.71 P.E. Kolattukudy: Plant cuticle and suberin, eLS (2001) doi:10.1038/npg.els.0001920

36.72 J.T. Martin, B.E. Juniper: *The Cuticles of Plants* (Edward Arnold, London 1970)

36.73 D.F. Cutler, K.L. Alvin, C.E. Price: *The Plant Cuticle* (Academic, London 1982)

36.74 G. Kerstiens: *Plant Cuticles: An Integrated Functional Approach* (BIOS Scientific, Oxford 1996)

36.75 M. Riederer, C. Müller: Biology of the Plant Cuticle. In: *Annual Plant Reviews*, Vol. 23 (Blackwell, Oxford 2006)

36.76 H. Bargel, C. Neinhuis: Tomato (*Lycopersicon esculentum* Mill.) fruit growth and ripening as related to the biomechanical properties of fruit skin and isolated cuticle, J. Exp. Bot. **56**, 1049–1060 (2005)

36.77 H. Bargel, K. Koch, Z. Cerman, C. Neinhuis: Structure–function relationships of the plant cuticle and cuticular waxes – A smart material?, Funct. Plant Biol. **33**, 893–910 (2006) doi:10.1071/FP06139

36.78 W. Barthlott: Scanning electron microscopy of the epidermal surface in plants. In: *Application of the Scanning em in Taxonomy and Functional Morphology. Systematics Associations' Special Volume*, ed. by D. Claugher (Clarendon, Oxford 1990) pp. 69–94

36.79 H.-J. Ensikat, T. Geisler-Wierwille, M. Weigend: A first report of hydroxylated apatite as structural biomineral in Loasaceae-plants' teeth against herbivores, Sci. Rep. **6**, 26073 (2016) doi:10.1038/srep26073

36.80 W. Barthlott: Morphogenese und Mikromorphologie komplexer Cuticular-Faltungsmuster an Blüten-Trichomen von Antirrhinum L. (Scrophulariaceae), Ber. Dt. Bot. Ges. **93**, 379–390 (1980)

36.81 L.H.P. Jones, K.A. Handreck: Silica in soils, plants, and animals, Adv. Agron. **19**, 107–149 (1967)

36.82 A.G. Sangster, M.J. Hudson, H.J. Tubb: Silicon deposition in higher plants. In: *Silicon in Agriculture*, ed. by L.E. Datnoff, G.H. Snyder, G.H. Korndörfer (Elsevier, Amsterdam 2001) pp. 85–114

36.83 R.K. Saeedur: *Calcium Oxalate in Biological Systems* (CRC, Boca Raton 1995) p. 375

36.84 K. Koch, B. Bhushan, W. Barthlott: Diversity of structure, morphology and wetting of plant surfaces, Soft Matter **4**, 1943–1963 (2008) doi:10.1039/b804854a

36.85 K. Koch, B. Bhushan, W. Barthlott: Multifunctional surface structures of plants: An inspiration for biomimetics, Prog. Mater. Sci. **54**, 137–178 (2009) doi:10.1016/j.pmatsci2008.07.003

36.86 C. Martin, B.J. Glover: Functional aspects of cell patterning in aerial epidermis, Curr. Opin. Plant Biol. **10**, 70–82 (2007)

36.87 C.A. Brewer, W.K. Smith, T.C. Vogelmann: Functional interaction between leaf trichomes, leaf wettability and the optical properties of water droplets, Plant Cell Environ. **14**, 955–962 (1991)

36.88 E. Rodriguez, P.L. Healey, I. Mehta: *Biology and Chemistry of Plant Trichomes* (Plenum, New York 1984)

36.89 H.D. Behnke: Plant trichomes-structure and ultrastructure: general terminology, taxonomic applications, and aspects of Trichome bacterial interaction in leaf tips of Dioscorea. In: *Biology and Chemistry of Plant Trichomes*, ed. by E. Rodriguez, P.L. Healey, I. Mehta (Plenum, New York 1984) pp. 1–21

36.90 W. Barthlott, D. Hunt: Seed-diversity in *Cactaceae* subfam. Cactoideae. In: *Succulent Plant Research*, Vol. 5, ed. by D. Hundt (David Hunt, Milborne Port 2000)

36.91 W. Barthlott, B. Große-Veldmann, N. Korotkova: Orchid seed diversity: A scanning electron microscopy survey, Englera **32**, 1–244 (2014)

36.92 W. Barthlott, S. Wiersch, Z. Colic, K. Koch: Classification of trichome types within species of the water fern *Salvinia*, and ontogeny of the egg-beater trichomes, Botany **87**, 830–836 (2009) doi:10.1139/B09-048

36.93 J.N. Israelachvili: *Intermolecular and Surface Forces*, 2nd edn. (Academic, London 1992)

36.94 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

36.95 P.G. De Gennes, F. Brochard-Wyart, D. Quere: *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves* (Springer, New York 2004)

36.96 B. Bhushan, Y.C. Jung: Wetting, adhesion and friction of superhydrophobic and hydrophilic leaves and fabricated micro/nanopatterned surfaces, J. Phys. Condens. Matter **20**, 225010 (2008)

36.97 M. Nosonovsky, B. Bhushan: *Green Tribology, Biomimetics, Energy Conservation and Sustainability* (Springer, Berlin, Heidelberg 2012)

36.98 E.Y. Bormashenko: Wetting of real surfaces. In: *De Gruyter Studies in Mathematical Physics*, Vol. 19 (Walter de Gruyter, Berlin, Boston 2013) p. 187

36.99 H.J. Butt, K. Graf, M. Kappl: *Physics and Chemistry of interfaces*, 3rd edn. (Wiley, Weinheim 2013)

36.100 F. Schellenberger, N. Encinas, D. Vollmer, H.-J. Butt: How water advances on superhydrophobic surfaces, Phys. Rev. Lett. **116**, 096101 (2016)

36.101 B. Bhushan: *Nanotribology and Nanomechanics – An Introduction*, 3rd edn. (Springer, Berlin, Heidelberg 2011)

36.102 C.W. Extrand: Model for contact angle and hysteresis on rough and ultraphobic surfaces, Langmuir **18**, 7991–7999 (2002)

36.103 Y.C. Jung, B. Bhushan: Wetting behavior during evaporation and condensation of water microdroplets on superhydrophobic patterned surfaces, J. Microsc. **229**, 127–140 (2008)

36.104 B. Bhushan, Y.C. Jung: Wetting study of patterned surfaces for superhydrophobicity, Ultramicroscopy **107**, 1033–1041 (2007)

36.105 P. Roach, N.J. Shirtcliffe, M.I. Newton: Progress in superhydrophobic surface development, Soft Matter **4**, 224–240 (2008) doi:10.1039/b712575p

36.106 X. Zhang, F. Shi, J. Niu, Y. Jiang, Z. Wang: Superhydrophobic surfaces: from structural control to functional application, J. Mater. Chem. **18**, 621–633 (2008)

36.107 C.W. Extrand: Origins of wetting, Langmuir **32**(31), 7697–7706 (2016) doi:10.1021/acs.langmuir.6b01935

36.108 A.V. Adamson: *Physical Chemistry of Surfaces* (Wiley, New York 1990)

36.109 M. Nosonovsky, B. Bhushan: Lotus versus rose: Biomimetic surface effects. In: *Green Tribology, Biomimetics, Energy Conservation and Sustainability*, ed. by M. Nosonovsky, B. Bhushan (Springer, Berlin, Heidelberg 2012) pp. 25–40

36.110 R.N. Wenzel: Resistance of solid surfaces to wetting by water, Ind. Eng. Chem. **28**, 988 (1936)

36.111 A.B.D. Cassie, S. Baxter: Wettability of porous surfaces, Trans. Faraday Soc. **40**, 546–551 (1944) doi:10.1039/tf9444000546

36.112 A. Marmur: Wetting on hydrophobic rough surfaces: To be heterogeneous or not to be?, Langmuir **19**, 8343–8348 (2003)

36.113 E.Y. Bormashenko: Wetting transitions on biomimetic surfaces, Philos. Trans. R. Soc. A **368**, 4695–4711 (2010)

36.114 L. Feng, Y. Zhang, J. Xi, Y. Zhu, N. Wang, F. Xia, L. Jiang: Petal effect: A superhydrophobic state with high adhesive force, Langmuir **24**, 4114–4119 (2008)

36.115 A.-L. Yarin: Drop impact dynamics: Splashing, spreading, receding, bouncing..., Annu. Rev. Fluid Mech. **38**, 159–192 (2006)

36.116 C. Josserand, S.T. Thoroddsen: Drop impact on a solid surface, Annu. Rev. Fluid Mech. **48**, 365–391 (2015)

36.117 J. Fukai, Y. Shiiba, T. Yamamoto, O. Miyatake, D. Poulikakos, C.M. Megaridis, Z. Zhao: Wetting effects on the spreading of a liquid droplet surface: Experiment and modeling colliding with a flat, Phys. Fluids **7**, 236–247 (1995) doi:10.1063/1.868622

36.118 C. Mundo, M. Sommerfeld, C. Tropea: Droplet-wall collisions: Experimental studies of the deformation and breakup process, Int. J. Multiph. Flow **21**, 151–173 (1995) doi:10.1016/0301-9322(94)00069-V

36.119 T. Mao, D. Kuhn, H. Tran: Spread and rebound of liquid droplets upon impact on flat surfaces, AIChE J. **43**, 2169–2179 (1997) doi:10.1002/aic.690430903

36.120 L. Xu, L. Barcos, S.R. Nagel: Splashing of liquids: interplay of surface roughness withsurrounding gas, Phys. Rev. E **76**, 066311 (2007) doi:10.1103/PhysRevE.76.066311

36.121 R. Rioboo, C. Tropea, M. Marengo: Outcomes from a drop impact on solid surfaces, At. Sprays **11**, 155–165 (2001) doi:10.1615/AtomizSpr.v11.i2.40

36.122 C. Motzkus, F. Gensdarmes, E. Géhin: Study of the coalescence/splash threshold of droplet impact on liquid films and its relevance in assessing airborne particle release, J. Colloid Interface Sci. **362**, 540–552 (2011) doi:10.1016/j.jcis.2011.06.031

36.123 T. Gilet, L. Bourouiba: Rain-induced ejection of pathogens from leaves: Revisiting the hypothesis of splash-on-film using high-speed visualization, Integr. Comp. Biol. **54**(6), 974–984 (2014)

36.124 K. Koch, R. Grichnik: Influence of surface structure and chemistry on water droplet splashing, Philos Trans. R. Soc. A **374**, 20160183 (2016) doi:10.1098/rsta.2016.0183

36.125 Y. Liu, L. Moevius, X. Xu, T. Qian, J.M. Yeomans, Z. Wang: Pancake bouncing on superhydrophobic surfaces, Nat. Phys. **10**(7), 515–519 (2014)

36.126 K. Koch, M. Bennemann, H.F. Bohn, D.C. Albach, W. Barthlott: Surface microstructures of daisy florets (Asteraceae) and characterization of their anisotropic wetting, Bioinspir. Biomim. **8**, 036005 (2013)

36.127 H.-J. Ensikat, P. Ditsche-Kuru, C. Neinhuis, W. Barthlott: Superhydrophobicity in perfection: The outstanding properties of the lotus leaf, Beilstein J. Nanotechnol. **2**, 152–161 (2011) doi:10.3762/bjnano.2.19

36.128 W. Barthlott, K. Riede, M. Wolter: Mimicry and ultrastructural analogy between the semiaquatic grasshopper *Paulinia acuminata* (Orthoptera: Pauliniidae) and its foodplant, the water-fern *Salvinia auriculata* (Filicateae: Salviniaceae), Amazoniana **13**, 47–58 (1994)

36.129 H. Bargel, W. Barthlott, K. Koch, L. Schreiber, C. Neinhuis: Plant cuticles: Multifunctional inter-

faces between plant and environment. In: *The Evolution of Plant Physiology*, ed. by A.R. Hemsley, I. Poole (Academic, London 2003) pp. 171–194
36.130 L. Schreiber, J. Schonherr: *Water and Solute Permeability of Plant Cuticles* (Springer, Berlin, Heidelberg 2009)
36.131 K. Koch, W. Barthlott: Superhydrophobic and superhydrophilic plant surfaces: An inspiration for biomimetic materials, Philos. Trans. R. Soc. A **367**, 1487–1509 (2009) doi:10.1098/rsta.2009.002
36.132 W. Barthlott: Epidermal and seed surface characters of plants: Systematic applicability and some evolutionary aspects, Nordic J. Bot. **1**, 345–355 (1981) doi:10.1111/j.1756-1051.1981.tb00704.x
36.133 P.G. Kevan, M.A. Lanet: Flower petal microtexture is a tactile cue for bees, Proc. Natl. Acad. Sci. USA **82**, 4750–4752 (1985)
36.134 S.N. Gorb: *Functional Surfaces in Biology: Adhesion Related Phenomena*, Vol. 1 (Springer, Berlin, Heidelberg 2009)
36.135 F. Exner, S. Exner: Die physikalischen Grundlagen der Blütenfärbungen, Sitzungsber. Kais. Akad. Wiss. Wien, Math.-nat. Kl. I **119**, 191–245 (1910)
36.136 Y. Toda: Physiological studies on *Schistostega osmundacea* (Dicks.) Mohr, J. Coll. Sci. **40**(5), 1–30 (1918)
36.137 Q. Kay, H. Daoud, C. Stirton: Pigment distribution, light reflection and cell structure in petals, Bot. J. Linn. Soc. **83**(1), 57–84 (1981)
36.138 H.M. Whitney, M. Kolle, P. Andrew, L. Chittka, U. Steiner, B. Glover: Floral iridescence, produced by diffractive optics, acts as cue for animal pollinators, Science **323**, 130–133 (2009)
36.139 D.G. Lloyd, S.C.H. Barret: *Floral Biology – Studies on Floral Evolution in Animal-Pollinated Plants* (Chapman & Hall, New York 1996)
36.140 B. Burr, D. Rosen, W. Barthlott: Untersuchungen zur Ultraviolettreflexion von Angiospermenblüten. III. Dilleniidae und Asteridae s. l, Trop. Subtrop. Pflanzenwelt **93**, 186 (1995), (Akad. Wiss. Lit. Mainz, F. Steiner, Stuttgart 1995)
36.141 B. Burr, W. Barthlott: Untersuchungen zur Ultraviolettreflexion von Angiospermenblüten. II. Magnoliidae, Ranunculidae, Hamamelididae, Caryophyllidae, Rosidae, Trop. Subtrop. Pflanzenwelt **87**, 193 (1993), (Akad. Wiss. Lit. Mainz, F. Steiner, Stuttgart 1993)
36.142 N. Biedinger, W. Barthlott: Untersuchungen zur Ultraviolettreflexion von Angiospermenblüten. I. Monocotyledoneae, Trop. Subtrop. Pflanzenwelt **86**, 122 (1993), (Akad. Wiss. Lit. Mainz, F. Steiner, Stuttgart 1993)
36.143 H. Whitney, M. Kolle, R. Alvarez-Fernandez, U. Steiner, B. Glover: Contributions of iridescence to floral patterning, Commun. Integr. Biol. **2**(3), 230–232 (2009)
36.144 S. Robinson, C.E. Lovelock, C.B. Osmond: Wax as a mechanism for protection against photoinhibition: A study of *Cotyledon orbiculata*, Bot. Acta **106**, 307–312 (1993)
36.145 C. Müller, M. Riederer: Plant surface properties in chemical ecology, Chem. Ecol. **3**, 2621–2651 (2005)
36.146 J.R. Ehleringer, O. Björkman: Pubescence and leaf spectral characteristics in a desert shrub *Encelia farinosa*, Oecologia **36**, 151–162 (1978)
36.147 M.G. Holmes, D.R. Keiller: Effects of pubescence and waxes on the reflectance of leaves in the ultraviolet and photosynthetic wavebands: A comparison of a range of species, Plant Cell. Env. **25**, 85–93 (2002)
36.148 D.M. Gates: Energy exchange and transpiration. In: *Water and Plant Life*, Ecological Studies, Vol. 19, ed. by O.L. Lange, L. Kappen, E.D. Schulze (Springer, Berlin, Heidelberg 1976) pp. 137–147
36.149 P.H. Schuepp: Model experiments on free convection heat and mass transfer of leaves and plant elements, Bound.-Layer Meteorol. **3**, 454–457 (1973)
36.150 P.H. Schuepp: Leaf boundary layers, New Phytol. **125**, 477–507 (1993)
36.151 H.G. Jones, E. Rotenberg: Energy, radiation and temperature regulation in plants. In: *Encyclopedia of Life Sciences* (Wiley, New York 2001) pp. 1–8
36.152 W. Barthlott, W. Schultze-Motel: Zur Feinstruktur der Blattoberflächen und systematischen Stellung der Laubmoosgattung *Rhacocarpus* und anderer *Hedwigiaceae*, Willdenowia **11**, 3–11 (1981)
36.153 H.G. Edelmann, C. Neinhuis, M. Jarvis, B. Evans, E. Fischer, W. Barthlott: Ultrastructure and chemistry of the cell wall of the moss *Rhacocarpus purpurascens* (Rhacocarpaceae): A puzzling architecture among plants, Planta **206**, 315–321 (1998)
36.154 F.T. Malik, R.M. Clement, D.T. Gethin, W. Krawszik, A.R. Parker: Nature's moisture harvesters: A comparative review, Bioinspir. Biomim. **9**, 031002 (2014)
36.155 M.A.K. Azad, D. Ellerbrok, W. Barthlott, K. Koch: Fog collecting biomimetic surfaces: Influence of microstructure and wettability, Bioinspir. Biomim. **10**, 016004 (2015) doi:10.1088/1748-3190/10/1/016004
36.156 M.A.K. Azad, W. Barthlott, K. Koch: Hierarchical surface architecture of plants as an inspiration for biomimetic fog collectors, Langmuir **31**(48), 13172–13179 (2015) doi:10.1021/acs.langmuir.5b02430
36.157 P.T. Martone, L. Kost, M. Boller: Drag reduction in wave-swept macroalgae: Alternative stratgies and new predictions, Am. J. Bot. **99**, 8006–8815 (2012)
36.158 W. Barthlott, J. Bertling, P. Schoppa, C. Vogt: Lufthaltende Schiffsbeschichtungen nach biologischem Vorbild zur Reibungsreduktion, Fraunhofer UMSICHT, Design4Science (2011)
36.159 W. Barthlott: Die Selbstreinigungsfähigkeit pflanzlicher Oberflächen durch Epicuticularwachse. In: *Klima- und Umweltforschung an der Universität Bonn* (Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn 1992) pp. 117–120
36.160 C.W. Extrand, S.I. Moon: Repellency of the lotus leaf: Contact angles, drop retention, and sliding angles, Langmuir **30**, 8791–8797 (2014) doi:10.1021/la5019482
36.161 A. Balmert, H.F. Bohn, P. Ditsche-Kuru, W. Barthlott: Dry under water: Comparative

morphology and functional aspects of air-retaining insect surfaces, J. Morphol. **272**, 442–451 (2011) doi:10.1002/jmor.10921

36.162 R.S. Seymour, P.G.D. Matthews: Physical gills in diving insects and spiders: Theory and experiment, J. Exp. Biol. **216**, 164–170 (2013)

36.163 K. Koch, H.F. Bohn, W. Barthlott: Hierarchical sculpturing of plant surfaces and superhydrophobicity, Langmuir **25**, 14116–14120 (2009) doi:10.1021/la9017322

36.164 M. Amabili, A. Giacomello, S. Meloni, C.M. Casciola: Unraveling the *Salvinia* paradox: design principles for submerged superhydrophobicity, Adv. Mater. Interfaces **2**, 1500248 (2015) doi:10.1002/admi.201500248

36.165 P. Ditsche, E. Gorb, M. Mayser, S. Gorb, T. Schimmel, W. Barthlott: Elasticity of the hair cover in air-retaining *Salvinia* surfaces, Appl. Phys. A **121**, 505–511 (2015) doi:10.1007/s00339-015-9439-y

36.166 D. Gandyra, S. Walheim, S. Gorb, W. Barthlott, T. Schimmel: The capillary adhesion technique: A versatile method for determining the liquid adhesion force and sample stiffness, Beilstein J. Nanotechnol. **6**, 11–18 (2015) doi:10.3762/bjnano.6.2

36.167 W. Konrad, C. Apeltauer, J. Frauendiener, W. Barthlott, A. Roth-Nebelsick: Applying methods from differential geometry to devise stable and persistent air layers attached to objects immersed in water, J. Bion. Eng. **6**, 350–356 (2009) doi:10.1016/S1672-6529(08)60133-X

36.168 M.J. Mayser, W. Barthlott: Layers of air in the water beneath the floating fern *Salvinia* are exposed to fluctuations in pressure, Integr. Comp. Biol. **54**, 1001–1007 (2014) doi:10.1093/icb/icu072

36.169 M.J. Mayser, H.F. Bohn, M. Reker, W. Barthlott: Measuring air layer volumes retained by submerged floating-ferns *Salvinia* and biomimetic superhydrophobic surfaces, Beilstein J. Nanotechnol. **5**, 812–821 (2014) doi:10.3762/bjnano.5.93

36.170 A. Solga, Z. Cerman, B.F. Striffler, M. Spaeth, W. Barthlott: The dream of staying clean: Lotus and biomimetic surfaces, Bioinspir. Biomim. **2**, 126–134 (2007) doi:10.1088/1748-3182/2/4/S02

36.171 S.D. Eigenbrode: Plant surface waxes and insect behaviour. In: *Plant Cuticles: An Integrated Functional Approach*, ed. by G. Kerstiens (BIOS Scientific, Oxford 1996) pp. 201–222

36.172 R.G. Beutel, S.N. Gorb: Ultrastructure of attachment specializations of hexapods (Arthropoda): Evolutionary patterns inferred from a revised ordinal phylogeny, J. Zool. Syst. **39**(4), 177–207 (2001) doi:10.1046/j.1439-0469.2001.00155.x

36.173 S. Gorb: *Attachment Devices of Insect Cuticle* (Springer, Berlin, Heidelberg 2001)

36.174 H.F. Bohn, W. Federle: Insect aquaplaning: Nepenthes pitcher plants capture prey with the peristome, a fully wettable water-lubricated anisotropic surface, Proc. Natl. Acad. Sci. USA **101**(39), 14138–14143 (2004)

36.175 U. Bauer, H.F. Bohn, W. Federle: Harmless nectar source or deadly trap: *Nepenthes* pitchers are activated by rain, condensation and nectar, Proc. R. Soc. B **275**, 259–265 (2008)

36.176 J. Gould: Learning from nature's best, Nature **519**, S2–S3 (2015) doi:10.1038/519S2a

36.177 A.J. Schulte, D.M. Droste, K. Koch, W. Barthlott: Hierarchically structured superhydrophobic flowers with low hysteresis of the wild pansy (*Viola tricolor*) – New design principles for biomimetic materials, Beilstein J. Nanotechnol. **2**, 228–236 (2011) doi:10.3762/bjnano.2.27

36.178 M. Hopkin: Butterflies boast ultrablack wings: Insects use optical trick to get the blackest black out of dark pigments, Nat. Sci. Update **28**, 040126–4 (2004)

36.179 M. Spinner, S.N. Gorb, A. Balmert, H. Bleckmann, G. Westhoff: Non-contaminating camouflage: Multifunctional skin microornamentation in the West African Gaboon viper (*Bitis rhinoceros*), PLoS ONE **9**, e91087 (2014) doi:10.1371/journal.pone.0091087

36.180 M.A.K. Azad: *Fog Collection on Plant Surfaces and Biomimetic Applications*, Ph.D. Thesis (University of Bonn, Bonn 2016)

36.181 M.A.K. Azad, T. Krause, L. Danter, A. Baars, K. Koch, W. Barthlott: Fog collection on polyethylene terephthalate (PET) fibers: Influence of cross section and surface structure, Langmuir (2017) doi:10.1021/acs.langmuir.7b00478

36.182 M. Nosonovsky, B. Bhushan: Energy transitions in superhydrophobicity: Low adhesion, easy flow and bouncing, J. Phys. Condens. Matter **20**, 395005 (2008)

36.183 G. Zouridakis, J.E. Moore, J. Maitland: *Biomedical Technology and Devices*, 2nd edn. (CRC, Boca Raton 2013)

36.184 D.J. Babu, M. Mail, W. Barthlott, J.J. Schneider: Superhydrophobic vertically aligned carbon nanotubes for biomimetic air retention under water (Salvinia effect), Adv. Mater. Interfaces (2017) doi:10.1002/admi.201700273

36.185 C.J. Weng, C.H. Chang, C.W. Peng, S.W. Chen, J.M. Yeh, C.L. Hsu, Y. Wei: Advanced anticorrosive coatings prepared from the mimicked xanthosoma sagittifolium-leaf-like electroactive epoxy with synergistic effects of superhydrophobicity and redox catalytic capability, Chem. Mater. **23**(8), 2075–2083 (2011)

36.186 S. Farhadi, M. Farzaneh, S.A. Kulinich: Anti-icing performance of superhydrophobic surfaces, Appl. Surf. Sci. **257**, 6264–6269 (2011) doi:10.1016/j.apsusc.2011.02.057

36.187 L. Cao, A.K. Jones, V.K. Sikka, J. Wu, D. Gao: Anti-icing superhydrophobic coatings, Langmuir **25**, 12444–12448 (2009) doi:10.1021/la902882b

36.188 Y.Y. Yan, N. Gao, W. Barthlott: Mimicking natural superhydrophobic surfaces and grasping the wetting process: A review on recent progress in preparing superhydrophobic surfaces, Adv. Colloid Interface Sci. **169**, 80–105 (2011) doi:10.1016/j.cis.2011.08.005

36.189 R. Fürstner, W. Barthlott, C. Neinhuis, P. Walzel: Wetting and self-cleaning properties of artificial

superhydrophobic surfaces, Langmuir **21**, 956–961 (2005) doi:10.1021/la0401011

36.190 K. Koch, B. Bhushan, Y.C. Jung, W. Barthlott: Fabrication of artificial Lotus leaves and significance of hierarchical structure for superhydrophobicity and low adhesion, Soft Matter **5**(7), 1386–1393 (2009)

36.191 M. Schwab, G. Noga, W. Barthlott: Bedeutung der Epicuticularwachse für die Pathogenabwehr am Beispiel von *Botrytis cinerea* – Infektionen bei Kohlrabi und Erbse, Gartenbauwissenschaft **60**, 102–109 (1995)

36.192 G.J. Noga, M. Knoche, M. Wolter, W. Barthlott: Changes in leaf micro-morphology induced by surfactant application, Angew. Bot. **61**, 521–528 (1987)

36.193 M. Wolter, W. Barthlott, M. Knoche, G.J. Noga: Concentration effects and regeneration of epicuticular waxes after treatment with Triton-X-100 surfactant, Angew. Bot. **62**, 53–62 (1988)

36.194 G. Noga, M. Wolter, W. Barthlott, W. Petry: Quantitative evaluation of epicuticular wax alterations as induced by surfactant treatment, Angew. Bot. **65**, 239–252 (1991)

36.195 G. Noga, M. Knoche, M. Wolter: The Impact of Triton X-100 surfactant on leaf micromorphology, Hortscience **23**(3), 808 (1988)

36.196 C. Neinhuis, M. Wolter, W. Barthlott: Epicuticular wax of *Brassica oleracea*: Changes in microstructure and ability to be contaminated of leaf surfaces after application of Triton X-100, Z. Pflanzenkrankh. Pflanzenschutz **99**, 542–549 (1992)

36.197 C. Zeiger, I.C.R. da Silva, M. Mail, M.N. Kavalenka, W. Barthlott, H. Hölscher: Microstructures of superhydrophobic plant leaves-inspiration for efficient oil spill cleanup materials, Bioinspir. Biomim. **11**(5), 056003 (2016)

36.198 M.A. Salem, W. Al-Zayadneh, H.F. Schulze, A.J. Cheruth: Effect of nanohydrophobic sand layer on Bermudagrass (*Cynodon* spp.) in urban landscaping, Urban Water J. **11**, 167–173 (2013) doi:10.1080/1573062X.2013.768684

36.199 J.-E. Melskotte, M. Brede, A. Ott, M. Mayser, W. Barthlott, A. Leder: Künstliche Luft haltende Oberflächen zur Reibungsreduktion am Schiff/Artificial air retaining surfaces for drag reduction on shiphulls. In: *Lasermethoden in der Strömungsmesstechnik*, Vol. 39, ed. by M. Brede, B. Ruck, D. Dopheide (Deutsche Gesellschaft für Laser-Anemometrie GALA e.V., Karlsruhe 2012) pp. 1–6

36.200 O. Tricinci, T. Terencio, B. Mazzolai, N.M. Pugno, F. Greco, V. Mattoli: 3D micropatterned surfaces inspired by *Salvinia molesta* via direct laser lithography, ACS Appl. Mater. Interfaces **7**(46), 25560–25567 (2015) doi:10.1021/acsami.5b07722

36.201 J.-E. Melskotte, M. Brede, A. Wolter, W. Barthlott, A. Leder: Schleppversuche an künstlichen, Luft haltenden Oberflächen zur Reibungsreduktion am Schiff. In: *Lasermethoden Strömungsmesstech., 21. Fachtag., München, 2013 (Tagungsbd.)*, ed. by C.J. Kähler, R. Hain, C. Cierpka, B. Ruck, A. Leder, D. Dopheide (Dt. Ges. Laser-Anemometrie GALA e.V., Karlsruhe 2013) pp. 53-1–51-7

36.202 P. Ditsche-Kuru, E.S. Schneider, J.-E. Melskotte, M. Brede, A. Leder, W. Barthlott: Superhydrophobic surfaces of the water bug *Notonecta glauca*: A model for friction reduction and air retention, Beilstein J. Nanotechnol. **2**, 137–144 (2011) doi:10.3762/bjnano.2.17

36.203 W. Federle, U. Maschwitz, B. Fiala, M. Riederer, B. Hölldobler: Slippery ant-plants and skilful climbers: Selection and protection of specific ant partners by epicuticular wax blooms in *Macaranga* (Euphorbiaceae), Oecologia **112**, 217–224 (1997)

36.204 P. Kim, M.J. Kreder, J. Alvarenga, J. Aizenberg: Hierarchical or not? Effect of the length scale and hierarchy of the surface roughness on omniphobicity of lubricant-infused substrates, Nano Lett. **13**(4), 1793–1799 (2013) doi:10.1021/nl4003969

36.205 T.S. Wong, S.H. Kang, S.K. Tang, E.J. Smythe, B.D. Hatton, A. Grinthal, J. Aizenberg: Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity, Nature **477**, 443–447 (2011) doi:10.1038/nature10447

36.206 C.A.E. Hamlett, N.J. Shirtcliffe, F.B. Pyatt, M.I. Newton, G. McHale, K. Koch: Passive water control at the surface of a superhydrophobic lichen, Planta **234**(6), 1267–1274 (2011) doi:10.1007/s00425-011-1475-z

36.207 W. Barthlott, W. Erdelen, M.D. Rafiqpoor: Biodiversity and technical innovations: bionics. In: *Concept and Value in Biodiversity. Routledge Studies in Biodiversity Politics and Management*, ed. by D. Lanzerath, M. Friele (Routledge, Oxon, New York 2014) pp. 300–315

37. Bioinspired Nanostructured Antibiofouling and Anti-inorganic Surfaces

Bharat Bhushan

In this chapter, biofouling and inorganic fouling are introduced. I then summarize fields susceptible to fouling followed by discussion of biofouling and inorganic fouling formation mechanisms, antifouling strategies from living nature, and current prevention and cleaning techniques. Finally, demonstrations of the antifouling uses of bioinspired rice leaf surfaces are presented. The data shows that micropatterned surfaces facilitate antifouling.

37.1 **Fields Susceptible to Fouling** 1307
37.2 **Biofouling and Inorganic Fouling Formation Mechanisms** 1310
37.2.1 Biofouling Formation 1310
37.2.2 Inorganic Fouling Formation 1310
37.2.3 Surface Factors 1311
37.3 **Antifouling Strategies from Living Nature** 1313
37.4 **Antifouling: Current Prevention and Cleaning Techniques** 1315
37.4.1 Prevention Techniques 1315
37.4.2 Self-Cleaning Surfaces and Cleaning Techniques 1318
37.5 **Bioinspired Rice Leaf Surfaces for Antifouling** 1318
37.5.1 Fabrication of Micropatterned Samples 1318
37.5.2 Antibiofouling Measurements 1319
37.5.3 Anti-inorganic Fouling Measurements . 1320
37.5.4 Results and Discussion 1320
37.5.5 Antibiofouling and Anti-inorganic Fouling Mechanisms 1323
37.6 **Closure** 1324
References 1324

Fouling is generally undesirable for most applications [37.1, 2]. Fouling includes biological fouling (commonly referred to as biofouling) and inorganic fouling. Biofouling is the accumulation of unwanted biological matter on surfaces, with biofilms created by microorganisms and macroscale biofouling (simply called macrofouling) created by macroorganisms. Inorganic fouling results from the accumulation of particles including deposits of dirt, corrosion, crystals, suspended particles, oil, and ice. The type and extent of fouling depends on the environment, inorganic deposits, and organisms. In biomedical contexts, fouling typically takes the form of biofilms [37.3–7], whereas marine and industrial biofouling typically includes macrofouling [37.8–11] and inorganic fouling [37.12–15].

Bixler and *Bhushan* [37.16] presented a review of the field. On the basis of the review, we begin with discussion of fields susceptible to fouling, followed by biofouling and inorganic fouling formation mechanisms, antifouling strategies from nature, current prevention and cleaning techniques in antifouling, and end with demonstrations of the use of bioinspired rice leaf surfaces for antifouling, based on *Bixler* et al. [37.17].

37.1 Fields Susceptible to Fouling

Fouling is prevalent in the medical, marine, and industrial fields, causing significant health risks and loss of resources [37.4, 9–11]. Examples of various types and problems incurred through fouling are shown in Table 37.1; the Table highlights the many applications susceptible to biofouling and inorganic fouling [37.16].

Figure 37.1 shows that biofouling in the human body can occur in regions such as prosthetic implants, biosensors, catheters, dental implants, and medical equipment [37.16]. Biofouling may lead to implant rejection, malfunction of biosensors, and spread of infectious diseases [37.4–7]. Biofilms on medical devices including a ventilation tube, needleless connector, and pacemaker wire are shown in Fig. 37.2 [37.16]. Biofouling can also occur in marine environments where noticeable marine growth

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_37

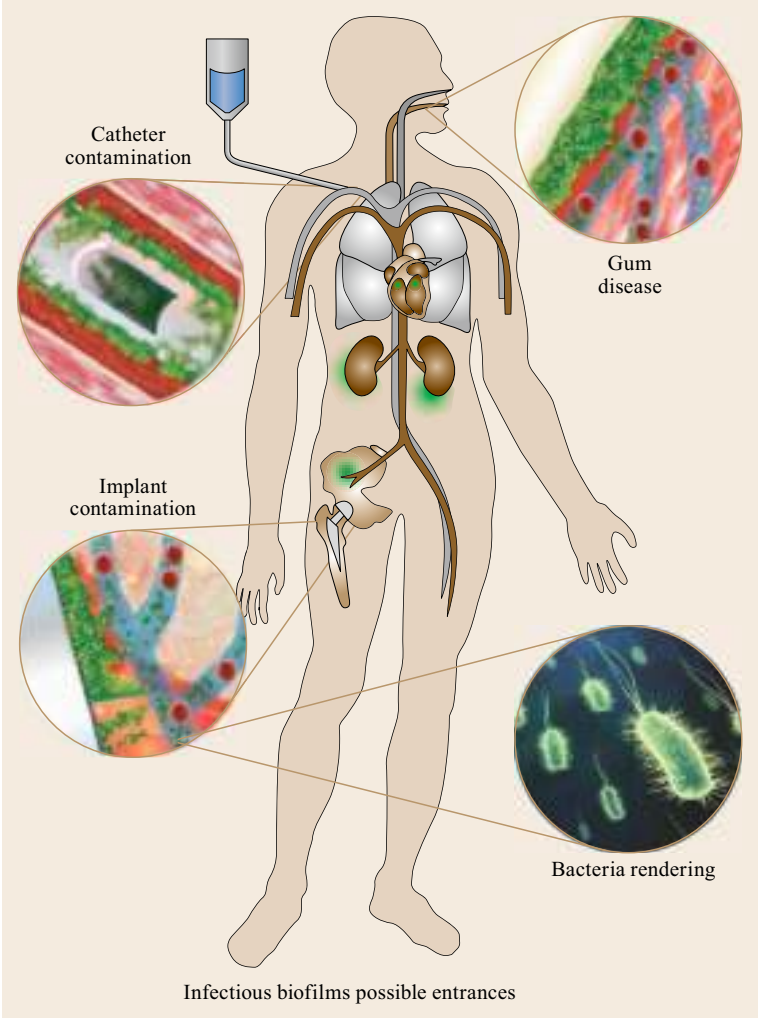


Fig. 37.1 Schematic highlighting areas and entry points of the body susceptible to infectious microorganisms leading to formation of biofilms (after [37.16])

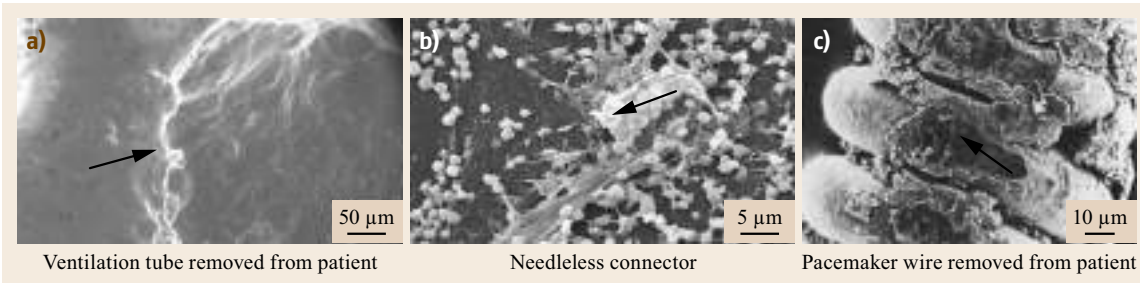


Fig. 37.2a–c Examples of biofilms affecting medical devices (after [37.16]). Images include a ventilation tube **(a)** [37.18], needleless connector **(b)** (adapted from Janice Carr), and a pacemaker wire **(c)** (after [37.19]). Such biofilms may lead to infections that require device removal since antibodies are often ineffective

appears on ships and underwater structures. This magnifies ship drag, corrosion, fuel consumption, and engine stress [37.10, 15, 20–22]. Industrial biofouling occurs in areas such as power plants, water treatment systems, and food or beverage industries, as shown in Fig. 37.3 [37.16]. Issues include pipe blockage, decreased reverse osmosis (RO) membrane flux, contaminated water, and reduced heat exchanger efficiency [37.7, 9, 12].

Inorganic fouling creates concern in many industries with decreased performance in applications ranging from power production to consumer products. For example, the interior surface of a heat exchanger tube from a nuclear reactor is shown in Fig. 37.4 [37.16], and corrosion deposits are readily apparent, which leads to inefficient performance [37.23], www.hcheattransfer.com. Inorganic fouling also includes undesirable fingerprints on transparent surfaces such as electronic touch screens (mobile phones, tablets, music players, GPS devices), eyeglasses, computer screens, and window glass [37.26]. Furthermore, inorganic fouling includes ice on airplane wings, helicopter blades, oil platforms, power lines, locks and dams, and wind turbines [37.27, 28].

Table 37.1 Fields susceptible to biofouling including common examples (adapted from [37.16])

Type	Problems
Medical	
Orthopedic implant	Removal due to infection
Respirator	Ventilator-associated pneumonia
Contact lens	Eye infection
Catheter	Urinary tract infections
Hemodialysis	Infectious break-outs
Teeth/dental implant	Periodontal disease, gingivitis
Biosensor	Failure from fibrous encapsulation
Marine	
Ship hull	Increased fuel consumption
Ship engine	Increased stress from extra drag
Marine platform	Increased marine structure load/stress/fatigue
Metal	Increased biocorrosion
Industrial	
Membrane	Reduced flux
Fluid flow	Frictional loss in pipes
Drinking water	Pathogens in potable water
Food, paper, and paint	Food spoilage and worker health risks
Metal cutting fluid	Filter blockage and worker health risks

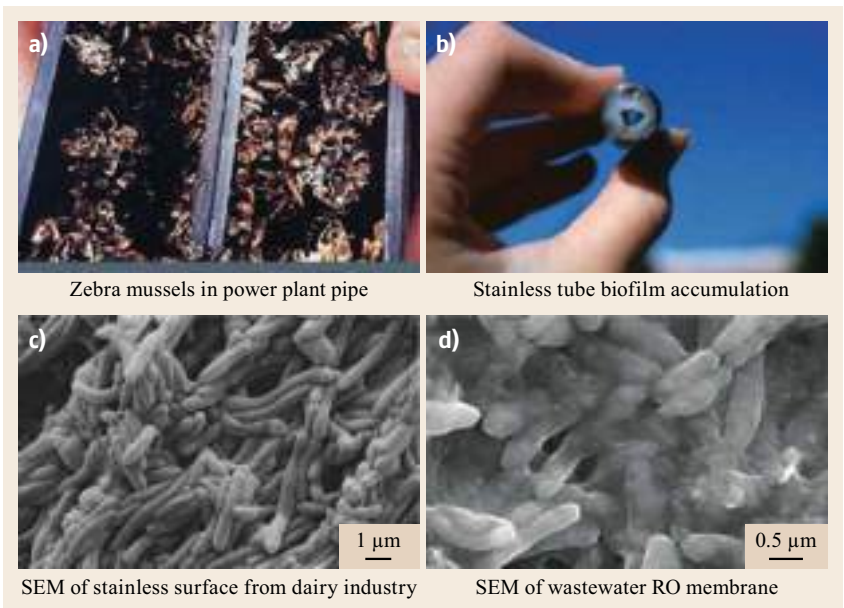


Fig. 37.3a–d Examples of industrial biofouling (after [37.16]). Zebra mussels (*Dreissena polymorpha*) clogging a power plant intake pipe **(a)** (photo courtesy of Peter Yates), biofilm accumulation in a stainless steel tube **(b)** (after [37.23]), a biofilm (*B. cereus*) on a dairy industry stainless steel plate **(c)** (after [37.24]), and a biofilm (*P. aeruginosa*) on a reverse osmosis (RO) membrane **(d)** (after [37.25])

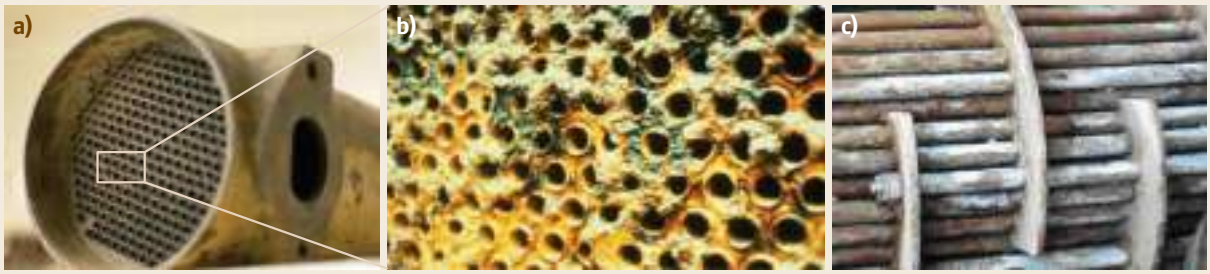


Fig. 37.4a–c Examples of inorganic fouling (after [37.16]). Inside of a clean nuclear power plant heat exchanger tube **(a)** compared to a fouled tube with corrosion deposits **(b)** (after [37.23]). Calcium carbonate crystallization fouling on the outside of a heat exchanger **(c)** (photo courtesy of H&C Heat Transfer Solutions, Inc)

37.2 Biofouling and Inorganic Fouling Formation Mechanisms

In this section, we discuss the type, formation, and morphology characteristics of biofouling and inorganic fouling (based on [37.16]).

37.2.1 Biofouling Formation

In biofilms, microorganisms adhere to one another and the substrate with an adhesive called the extracellular polymeric substance (EPS). The biofilm continues to grow and become more diverse by attracting additional microorganisms through chemical *messages*. Colonization is the process of organisms collecting and growing on a surface, and the morphology is characterized by the thickness, density, structure, composition, bioadhesive strength, and weight of the fouling organisms [37.7, 11, 18, 19, 29, 30]. In the marine environment, this usually starts with the formation of a biofilm that attracts larger macrofoulers [37.3, 10, 11, 31]. For instance, tubeworms prefer to settle on biofilms, whereas bryozoans and barnacles do not require a biofilm [37.32]. Figure 37.5 illustrates the five-stage colonization process, which includes initial attachment, irreversible attachment, initial growth, final growth, and dispersion (adapted from D. Davis, [37.16]). Initial attachment starts the colonization process, which begins within a few days to a few weeks in a nutrient-rich environment. The biofilm covered surface then attracts other organisms that may have been previously deterred. Initial attachment of microorganisms is reversible, but once they secrete the EPS, the bond becomes irreversible. This permanent attachment allows initial growth, final growth, and dispersion [37.3, 10, 31, 33].

Biofouling growth rates depend on the organism, substrate, flow velocity, shear stress, and temperature [37.10, 34, 35]. Organism transportation to a surface is either passive from the current or active from the propagule, juvenile, or adult organism. Active mechanisms include electrostatic repulsion, Brownian motion, turbulent pulsations, and cell outgrowths [37.10, 32]. Underwater environments are ideal for biofouling since currents deliver nutrients and carry away wastes, promoting colonization by planktonic and sessile organisms.

The tenacity of the biofouling is dependent upon bioadhesion [37.3, 10, 31, 33]. Bioadhesion is the adhesion strength of biofouling on a hard surface [37.36]. This depends on the organism type, the substrate, and the separating fluid [37.37], and is related to influences of electrostatic forces and surface wettability [37.38–41]. Biofilm bioadhesion is a two-stage process, starting with the initial attachment and then the irreversible attachment. Initial attachment is controlled by a physical adhesion between the microorganism and the substrate. Also known as adsorption, the initial colonists attach to a surface through weak, reversible van der Waals bonds, which are slightly stronger than the electrostatic repulsive force. Irreversible attachment is accomplished with secretion of the EPS, which exhibits a sponge-like matrix. This adhesive permanently bonds the microorganisms to one another and collectively to the surface [37.3, 10, 42].

37.2.2 Inorganic Fouling Formation

Inorganic fouling is composed of nonbiological particles that may form in addition to or independently of biofouling. Particles originate from corrosion, crystallization, suspended particles, oil, and ice. For instance, salts from aqueous solutions crystallize and deposit

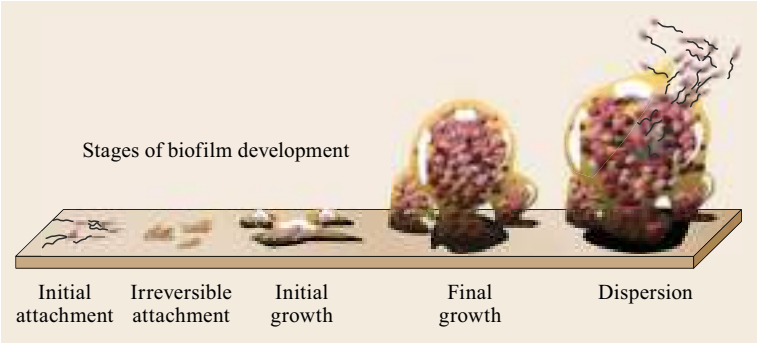


Fig. 37.5 Biofilm development process. Schematic illustrates the five-stage colonization process (adapted from D. Davis)

on surfaces. Other deposits may result from minerals found in water such as magnesium, calcium, and barium [37.7, 9, 12, 16].

Types of inorganic fouling include particulate, freeze, and gas stream particulate fouling. Particulate fouling occurs when suspended solid particles deposit onto a heat transfer surface [37.43]. Deposition of crystals from freeze fouling occurs in locations such as cold-region oil pipelines when waxy hydrocarbons make contact with cold pipe walls. Gas stream particulate fouling occurs in gas lines, reactors, combustion chambers, and heat exchangers. This includes mineral, organic, and inorganic particles, which are common in oil or gas combustion systems [37.44]. Biofouling may initiate inorganic fouling, where biocorrosion causes the formation of corrosion particles. Such fouling is prevalent in boilers, cooling condensers, desalination plants, food-processing equipment, geothermal plants, and oil-production equipment [37.14]. Heat exchangers can develop hard deposits called *scale* or more porous deposits such as *sludge* [37.13].

37.2.3 Surface Factors

Biofouling and inorganic fouling depend on various surface factors including wettability, microtexture, and contours [37.8, 10, 11, 16, 31, 45]. For instance, bryozoan and mussel larvae prefer hydrophobic surfaces [37.46]; hydroids, bryozoans, and ascidians prefer microtextured surfaces; larvae, sponges, barnacles, ascidians, and mollusks prefer light-colored surfaces; barnacles prefer convex contours; and calcareous sponges prefer concave contours [37.10].

Surface wettability influences fouler colonization, with surfaces ranging from superhydrophobic to superhydrophilic [37.8, 10, 11, 31, 45, 47]. Water droplets on a hydrophobic surface will *bead up*, whilst droplets on a hydrophilic surface will spread out evenly.

Low adhesive, superhydrophobic surfaces promote contaminant removal with the application of liquid through *self-cleaning*. Sheeting action of sliding liquid on a superhydrophilic surface can promote self-cleaning, where the liquid helps collect and remove fouling and disrupt microorganism settlement [37.48]. In general though, microorganisms prefer to colonize hydrophilic surfaces, although some actually prefer hydrophobic surfaces. For example, *Ulva linza* prefer hydrophobic surfaces, whereas *Balanus amphitrite* prefer hydrophilic surfaces. Furthermore, microorganisms such as *Ulva linza* attach more easily to hydrophobic surfaces, but experience decreased adhesive strength [37.40].

Surface microtexture influences organisms such as hydroids, bryozoans, and ascidians that seek shelter against strong currents by settling in grooves, pits, cracks, and crevices [37.8, 10]. Microorganisms prefer to settle in areas slightly larger than themselves for maximum protection and surface area contact with the substrate. The antifouling occurs when the gap between features is less than the approximate size of fouling microorganisms [37.16, 40, 49–52]. Fewer attachment points between the microorganism and substrate translate into lower bioadhesive strength [37.53]. Bioadhesion is also affected by the effectiveness of EPS flowing into crevices formed by surface roughness, which depends on the adhesive viscosity. Furthermore, when the adhesive only contacts the surface asperity peaks or tops of microtexture ridges, the applied force to break bonds is significantly reduced [37.54].

Several bioassay antifouling experiments have been performed with different microorganisms and microstructured surfaces – using pillars, ribs, and a combination thereof with various diameters, widths, lengths, heights, and pitches [37.17, 49–53, 55, 56]. Laboratory research suggests that samples with appropriately sized topographies prevent colonization by various microorganisms including green algae *Ulva* spores [37.49, 51], porcine cardiovascular endothelial cells [37.49], and *Balanus amphitrite* cyprids [37.50]. Such microorganisms have been exposed to samples and the antifouling effectiveness is determined by counting settled cells using light [37.49] and fluorescence microscopy [37.51].

Up to 85% spore settlement reduction is reported with microsized ribs when compared to flat control samples [37.49]. This reduction is attributed to the microsized feature spacing that is slightly smaller than the microorganisms under investigation. The so-called engineered antifouling microtopographies are designed to target specific microorganisms based on their shape and size [37.50, 51]. *Bixler* et al. [37.17] reported that reduction in biofouling occurs even if the microsized feature spacing is larger than the microorganisms. Sharp feature edges discourage the settling and eventual colonization of microorganisms on the surface.

Figure 37.6 shows surface topography antifouling mechanisms demonstrating wettability and texture properties that influence microorganism settlement and colonization [37.16]. Microorganisms search for an ideal site to settle, which can depend on parameters such as wettability and texture [37.38, 40]. During the search process, microorganisms will settle, move on, or expire in the process. In general, biofouling on superhydrophobic surfaces wash off more easily (self-cleaning) as compared to superhydrophilic surfaces. The images at the bottom of Fig. 37.6 show a variety of examples found in nature, including the hydrophilic elephant ear (*Alocasia odora*) [37.39], microtextured shark skin (*Carcharhinus galapagensis*) [37.57], hydrophobic water fern (*Regnellidium diphyllum*) [37.39], and superhydrophobic nanocolumns on an insect wing (*Cicada orni*) [37.58].

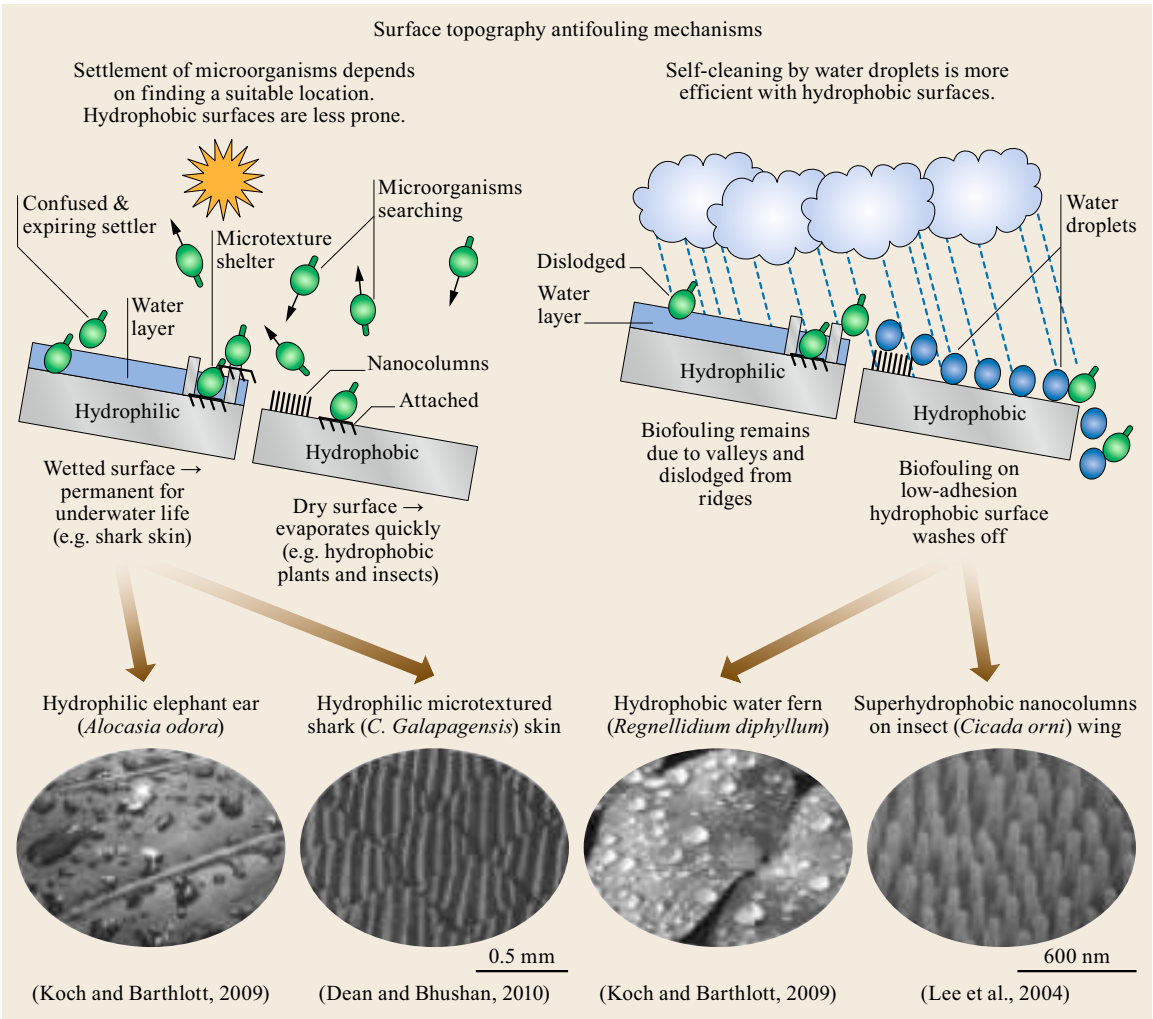


Fig. 37.6 Schematic shows surface topography antifouling mechanisms demonstrating wettability and texture properties that influence microorganism settlement and colonization (adapted from [37.16]; *Koch* and *Barthlott*, 2009 [37.39]; *Dean* and *Bhushan*, 2010 [37.57]; *Lee* et al., 2004 [37.58])

37.3 Antifouling Strategies from Living Nature

In living nature, many plants and animals thrive due to their antifouling and self-cleaning characteristics [37.16, 32, 39, 59–62]. *Bixler* and *Bhushan* [37.16] provide a number of examples of antifouling strategies from living nature, including a variety of both ambient and aquatic species. Common antifouling mechanisms include wettability, low drag, low adhesion, microtexture, grooming/mechanical cleaning, and chemical secretions [37.16].

Desirable wettability properties include superhydrophobic and superoleophobic low-adhesion self-cleaning surfaces that resist fouling, as liquids wash away contaminant particles [37.16]. Figure 37.7a shows various superhydrophobic examples including the lotus leaf, rice leaf, butterfly wings, and hydrophobic pigeon feathers. Figure 37.7b shows examples of superoleophobic-in-water fish scales and shark skin, which exhibit low drag during swimming.

Superhydrophobic lotus leaves exhibit antifouling with self-cleaning properties [37.64, 65]. It is believed that since lotus plants thrive in humid, marshy environments, self-cleaning prevents unwanted fouling, which may inhibit photosynthesis. Lotus leaf hierarchical micropapillae, which are microbumps superimposed with low surface energy waxy nanotubules, provide superior self-cleaning properties [37.64, 66]. In nature, rain droplets impact the lotus leaf surface and effectively roll off due to the water repellency effect. This leads to self-cleaning, where droplets collect and remove any contaminant particles (Fig. 37.8a). It should be noted that many other plant species also exhibit self-cleaning properties due to superhydrophobic and low adhesion characteristics.

Fish skin and the skin of fast-swimming sharks exhibit low drag, antifouling, and self-cleaning properties [37.57, 67]. Fish skin is covered by oriented scales with concentric rings overlapping and hinged such that water flow is from head to tail. Sharks exhibit low drag skin in order to swim quickly to catch prey, and thus antifouling is necessary in order to maintain the low drag. Shark skin consists of microstructured riblets on dermal denticles to reduce drag and prevent biofouling, and in comparison to other aquatic species such as whales, which are often covered with barnacles, sharks remain clean (Fig. 37.8b) [37.16, 67].

Rice leaves and butterfly wings exploit a combination of antifouling properties from the shark skin and lotus leaf [37.17, 63]. It is believed that since rice plants thrive in humid, marshy environments, self-cleaning prevents unwanted fouling, which may inhibit photosynthesis. Since butterflies are fragile and unable to clean their wings, such properties are critical to maintain structural coloration and flight control [37.68]. Hierarchical structures consisting of micropapillae superimposed with waxy nanobumps in rice leaves and microgrooves on top of shingle-like scales in butterfly wings provide superhydrophobicity and low adhesion exhibiting the lotus effect. Sinusoidal grooves in rice leaves and aligned, shingle-like scales in butterfly wings provide anisotropic flow exhibiting the shark skin effect. A combination of anisotropic flow, superhydrophobicity, and low adhesion leads to drag reduction and antifouling. Because of these anisotropic flow characteristics, droplets roll down the blade of the rice leaf and axially away from the butterfly body [37.63, 69]. In addition, the hierarchical structure reduces coverage and colonization of biofouling products [37.17].

Pigeon feathers are hydrophobic and remain clean [37.70]. The feathers are formed by barbs and barbules. The hydrophobic properties of the outer feathers (pennae) are due to a combination of this complex feather microstructure and the presence of preen oil (preen wax) spread mechanically by the bird during the preening procedure. Air pockets are then trapped by the feather microstructure, responsible for hydrophobicity.

Certain microtextured surfaces resist biofouling as organisms seek ideal surface features for settlement, and may be deterred if no suitable surface is found. Various examples are shown in Fig. 37.6.

Grooming/other mechanical cleaning is the physical removal of biofouling from the host, which effectively controls slow- and fast-growing biofouling. Decapods and crustaceans groom other creatures with special brush structures which remove foulers from gills and appendages. Echinoderms and bryozoans use special structures called pedicellaria to groom macroepibionts, while crayfish depend on *Branciobdellid annelids* to feed on foulers in their gills [37.16].

Any physical action that prevents or removes fouling may be present in conjunction with other antifouling mechanisms, such as low drag, low adhesion, and chemical secretions. Figure 37.9a illustrates a mechanical cleaning example, combining wiping with chemical secretions, as demonstrated in the human eye via the lachrymal gland secreting tears, puncta collecting tears, and the eyelid wiping the cornea [37.71]. Chemical secretion methods range from preventing to removing biofouling. Figure 37.9b shows the red seaweed (*Delisea pulchra*), which uses a halogenated furanone to manipulate colonizing bacteria *attraction* messages [37.72]. Other chemical examples include

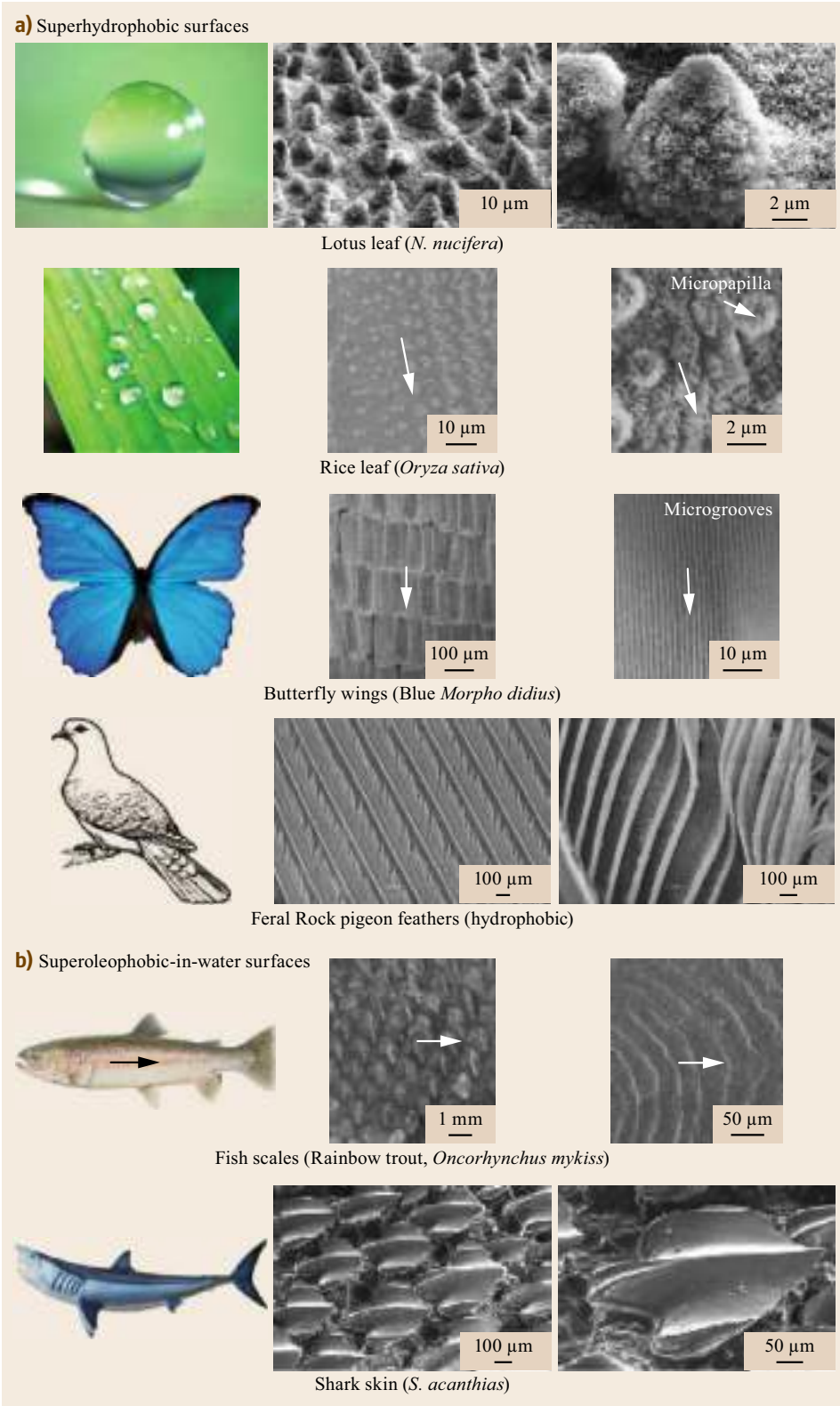


Fig. 37.7a,b Examples of antifouling in nature in regard to wettability properties. Water promotes self-cleaning on low-adhesion superhydrophobic surfaces, as shown **(a)** in the lotus leaf, rice leaf, butterfly wings, and pigeon feather (hydrophobic) examples (after [37.16]). **(b)** Fish scales and shark skin are superoleophobic when submerged in water, and exhibit low drag and antifouling. Reprinted from [37.63] with permission of Taylor and Francis Ltd.



Fig. 37.8a–c Shark skin and lotus leaf effects. **(a)** The lotus effect is illustrated by a water droplet carrying off contaminants from the lotus leaf surface [37.66]. These images show biofouling on a whale **(b)** and clean shark skin **(c)**, even though both live in similar environments. The inset SEM image shows microstructured riblets on Mako shark skin which provide low drag and antifouling properties (reprinted from [37.67] with permission from Wiley)

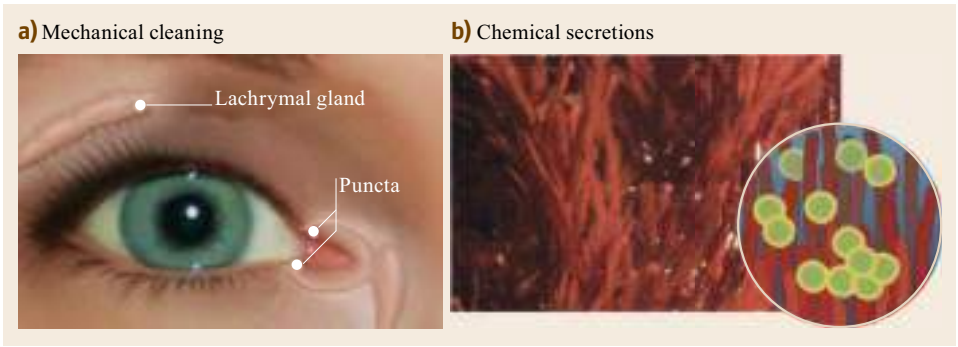


Fig. 37.9a,b Antifouling examples from nature using mechanical cleaning and chemical secretions. **(a)** Human eye combines wiping with chemical secretions to clean the cornea. **(b)** Red seaweed using a halogenated furanone to manipulate colonizing bacteria *attraction* messages [37.16]

snails that leave a predatory mucous trail, antimicrobial coral egg shells, mucus on shark skin, and antifouling chemicals produced by the bacteria *Roseobacter gallaeciensis* [37.16]. Furthermore, blood cell protein adsorption and cell adhesion are inhibited with phosphorylcholine [37.16].

37.4 Antifouling: Current Prevention and Cleaning Techniques

Bixler and *Bhushan* [37.16] report that reducing or eliminating fouling is accomplished through a variety of methods in medical, marine, and industrial applications. Common controls exist, such as microorganism-resistant coatings and self-cleaning low-adhesion surfaces. Improving biocompatibility can also reduce medical biofouling, since microorganisms will be less likely to target any implanted device.

37.4.1 Prevention Techniques

Medical biofouling is of great concern for the health of patients [37.16]. Fouling of medical devices can significantly alter the properties of the fouled materials. Multiple coatings and antifouling techniques have been reported [37.24], and many incorporate hydrophobic polymers with antimicrobial properties [37.16, 73]. Example applications include catheters, endotracheal tubes [37.74], and orthopedic hip implants. Superhydrophobic and low-adhesion *self-cleaning* surfaces are believed to minimize protein bioadhesion, and antimicrobial surfaces reduce ambient microorganisms [37.33]. Surfaces with superhydrophilic polyethylene glycol or polyethylene oxide exhibit less protein adsorption and bioadhesion [37.54, 75]. Other coating examples include self-assembled monolayers (SAMs), which are desirable since they coat surfaces with a molecularly thick biocompatible film. This is especially important for applications such as biomedical micro/nanoelectromechanical systems (BioMEMS/NEMS) [37.1, 76]. Figure 37.10a from *Bixler* and *Bhushan* [37.16] shows an orthopedic hip implant cov-

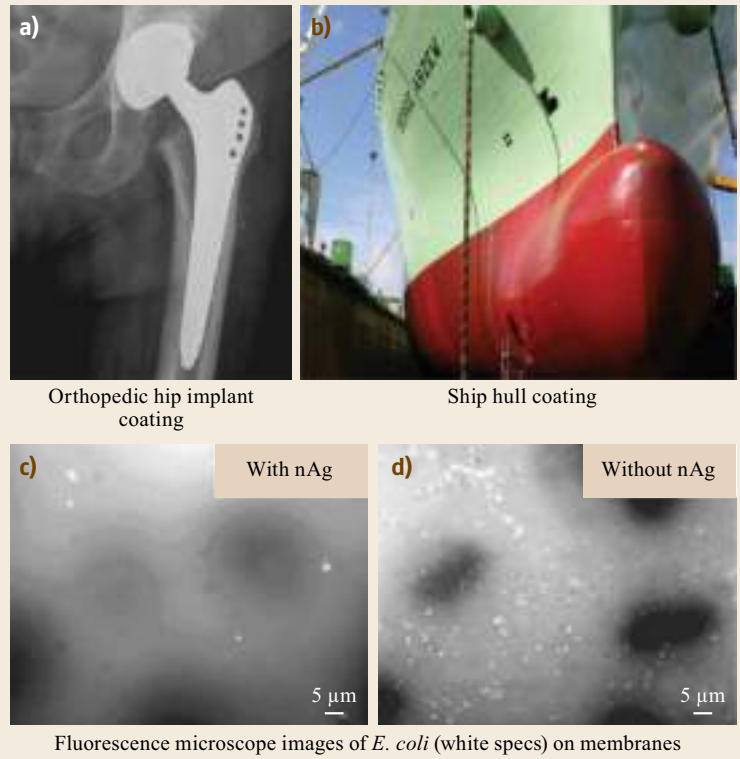


Fig. 37.10a–d Medical, marine, and industrial antifouling coating examples [37.16]. Orthopedic hip implants are covered with biocompatible coatings designed to reduce biofouling, wear rate, and promote bone growth **(a)** (photo courtesy of Ir Jef Vleugels). Marine foul-release organic tin-free hull coating called *Green Ocean Coating Heavy Duty* minimizes manual cleaning **(b)** (photo courtesy of Bayer MaterialScience). Polysulfone ultrafiltration membrane treated with silver nanoparticles (nAg) shows antifouling benefits with fewer *E. coli* attached **(c,d)** (reprinted from [37.77] with permission from Elsevier)

ered with a biocompatible coating designed to control biofouling, wear rate, and promote bone growth [37.78, 79].

Additional medical prevention techniques include hydrophilic red blood cell plasma membranes that reduce biofouling in medical devices and other synthetic membranes [37.80]. For instance, Advanced Hydro hydrophilic polymer coatings on water purification and filtration membranes provide a 30–50% reduction in fouling (www.advancedhydro.net). Clean Membranes utilize a *smart-comb* copolymer in ultrafiltration membranes to provide antifouling and increased permeability (www.cleanmembranes.com). Superhydrophilic and antimicrobial nanoparticles in a polyamide thin film coating provide antifouling for reverse osmosis (RO) membranes (www1.cnsi.ucla.edu). Furthermore, studies show success using titanium dioxide (TiO_2) nanoparticles as a thin film on reverse osmosis membranes to control biofouling due to the photocatalytic effects that reduce bacteria [37.81]. Similarly, ultrafiltration membranes impregnated with antimicrobial and hydrophilic silver nanoparticles (nAg) provide antifouling benefits. Figure 37.10c,d from *Bixler* and *Bhushan* [37.16] compares such treated and untreated polysulfone ultrafiltration membranes, showing fewer *E. coli* (*Escherichia coli*) attached to the nAg membrane [37.77].

Tributyltin coatings have been efficacious in marine antifouling applications, and were commonplace. Unfortunately, toxicity and other properties make tributyltin coatings unsuitable for in vivo use. Because of tributyltin toxicity to marine organisms, antifouling preparations containing it were banned in 2008 by the International Maritime Organization, adding urgency to the development of more environmentally friendly antifouling and foul-release coatings. As a result, many researchers have been exploring environmentally friendly methods such as new nontoxic antifouling paints and foul-release coatings [37.8, 10, 11, 82]. Marine foul-release coatings are nontoxic hydrophobic surfaces that promote self-cleaning. Nano/microtextured surfaces control surface energy, charge, conductivity, porosity, roughness, wettability, and friction [37.83]. Hydrophobic materials such as fluoropolymers and silicone elastomers are nontoxic, adhesion resistant, environmentally friendly, and easy to clean [37.84]. Siloxane elastomers provide a combination of lower elastic modulus and lower surface energy, creating a hydrophobic, low-stick surface. Silicone foul-release coatings are used for some

quick-moving boats, but wear away over time [37.54, 87]. Figure 37.10b from *Bixler* and *Bhushan* [37.16] shows a foul-release organic tin-free hull coating called *Green Ocean Coating Heavy Duty* by Advanced Marine Coatings, whose smoothness and hardness reduces cleaning cycles (www.bayermaterialscience.com).

Industrial antifouling is accomplished similarly to marine antifouling, with coatings that prevent and remove fouling [37.16]. For instance, industrial heat-

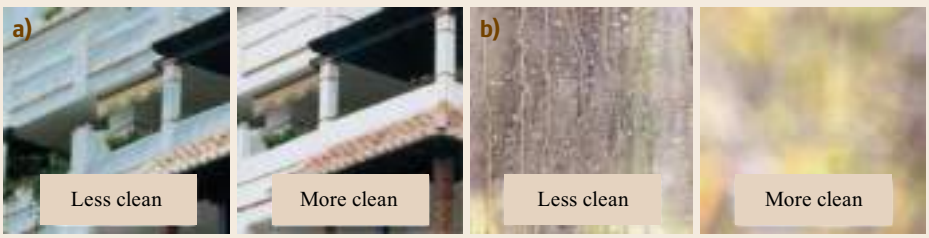


Fig. 37.11a–d Effectiveness of anti-inorganic fouling coatings [37.16]. Demonstrated is the improved cleanliness with Lotusan hydrophobic self-cleaning paint **(a)** (www.stocorp.com), and improved clarity with the SunClean hydrophilic self-cleaning glass **(b)** (www.ppg.com)

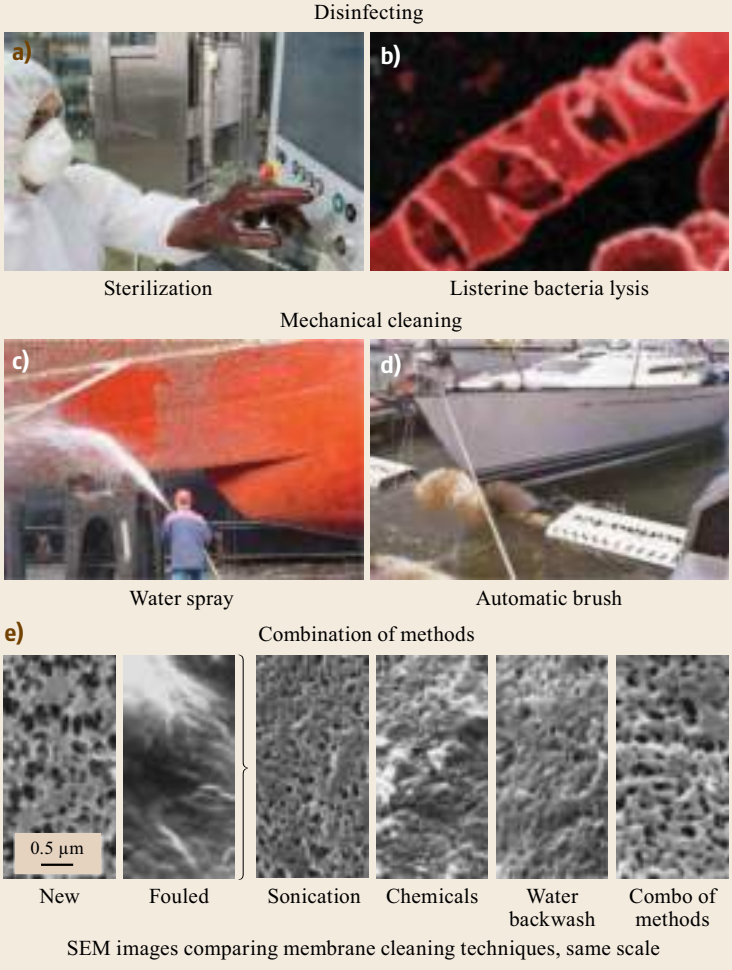


Fig. 37.12a–e Medical, marine, and industrial biofouling cleaning examples [37.16]. Steam sterilization medical autoclave **(a)** (photo courtesy of Life Science Outsourcing), Listerine effect on bacteria **(b)** (adapted from Christoph Schaudinn), water spray on a low-stick ship hull **(c)**, automatic brush washing system **(d)** [37.85], and microfiltration membrane using sonication, chemicals, and water backwash **(e)** (reprinted from [37.86] with permission from Elsevier)

exchanger applications utilize SAMs to reduce bioadhesive strength between foulers and a heated surface, using low surface energy hydrophobic properties. Materials such as silicon can withstand temperatures up to 200 °C whilst materials such as hexadecyl disulfide can withstand up to 225 °C. SAMs furthermore offer corrosion protection by limiting the oxygen and water diffusion at the substrate surface. This technique could be widely used for heat exchangers but needs long-term durability testing [37.88].

37.4.2 Self-Cleaning Surfaces and Cleaning Techniques

Even though fouling prevention is the goal, in many cases fouling is not completely preventable and a surface must be cleaned. In response, several self-cleaning coatings have been developed for various applications. *Bixler* and *Bhushan* [37.16] report example self-cleaning products including hydrophobic paints, roof tiles, and fabrics as well as self-cleaning hydrophilic windows and membranes. Examples are Lotusan paints by Sto (www.stocorp.com), Tegotop paints by Degussa (www.goldschmidt.com), Erlus Lotus roof tiles (www.erlus.com), NanoSphere fabrics by Schoeller Technologies (www.schoeller-textiles.com), and SunClean glass by PPG Industries (www.ppg.com). Self-cleaning is demonstrated with the hydrophobic Lotusan roll on and Tegotop spray paints that create a microstructure surface (www.stocorp.com; www.goldschmidt.com), hydrophobic Erlus Lotus roof tiles with an etched surface finish (www.erlus.com), and hydrophobic NanoSphere fabrics with integrated nanoparticles (www.schoeller-textiles.com). Self-cleaning hydrophilic SunClean glass with titanium dioxide uses the sheeting action of water to efficiently gather and remove contaminants (www.ppg.com). Figure 37.11 shows the improved cleanliness with Lotusan hydrophobic self-cleaning paints on concrete and stucco (www.stoco2rp.com), and the improved clarity with SunClean hydrophilic self-cleaning glass (www.ppg.com) [37.16].

Physical cleaning and chemical disinfecting are often considered effective methods to remove fouling [37.7, 9, 11, 16, 89, 90]. Most employ mechanical methods such as a manual scrubbing to disrupt and remove the fouling. Figure 37.12 illustrates common methods of cleaning and disinfecting such as medical sterilization, chemical disinfection, high-pressure water spray, automatic brush scrubbing, and membrane cleaning techniques [37.16]. Included are a steam sterilization medical autoclave, Listerine antiseptic mouth rinse, water spray on a low-adhesion ship hull, automatic brush washing system, and microfiltration membrane using sonication, chemicals, and water backwash [37.16, 86]. Disinfecting is commonly employed in dental applications with antiseptic mouth rinses such as Listerine. For example, the active ingredient chlorhexidine attacks and reduces dental biofilm to control plaque and gingivitis in the oral cavity [37.91]. The biofilm is reduced when the individual bacteria walls rupture (www.jjdentalprofessional.com). Mouth rinses combined with brushing and flossing increase the effectiveness, since such *mechanical methods* disrupt the biofilm [37.90].

37.5 Bioinspired Rice Leaf Surfaces for Antifouling

In this section, we present data on an antibiofouling study carried out by *Bixler* et al. [37.17] on rice leaf-inspired surfaces. The rice leaf is particularly intriguing due to its relatively simple micropapillae morphology. Bioinspired rice leaf surfaces using microsized pillar patterns can be produced with silicon master patterns fabricated by photolithography [37.92]. Sample geometrical dimensions are based on actual rice leaf morphology. Actual rice leaf surfaces are covered by a sinusoidal pattern of micropapillae with height 2–4 μm, diameter 2–4 μm, pitch 5–10 μm, and peak radius 0.5–1 μm [37.63, 69].

37.5.1 Fabrication of Micropatterned Samples

The bioinspired rice leaf surface shown in Fig. 37.13 was created to replicate the rice and butterfly wing effect, and was designed to include anisotropic flow and superhydrophobic characteristics. Samples were created using silicon master patterns and the soft-lithography technique [37.92]. Silicon master patterns with pits (negative versus positive features) were fabricated in order to emboss polymer films and to accurately produce master patterns. Polydimethylsiloxane (PDMS) replicas were produced from silicon master patterns using a three-step soft-lithography procedure [37.92]. PDMS (184 Silicone Elastomer by Dow Corning) was chosen due to its low surface energy property which leads to high contact angle. A nanostructured superhydrophobic coating with 50 nm hydrophobized silica nanoparticles and methylphenyl silicone resin was applied to make them superhydrophobic.

Micropatterned plastic sheets were also produced using a commercial hot embossing procedure in a clean room environment [37.92]. The hot embosser pro-

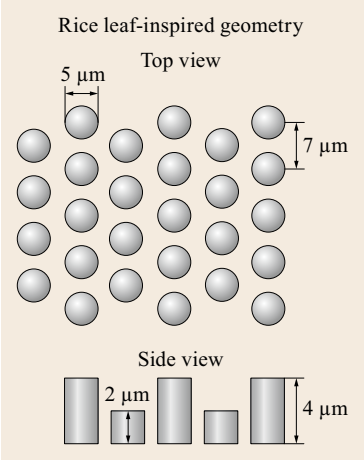


Fig. 37.13 Rice leaf-inspired geometry for antifouling. Shown are micropatterned surfaces with dual height pillars. Pillars are arranged in a hexagonal array in order to maintain consistent gaps to promote antifouling (adapted from [37.92])

duced single-sided embossed plastic sheets from a silicon master pattern using a combination of heat and pressure. Poly(methyl methacrylate) (PMMA) plastic sheets were selected, which were optically clear and 175 μm in thickness. PMMA molds relatively easily due to its softening temperature.

37.5.2 Antibiofouling Measurements

For studies to evaluate antibiofouling benefits, *Bixler* et al. [37.17] considered several foulers primarily based on their size and availability. These include zoospores (*Enteromorpha*) [37.55] and (*Ulva linza*) [37.93], proteins [37.94], *E. coli* [37.95], and MRSA (*Methicillin-resistant Staphylococcus aureus*) [37.96]. They selected the microorganism *E. coli* due to its prevalence in the medical industry, availability, and access to facilities for safe handling. The multistep antibiofouling measurement process is shown in Fig. 37.14 with each process from *E. coli* broth creation to the final counting [37.17]. The dimensions of single wild-type *E. coli* cells were approximately 0.5–0.75 μm wide and 2 μm long. When actively growing and dividing, the cells stretch to ≈ 4 μm before dividing into two cells.

The bioassay procedure began with a stock solution of *E. coli* (HB2151 *E. coli* clones from The Medical Research Council, Cambridge, England) streaked onto TYE (Tryptone and Yeast Extract) dehydrated agar plates for bacterial colony growth. The plates were then placed in an incubator (Labnet, model 311 DS) at 37 °C overnight, and then stored at 4 °C until use. To prepare the *E. coli* broth, a single colony was picked from streak plates and inoculated into 50 ml sterile LB (Luria Broth, Lennox from Fisher Scientific, Pittsburgh, PA) in a beveled flask. *E. coli* was grown for 16 hours with shaking at 250 rpm at a temperature of 37 °C. After incubation, the *E. coli* solution was diluted to 1/100

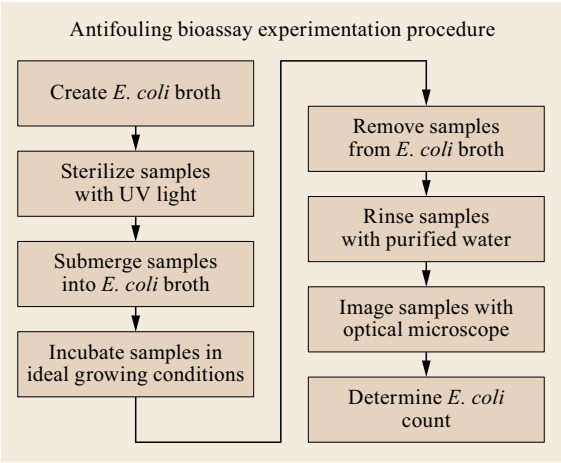


Fig. 37.14 Antifouling bioassay experimental procedure with *E. coli* cells in the laboratory. As shown, samples are sterilized, submerged into broth, placed in an incubator, rinsed to remove loose cells, and imaged with an oil immersion light microscope to determine cell counts (after [37.17])

and 1/10 into sterile LB broth and placed on ice until use.

Bixler et al. [37.17] placed the flat control sample (PDMS), a replica sample (PDMS), and an embossed sample (PMMA) in the bottom of a sterile, polystyrene culture dish and sterilized for 10 min via ultraviolet exposure (30W T8 Hot Cathode Germicidal Sterilamp fluorescent light bulb, Philips TUV30W G30T8 360164). *E. coli* broth was added to the culture dish until all samples were covered and then incubated at 37 °C for either 2 or 4 hours. The aforementioned concentrations and incubation times were determined from trial and error with samples to achieve an appropriate level of *E. coli* growth. The goal was to visually determine a difference at 1000× magnification between flat control and other samples under investigation.

After incubation, samples were removed from the *E. coli* broth and rinsed at a 45° angle. Rinsing removed any unattached *E. coli* from the surface prior to imaging. An autopipette (Pipet-Aid from Drummond Scientific, Broomall, PA) was clamped and held on a stand over the rinsing station with a liquid fall height (distance between pipette tip end and sample) of 7 cm. Samples were rinsed with 10 ml of sterile PBS at a flow rate of approximately 2 ml/s, and then immediately placed onto a glass microscope slide with a glass cover slip on top. Using a light microscope and CCD (charge coupled device) camera (Nikon, Diagnostics Instruments, Inc.), a 140 μm by 140 μm area of each sample was imaged and flat samples analyzed with image processing software (SPIP 5.1.11, Image Metrol-

ogy A/S, Horshølm, Denmark). Images were collected with a 100× oil immersion objective lens for a total magnification of 1000×. For flat control samples, image processing software (SPIP 5.1.11, Image Metrology A/S, Horshølm, Denmark) was utilized to determine total coverage area of *E. coli*. For the samples, the *E.coli* was manually counted five times in circular areas 35 μm in diameter using two samples within the same bioassay.

37.5.3 Anti-inorganic Fouling Measurements

To understand the anti-inorganic fouling effectiveness of the samples, self-cleaning experiments were conducted [37.92]. These experiments included contaminating samples, employing a wash technique, and determining the percentage of inorganic particles removed. Depositing contaminate particles on the tilted (45°) replicas involved a glass contamination chamber (0.3 m in diameter and 0.6 m high). A tray containing 0.2 g of hydrophilic silicon carbide (SiC) contaminants (400 mesh particle size by Aldrich, with sizes ranging from 10–15 μm) was placed in the top chamber with an air hose directed at the center. These particles were chosen because of their similar properties to natural dirt (shape, size, and hydrophilicity). Contaminants were blown with laboratory air for 10 s at 300 kPa, and then allowed to settle for 30 s before the separator panel was removed.

After 30 min, the samples were removed and subjected to prewash experiment particle analysis. Using a light microscope and CCD camera, a 280 μm by 210 μm area of each sample was imaged and analyzed with SPIP image processing software. This process was performed before and after each wash experiment. Wash experiments consisted of exposing the tilted (45°) replicas to water droplets falling from specified height and drip rates (total duration of 2 min using 10 mL water). Droplet velocity was approximately 5.6 m/s at a height of 0.4 m. This translates into pressures of 4000 Pa.

37.5.4 Results and Discussion

Antibiofouling

To understand the antifouling effectiveness of the various samples, *Bixler* et al. [37.17] conducted a series of bioassay experiments. Figure 37.15 shows light microscope images of the *E. coli* colonized on a flat sample at two magnifications [37.17]. As shown, the *E. coli* are long cylindrically shaped microorganisms and tend to cluster in groups and align themselves from head to tail.

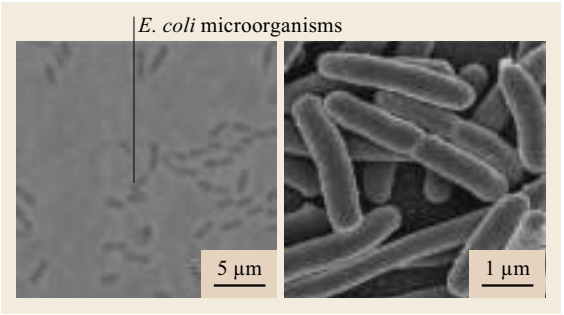


Fig. 37.15 Images of *E. coli* on flat surfaces after the bioassay procedure [37.17] and at a higher magnification (adapted from www.universityofcalifornia.edu)

Figure 37.16a shows light microscope images of flat PDMS samples at the two concentrations and two incubation times (with *E. coli* highlighted by the SPIP software) [37.17]. Figure 37.16b shows the tabulated results of percentage coverage area of the *E. coli* on the surfaces for 1/100 and 1/10 concentrations at 2 and 4 h incubation time. The greatest coverage occurs at the higher incubation time as *E. coli* growth occurs exponentially over time. Furthermore, as expected, the lower concentration does not provide as much coverage area, however the differences are within 10% for both incubation times.

Figure 37.17 shows light microscopic images of the flat PDMS control sample, replica sample, and the embossed sample. Antifouling is determined by counting the total number of *E. coli* on the flat control samples and comparing that to the other samples, with results reported in percentage *E. coli* reduced coverage. Figure 37.18 shows the percentage reduced coverage compared to the flat control samples. As anticipated due to the discontinuities of the various surfaces, patterned samples provided antifouling benefits. In general, the antifouling benefit is greater with the 2 h incubation duration as compared to the 4 h duration. It is believed that the longer duration leads to more significant coverage, and the features provide less overall benefit. It should be noted that the replica sample and embossed sample performed similarly, indicating minimal differences between samples fabricated from PDMS or PMMA.

Anti-inorganic Fouling

To understand the anti-inorganic fouling effectiveness of the various samples, *Bixler* and *Bhushan* [37.92] conducted a series of self-cleaning wash experiments. Figure 37.19 presents the before and after images for flat PDMS and PMMA control samples, the replica sample, and the embossed sample [37.92]. These images were used with SPIP imaging software to quantify the percentage of particles removed. Figure 37.20

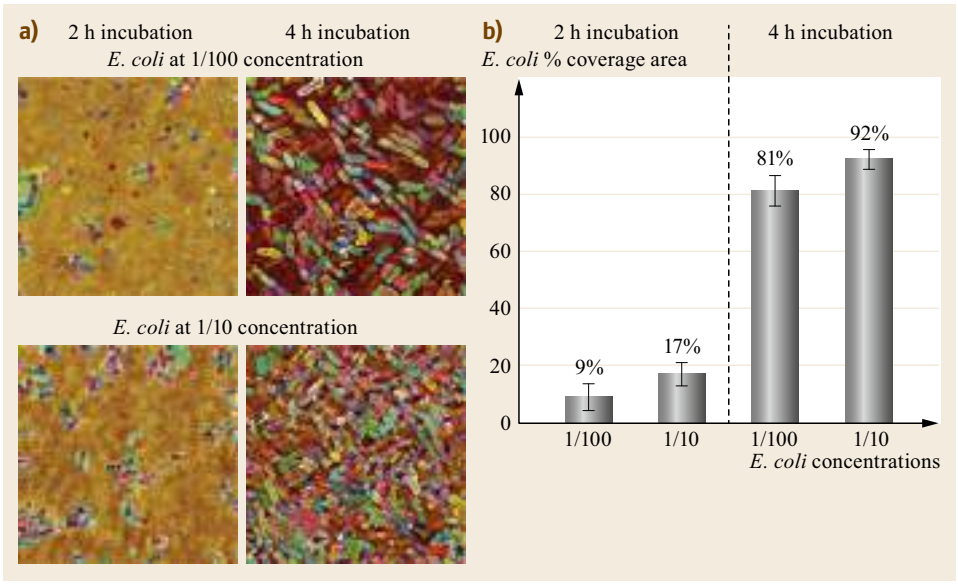


Fig. 37.16 **(a)** Light microscope images of flat PDMS control samples and **(b)** percent coverage area after *E. coli* bioassay with two concentrations and two incubation times. *E. coli* covers the majority of the surface after 4 hours in ideal growing conditions. This baseline serves to compare the replica sample and embossed sample when calculating percentage reduced coverage (adapted from [37.17])

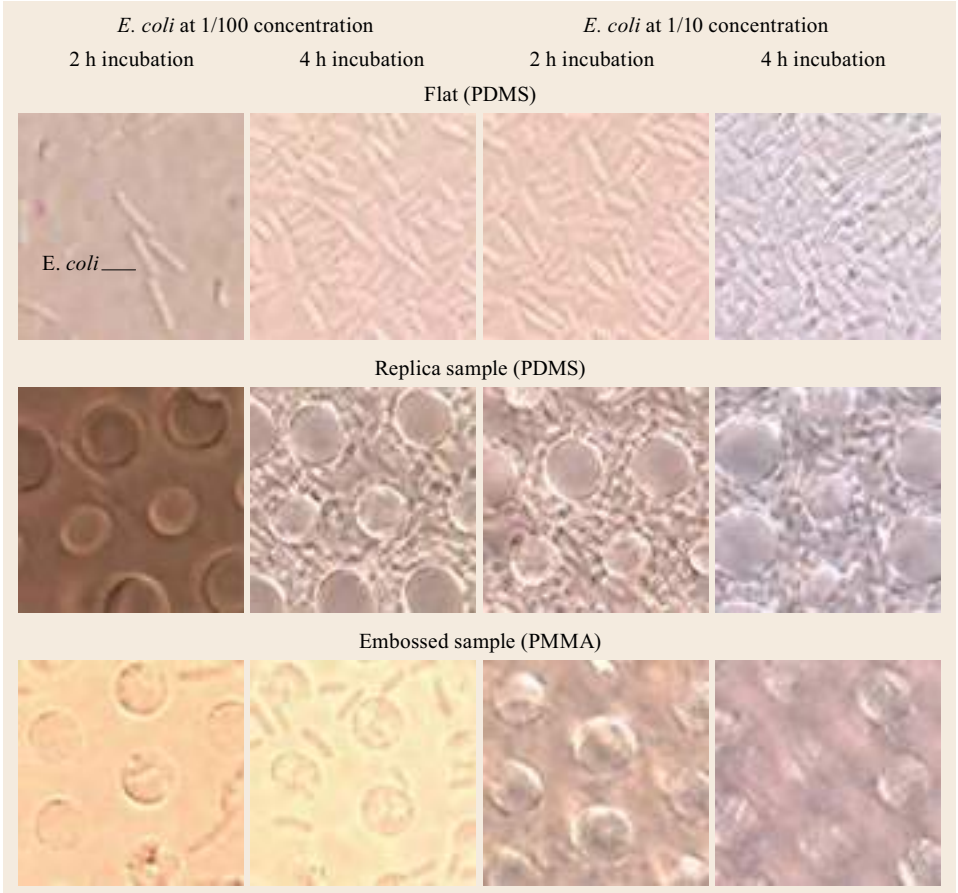


Fig. 37.17 Light microscope images using oil immersion techniques at 1000× magnification of samples after bioassay experiments using *E. coli*. Images of flat PDMS, the replica sample, and the embossed sample are shown with two *E. coli* concentrations and two incubation times. The long and narrow cylindrically shaped objects are the *E. coli*. Antifouling analysis was conducted using the images and particle analysis software to quantify effectiveness (adapted from [37.17])

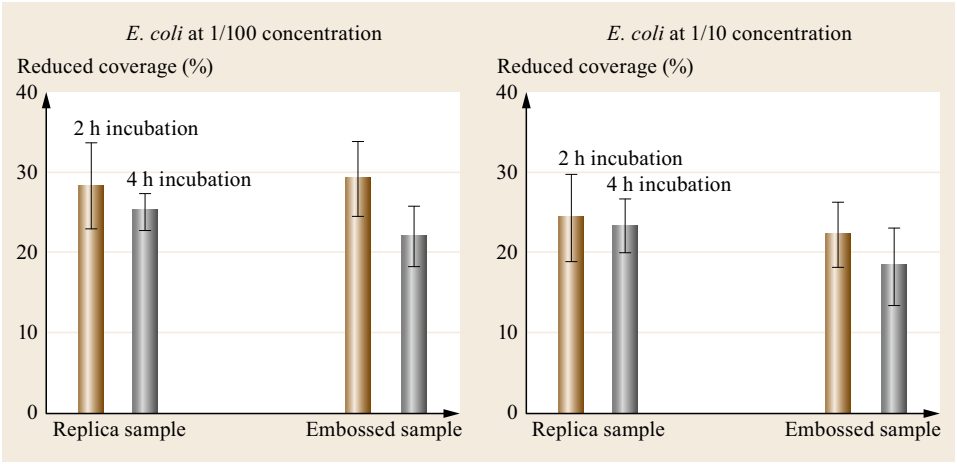


Fig. 37.18 Antibiofouling results from *E. coli* bioassay experiments. Results are presented for flat PDMS, the replica sample, and the embossed sample. The percentage reduced coverage is calculated from the flat PDMS control samples. It is believed that antifouling efficiency improves with high CA and low CAH. Error bars show ±1 standard deviation (adapted from [37.17])

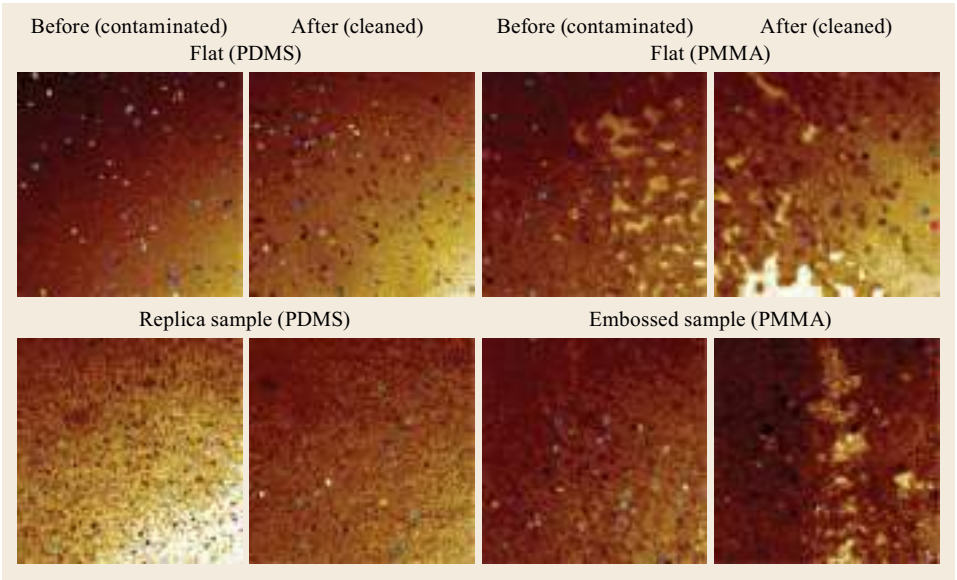


Fig. 37.19 Light microscope images of samples before and after anti-inorganic fouling wash experimentation, with specs representing silicon carbide (SiC) particles [37.92]. Results are presented for flat PDMS and PMMA, the replica sample, and the embossed sample. Self-cleaning analysis was conducted using the images and particle analysis software to quantify particle removal

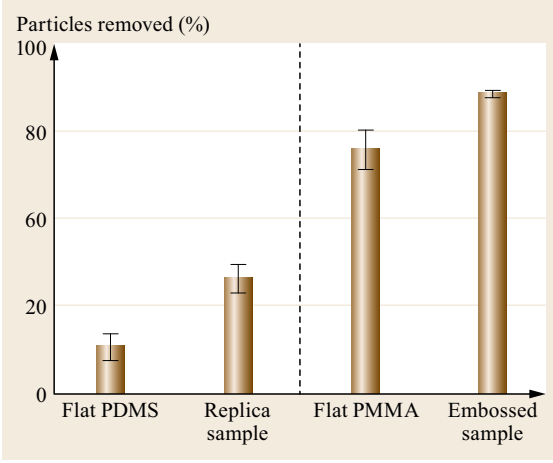


Fig. 37.20 Anti-inorganic fouling wash experimentation results (adapted from [37.92]). Results are presented for the flat PDMS, the replica sample, the flat PMMA sample, and the embossed sample. The embossed sample shows the greatest self-cleaning efficiency followed by the replica sample. Micropatterned samples are superior in self-cleaning efficiency compared to their respective flat samples. Error bars show ±1 standard deviation ◄

shows self-cleaning data for wash experiments. Self-cleaning is more evident with a patterned sample and an embossed sample, as compared to flat samples [37.92]. To understand wettability of samples, *Bixler* and *Bhushan* [37.92] made contact angle measurements. They measured contact angle (CA) and contact angle hysteresis (CAH) values. The replica sample

was found to be superhydrophobic ($\approx 150°$) with low CAH, whereas the embossed sample was hydrophobic ($\approx 101°$). High CA and low CAH is believed to amplify the self-cleaning abilities of the samples, as it is believed that the droplets are able to roll and collect the particles after impact.

37.5.5 Antibiofouling and Anti-inorganic Fouling Mechanisms

Bixler et al. [37.17] presented conceptual models to describe antibiofouling mechanisms. Figure. 37.21 shows schematic illustrations of colonization of microorganisms exposed to flat and patterned samples. Microorganisms are unable to effectively colonize the patterned morphology as well as the flat control sample. If the length of the microorganism is larger than the pillar spacing, they cannot invade the spacing and instead land on top of the pillars. If the length of the microorganism is smaller than the spacing, as is the case with *E. coli*, the sharp edges of the pillars are believed to discourage the settling and eventual colonization of microorganisms on the surface. The edges provide an inconvenient landing zone for settlement, and thus the microorganism must continue the search for an ideal location to colonize. It is believed during this process that many *E. coli* that do not attach remain metabolically active throughout the experiment while others may expire. In many cases, the *E. coli* may be able to settle in clusters and around the bases of the pillars, which provide multiple attachment points of contact.

When the microorganism concentrations are low and the incubation times are short, only a few microorganisms are able to colonize the sample surface. However, when the concentrations are high and the incubation times are long, the microorganisms eventually will cover over the pillars and create a biofilm. Nevertheless, the discontinuities found with the sharp-edged microsized pillars are demonstrated to provide antifouling benefit and extend the time to which a particular surface becomes colonized.

An illustration of anti-inorganic fouling mechanisms is shown in Fig. 37.22 [37.92]. The schematic shows that foulers may be removed from a surface by water droplets or fluid flow when the rice and butterfly wing effect is present. Such a pattern is believed to encourage self-cleaning and directional fluid control, which is believed to provide antifouling benefit.

To summarize, there are a number of examples of micropatterned surfaces in the literature that show that they exhibit antibiofouling if the microsized feature spacing is slightly smaller than the microorganisms. The data presented in this section shows that even if the spacing is larger, they exhibit antibiofouling effects.

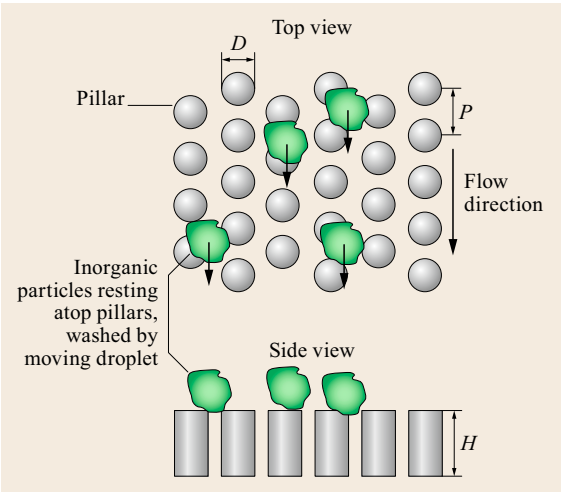


Fig. 37.22 Anti-inorganic fouling mechanisms with rice leaf-inspired samples (adapted from [37.92]. Model shown utilizes uniformly spaced microsized single-height pillars in a hexagonal array. Inorganic fouling may be washed away in fluid flow via self-cleaning

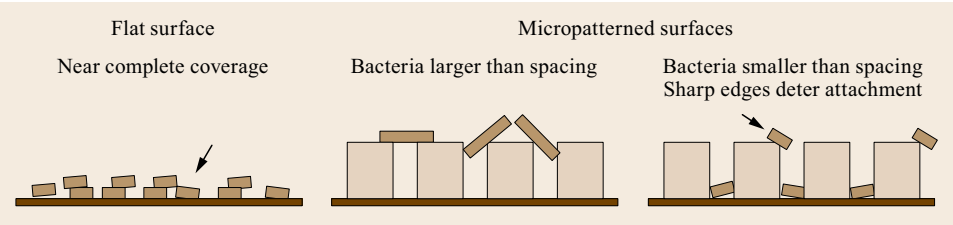


Fig. 37.21 Conceptual models of antibiofouling mechanisms with micropatterned samples. Shown are colonization on flat and micropatterned surfaces. If the length of the microorganism is larger than the pillar spacing, it cannot invade the spacing and instead it lands on top of the pillars. If the length of the microorganism is smaller than the spacing, microstructured discontinuities and sharp edges are believed to discourage colonization – where they may expire in the process of searching for an ideal landing site

37.6 Closure

In this chapter, biofouling and inorganic fouling were introduced. Next, fields susceptible to fouling were summarized, followed by discussion of biofouling and inorganic fouling formation mechanisms, antifouling strategies from living nature, and current prevention and cleaning techniques. Finally, demonstrations of the use of bioinspired rice leaf surfaces for anti-fouling were presented. Data showed that micropatterned surfaces facilitate antifouling.

References

37.1 B. Bhushan: Biomimetics: Bioinspired Hierarchical-Structured Surfaces for Green Science and Technology, 2nd edn. (Springer International, Cham 2016)

37.2 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Cham 2016)

37.3 P. Stoodley, K. Sauer, D.G. Davies, J.W. Costerton: Biofilms as complex differentiated communities, Annu. Rev. Microbiol. **56**, 187–209 (2002)

37.4 T. Vo-Dinh (Ed.): *Nanotechnology in Biology and Medicine* (CRC, Boca Raton 2007)

37.5 M.J. Schulz, V.N. Shanov, Y. Yun (Eds.): *Nanomedicine Design of Particles, Sensors, Motors, Implants, Robots, and Devices* (Artech House, Boston 2009)

37.6 M. Shirtliff, J.G. Leid (Eds.): *The Role of Biofilms in Device-Related Infections* (Springer, Berlin, Heidelberg 2009)

37.7 J. Chan, S. Wong: *Biofouling Types, Impact and Anti-Fouling* (Nova Science, New York 2010)

37.8 M. Fingerman, R. Nagabhushanam, M.F. Thompson (Eds.): *Recent Advances in Marine Biotechnology* (Science, Enfield 1999)

37.9 J. Walker, S. Surman, J. Jass (Eds.): *Industrial Biofouling Detection, Prevention and Control* (Wiley, New York 2000)

37.10 A.I. Railkin: *Marine Biofouling Colonization Processes and Defenses* (CRC, Boca Raton 2004)

37.11 C. Hellio, D. Yebra (Eds.): *Advances in Marine Antifouling Coatings and Technologies* (CRC, Boca Raton 2009)

37.12 E.F.C. Somerscales, J.G. Knudsen (Eds.): *Fouling of Heat Transfer Equipment* (Hemisphere, Washington 1981)

37.13 T.R. Bott: Crystallization fouling – Basic science and models. In: *Fouling Science and Technology*, ed. by L.F. Melo, T.R. Bott, C.A. Bernardo (Kluwer Academic, Dordrecht 1988) pp. 251–260

37.14 A.M. Pritchard: Deposition of hardness salts. In: *Fouling Science and Technology*, ed. by L.F. Melo, T.R. Bott, C.A. Bernardo (Kluwer Academic, Dordrecht 1988) pp. 261–274

37.15 L.F. Melo, T.R. Bott, C.A. Bernardo (Eds.): *Fouling Science and Technology* (Kluwer Academic, Dordrecht 1988)

37.16 G.D. Bixler, B. Bhushan: Biofouling lessons from nature, Phil. Trans. R. Soc. A **370**, 2381–2417 (2012)

37.17 G.D. Bixler, A. Theiss, B. Bhushan, S.C. Lee: Anti-fouling properties of microstructured surfaces bio-inspired by rice leaves and butterfly wings, J. Colloid Interf. Sci. **419**, 114–133 (2014)

37.18 A. Trinidad, A. Ibanez, D. Gomez, J.R. Garcia-Berrocal, R. Ramierz-Camacho: Application of environmental scanning electron microscopy for study of biofilms in medical devices, Micros. Sci. Technol. Appl. Educ. **1**, 204–210 (2010)

37.19 T.J. Marrie, J.W. Costerton: Morphology of bacterial attachment to cardiac pacemaker leads and power packs, J. Clin. Microbiol. **19**, 911–914 (1984)

37.20 M. Copisarow: Marine fouling and its prevention, Science **101**, 406–407 (1945)

37.21 Anonymous: *Marine Fouling and its Prevention* (Woods Hole Oceanographic Institute, US Naval Institute, Annapolis 1952)

37.22 D.L. Ray (Ed.): *Marine Boring and Fouling Organisms* (Univ. of Washington Press, Seattle 1959)

37.23 A.B. Cunningham, J.E. Lennox, R.J. Ross: Biofilms in Industrial Environments, http://biofilmbook.hypertextbookshop.com/public_version/contents/contents.html (2008) retrieved June 27, 2011

37.24 M. Simoes, L.C. Simoes, M.J. Vieira: A review of current and emergent biofilm control strategies, LWT – Food Sci. Technol. **43**, 573–583 (2010)

37.25 M. Herzberg, M. Elimelech: Biofouling of reverse osmosis membranes: Role of biofilm-enhanced osmotic pressure, J. Membr. Sci. **295**, 11–20 (2007)

37.26 Y. Wang, B. Bhushan: Wear-resistant and anti-smudge superoleophobic coating on polyethylene terephthalate substrate using SiO_2 nanoparticles, ACS Appl. Mater. Interf. **7**, 743–755 (2015)

37.27 L. Cao, A.K. Jones, V.K. Sikka, J. Wu, D. Gao: Anti-icing superhydrophobic coatings, Langmuir **25**, 12444–12448 (2009)

37.28 A.J. Meuler, J.D. Smith, K.K. Varanasi, J.M. Mabry, G.H. McKinley, R.E. Cohen: Relationships between water wettability and ice adhesion, ACS Appl. Mat. Interf. **2**, 3100–3110 (2010)

37.29 C.W. Keevil, A. Godfree, D. Holt, C. Dow (Eds.): *Biofilms in the Aquatic Environment* (The Royal Society of Chemistry, Cambridge 1999)

37.30 H.C. Flemming, U. Szewzyk, T. Griebe (Eds.): *Biofilms Investigative Methods and Applications* (Technomic, Lancaster 2000)

37.31 J.W. Costerton: *Springer Series on Biofilms* (Springer, Berlin, Heidelberg 2008)

37.32 E. Ralston, G. Swain: Bioinspiration – The solution for biofouling control?, Bioinsp. Biomim. **4**, 1–9 (2009)

37.33 D. Monroe: Looking for chinks in the armor of bacterial biofilms, PLoS Biol. **5**, 2458–2461 (2007)

37.34 T. Griebe, H.C. Flemming: Rotating annular reactors for controlled growth of biofilms. In: *Biofilms Investigative Methods and Applications*, ed. by H.C. Flemming, U. Szewzyk, T. Griebe (Technomic, Lancaster 2000) pp. 1–22

37.35 G. Jones: The battle against marine biofouling: A historical review. In: *Advances in Marine Antifouling Coatings and Technologies*, ed. by C. Hellio, D. Yebra (CRC, Boca Raton 2009) pp. 19–45

37.36 M.L.B. Palacio, B. Bhushan: Bioadhesion: A review of concepts and applications, Phil. Trans. Royal Soc. A **370**, 2321–2347 (2012)

37.37 M.E. Callow, R.A. Pitchers, A. Milne: The control of fouling by non-biocidal systems. In: *Algal Biofouling*, ed. by L.V. Evans, K.D. Hoagland (Elsevier Science, Amsterdam 1986) pp. 145–158

37.38 X. Feng, L. Jiang: Design and creation of superwetting/antiwetting surfaces, Adv. Mater. **18**, 3063–3078 (2006)

37.39 K. Koch, W. Barthlott: Superhydrophobic and superhydrophilic plant surfaces: An inspiration for biomimetic materials, Phil. Trans. R. Soc. **367**, 1487–1509 (2009)

37.40 A.J. Scardino: Surface modification approaches to control marine biofouling. In: *Advances in Marine Antifouling Coatings and Technologies*, ed. by C. Hellio, D. Yebra (CRC, Boca Raton 2009) pp. 664–692

37.41 X. Sheng, Y.P. Ting, S.O. Pehkonen: Force measurements of bacterial adhesion to metals using cell probe in atomic force microscopy. In: *Biofouling Types, Impact and Anti-Fouling*, ed. by J. Chan, S. Wong (Nova Science, New York 2010) pp. 129–154

37.42 H.C. Flemming, J. Wingender, R. Moritz, W. Borchard, C. Mayer: Physio-chemical properties of biofilms – A short review. In: *Biofilms in the Aquatic Environment*, ed. by C.W. Keevil, A. Godfree, D. Holt, C. Dow (The Royal Society of Chemistry, Cambridge 1999) pp. 1–12

37.43 N. Epstein: Particulate fouling of heat transfer surfaces: Mechanisms and models. In: *Fouling Science and Technology*, ed. by L.F. Melo, T.R. Bott, C.A. Bernardo (Kluwer Academic, Dordrecht 1988) pp. 143–164

37.44 T.R. Bott: Crystallization of organic materials. In: *Fouling Science and Technology*, ed. by L.F. Melo, T.R. Bott, C.A. Bernardo (Kluwer Academic, Dordrecht 1988) pp. 275–280

37.45 P. Vadgama (Ed.): *Surfaces and Interfaces for Biomaterials* (CRC, Boca Raton 2005)

37.46 D.P. Gordon, S.F. Mawatari: *Atlas of Marine-Fouling Bryozoa of New Zealand Ports and Harbours* (New Zealand Oceanographic Institute, Wellington 1992)

37.47 B. Bhushan, Y.C. Jung: Natural and biomimetic artificial surfaces for superhydrophobicity, self-cleaning, low adhesion, and drag reduction, Prog. Mater. Sci. **56**, 1–108 (2011)

37.48 T. Harder, L.H. Yee: Bacterial adhesion and marine fouling. In: *Advances in Marine Antifouling Coatings and Technologies*, ed. by C. Hellio, D. Yebra (CRC, Boca Raton 2009) pp. 113–131

37.49 M.L. Carman, T.G. Estes, A.W. Feinburg, J.F. Schumacher, W. Wilkerson, L.H. Wilson, M.E. Callow, J.A. Callow, A.B. Brennan: Engineered antifouling microtopographies – Correlating wettability with cell attachment, Biofouling **22**, 11–21 (2006)

37.50 J.F. Schumacher, N. Aldred, M.E. Callow, J.A. Finlay, J.A. Callow, A.S. Clare, A.B. Brennan: Species-specific engineered antifouling topographies: Correlations between the settlement of algal zoospores and barnacle cyprids, Biofouling **23**, 307–317 (2007)

37.51 J.F. Schumacher, M.L. Carman, T.G. Estes, A.W. Feinberg, L.H. Wilson, M.E. Callow, J.A. Callow, J.A. Finlay, A.B. Brennan: Engineered antifouling microtopographies – Effect of feature size, geometry, and roughness on settlement of zoospore of the green alga *Ulva*, Biofouling **23**(1), 55–62 (2007)

37.52 A.B. Brennan, R.H. Baney, M.I. Carman, T.G. Estes, A.W. Feinberg, L.H. Wilson, J.F. Schumacher: Surface Topographies for Non-Toxic Bioadhesion Control, United States Patent 7 650 848 (2010)

37.53 A.J. Scardino, J. Guenther, R. de Nys: Attachment point theory revisited: The fouling response to a microtextured matrix, Biofouling **24**, 45–53 (2008)

37.54 T. Vladkova: Surface modification approach to control biofouling. In: *Springer Series on Biofilms*, ed. by J.W. Costerton (Springer, Berlin, Heidelberg 2008) pp. 135–162

37.55 M.E. Callow, A.R. Jennings, A.B. Brennan, C.E. Seegert, A. Gibson, L. Wilson, A. Feinberg, R. Baney, J.A. Callow: Microtopographic cues for settlement of zoospores of the green fouling alga enteromorpha, Biofouling **18**, 237–245 (2002)

37.56 L. Hoipkemeier-Wilson, J.F. Schumacher, M.L. Carman, A.L. Gibson, A.W. Feinberg, M.E. Callow, J.A. Finlay, J.A. Callow, A.B. Brennan: Antifouling potential of lubricious, micro-engineered, PDMS elastomers against zoospores of the green fouling alga ulva *(Enteromorpha)*, Biofouling **20**, 53–63 (2004)

37.57 B. Dean, B. Bhushan: Shark-skin surfaces for fluid-drag reduction in turbulent flow: A review, Phil. Trans. R. Soc. A **368**, 4775–4806 (2010)

37.58 W. Lee, M.K. Jin, W.C. Yoo, J.K. Less: Nanostructuring of a polymeric substrate with well-defined nanometer-scale topography and tailored surface wettability, Langmuir **20**, 7665–7669 (2004)

37.59 A. Kesel, R. Liedert: Learning from nature: Non-toxic biofouling control by shark skin effect, Comp. Biochem. Physiol. A **146**, S130 (2007)

37.60 K. Koch, B. Bhushan, W. Barthlott: Diversity of structure, morphology, and wetting of plant surfaces, Soft Matter **4**, 1943–1963 (2008)

37.61 K. Koch, B. Bhushan, W. Barthlott: Multifunctional surface structures of plants: An inspiration for biomimetics, Prog. Mater. Sci. **54**, 137–178 (2009)

37.62 B. Bhushan: Biomimetics: Lessons from nature – An overview, Phil. Trans. R. Soc. A **367**, 1445–1486 (2009)

37.63 G.D. Bixler, B. Bhushan: Rice- and butterfly-wing effect inspired low drag and antifouling surfaces: A review, Crit. Rev. Solid State Mater. Sci. **40**, 1–37

(2015)
37.64 W. Barthlott, C. Neinhuis: Purity of the sacred lotus, or escape from contamination in biological surfaces, Planta **202**, 1–8 (1997)
37.65 C. Neinhuis, W. Barthlott: Characterization and Distribution of Water-repellent, Self-cleaning Plant Surfaces, Ann. Bot. **79**, 667–677 (1997)
37.66 B. Bhushan, Y.C. Jung, K. Koch: Micro-, nano- and hierarchical structures for superhydrophobicity, self-cleaning and low adhesion, Phil. Trans. Roy. Soc. A **367**, 1631–1672 (2009)
37.67 G.D. Bixler, B. Bhushan: Fluid drag reduction with shark-skin riblet inspired microstructured surfaces, Adv. Funct. Mater. **23**, 4507–4528 (2013)
37.68 T. Wagner, C. Neinhuis, W. Barthlott: Wettability and contaminability of insect wings as a function of their surface sculptures, Acta Zoologica **77**, 213–225 (1996)
37.69 G.D. Bixler, B. Bhushan: Fluid drag reduction and efficient self-cleaning with rice leaf and butterfly wing bioinspired surfaces, Nanoscale **5**, 7685–7710 (2013)
37.70 E. Bormashenko, Y. Bormashenko, T. Stein, G. Whyman, E. Bormashenko: Why do pigeon feathers repel water? Hydrophobicity of pennae, cassie-baxter wetting hypothesis, and cassie-wenzel capillary-induced wetting transition, J. Colloid Interf. Sci. **311**, 212–216 (2007)
37.71 D. Menton: The Seeing Eye, *AnswersMagazine.com*, July–Sept. 76–79 (2008)
37.72 J. Yen: Marine dynamics. In: *Bulletproof Feathers How Science Uses Nature's Secrets to Design Cutting-Edge Technology*, ed. by R. Allen (Univ. Chicago Press, Chicago 2010) pp. 22–43
37.73 K. LoVetri, P.V. Gawande, N. Yakandawala, S. Madhyastha: Biofouling and anti-fouling of medical devices. In: *Biofouling Types, Impact and Anti-Fouling*, ed. by J. Chan, S. Wong (Nova Science, New York 2010) pp. 105–128
37.74 N. Dror, M. Mandel, Z. Hazan, G. Lavie: Advances in microbial biofilm prevention on indwelling medical devices with emphasis on usage of acoustic energy, Sensors **9**, 2538–2554 (2009)
37.75 S. Sharma, K.C. Popat, T.A. Desai: Design and biological applications of nanostructured poly(ethylene glycol) films. In: *Nanotechnology in Biology and Medicine Methods, Devices, and Applications*, ed. by T. Vo-Dinh (CRC, Boca Raton 2007) p. 39-1
37.76 B. Bhushan, D. Hansford, K.K. Lee: Surface modification of silicon and polydimethylsiloxane surfaces with vapor-phase-deposited ultrathin fluorosilane films for biomedical nanodevices, J. Vac. Sci. Technol. A **24**, 1197–1202 (2006)
37.77 K. Zodrow, L. Brunet, S. Mahendra, D. Li, A. Zhang: Polysulfone ultrafiltration membranes impregnated with silver nanoparticles show improved biofouling resistance and virus removal, Water Res. **43**, 715–723 (2009)
37.78 Anonymous: *Why Is Titanium the Metal of Choice for Medical Applications from Head to Toe?* (International Titanium Association, Boulder 1999)
37.79 I.R. Vleugels: Multi-functional coating solution, Euprofile, Public Serv. Rev.: EU **18**, 1–2 (2006)
37.80 Z. Cui, Y. Wan: Biofouling in membrane separation systems. In: *Surfaces and Interfaces for Biomaterials*, ed. by P. Vadgama (CRC, Boca Raton 2005) pp. 493–544
37.81 S.H. Kim, S.Y. Kwak, B.H. Sohn, T.H. Park: Design of TiO_2 nanoparticle self-assembled aromatic polyamide thin-film-composite (TFC) membrane as an approach to solve biofouling problem, J. Membr. Sci. **211**, 157–165 (2003)
37.82 M.E. Callow: The status and future of biocides in marine biofouling prevention. In: *Recent Advances in Marine Biotechnology*, ed. by M. Fingerman, R. Nagabhushanam, M.F. Thompson (Science, Enfield 1999) pp. 109–126
37.83 A.S. Clare, N. Aldred: Surface colonization by marine organisms and its impact on antifouling research. In: *Advances in Marine Antifouling Coatings and Technologies*, ed. by C. Hellio, D. Yebra (CRC, Boca Raton 2009) pp. 46–79
37.84 N. Clarkson: The antifouling potential of silicone elastomer polymers. In: *Recent Advances in Marine Biotechnology*, ed. by M. Fingerman, R. Nagabhushanam, M.F. Thompson (Science, Enfield 1999) pp. 87–108
37.85 P. Willemsen: Biofouling in European aquaculture: Is there an easy solution?, Eur. Aquac. Soc. Spec. Publ. **35**, 101–118 (2005)
37.86 A.L. Lim, R. Bai: Membrane fouling and cleaning in microfiltration of activated sludge wastewater, J. Membr. Sci. **216**, 279–290 (2003)
37.87 T. Vladkova: Surface engineering for non-toxic biofouling control, J. Univ. Chem. Technol. Metallurgy **43**, 239–256 (2007)
37.88 R. Venkatesan, P.S. Murthy: Macrofouling control in power plants. In: *Springer Series on Biofilms*, ed. by J.W. Costerton (Springer, Berlin, Heidelberg 2008) pp. 265–290
37.89 C.P. Cologer: Six year interaction of underwater cleaning with copper based antifouling paints on navy surface ships, Nav. Eng. J. **96**, 200–208 (1984)
37.90 N. Sharma, C.H. Charles, M.C. Lynch, J. Qaqish, J.A. McGuire, J.G. Galustians, L.D. Kumar: Adjunctive benefit of an essential oil-containing mouth rinse in reducing plaque and gingivitis in patients who brush and floss regularly, J. Am. Dent. Assoc. **135**, 496–504 (2004)
37.91 C.H. Charles, K.M. Mostler, L.L. Bartels, S.M. Mankodi: Comparative antiplaque and antigingivitis effectiveness of an chlorhexidine and an essential oil mouth rinse: 6-Month clinical trial, J. Clin. Periodontol. **31**, 878–884 (2004)
37.92 G.D. Bixler, B. Bhushan: Rice- and butterfly-wing effect inspired self-cleaning and low drag micro/nanopatterned surfaces in water, oil, and air flow, Nanoscale **6**, 76–96 (2014)
37.93 J.A. Finlay, M.E. Callow, M.P. Schultz, G.W. Swain, J.A. Callow: Adhesion strength of settled spores of the green alga enteromorpha, Biofouling **18**, 251–256 (2002)

37.94 I. Banerjee, R.C. Pangule, R.S. Kane: Antifouling coatings: Recent developments in the design of surfaces that prevent fouling by proteins, bacteria, and marine organisms, Adv. Mater. **23**, 690–718 (2011)

37.95 H. Ma, C.N. Bowman, R.H. Davis: Membrane fouling reduction by backpulsing and surface modification, J. Membr. Sci. **173**, 191–200 (2000)

37.96 J.T. Fletcher, J.A. Finlay, M.E. Callow, J.A. Callow, M.R. Ghadiri: A combinatorial approach to the discovery of biocidal six-residue cyclic D,L-α-peptides against the bacteria methicillin-resistant Staphylococcusaureus (MRSA) and *E. coli* and the biofouling algae Ulva linza and Navicula perminuta, Chem. Eur. J. **13**, 4008–4013 (2007)

Part I Micro/NEMS

Part I Micro/Nanodevice Reliability

38 MEMS/NEMS and BioMEMS/BioNEMS: Tribology, Mechanics, Materials and Devices
Bharat Bhushan, Columbus, USA

39 Friction and Wear in Micro- and Nanomachines
Maarten P. de Boer, Pittsburgh, USA
Sameer S. Shroff, Pittsburgh, USA
Frank W. DelRio, Boulder, USA
W. Robert Ashurst, Auburn, USA

40 Failure Mechanisms in MEMS/NEMS Devices
W. Merlijn van Spengen, Amsterdam, The Netherlands
Robert Modliński, Munich, Germany
Robert Puers, Leuven, Belgium
Anne Jourdain, Leuven, Belgium

41 Mechanical Properties of Micromachined Structures
Harold Kahn, Charlottesville, USA

42 High Volume Manufacturing and Field Stability of MEMS Products
T. Kieran Nunan, Carlisle, USA
Mark G. da Silva, Wilmington, USA

43 Packaging and Reliability Issues in Micro/Nano Systems
Yu-Chuan Su, Hsinchu, Taiwan
Jongbaeg Kim, Seoul, Republic of Korea
Yu-Ting Cheng, Hsinchu, Taiwan
Mu Chiao, Vancouver, Canada
Liwei Lin, Berkely, USA

38. MEMS/NEMS and BioMEMS/BioNEMS: Tribology, Mechanics, Materials and Devices

Bharat Bhushan

A number of micro/nanoelectromechanical systems (MEMS/NEMS) and bioMEMS/bioNEMS are used in commercial applications and/or are under development. Surface area-to-volume ratio in MEMS/NEMS is large, and surface forces such as adhesion, friction, and meniscus and viscous forces become very large compared to inertial and electromagnetic forces. Some devices are designed to execute expected functions with short durations, typically in the millisecond to picosecond range. The expected life of the devices for high-speed contacts can vary from a few hundred thousand to many billions of cycles, e.g., over a hundred billion cycles for digital micromirror devices (DMDs). This puts serious requirements on materials in systems involving relative motion. Thus, there is a need for a fundamental understanding of adhesion, friction/stiction, wear, lubrication, and the role of surface contamination and environment, all on the nanoscale. Most mechanical properties are known to be scale dependent. Therefore, the properties of nanoscale structures need to be measured. For bioMEMS/bioNEMS, adhesion between biological molecular layers and the substrate, and friction and wear of biological layers can be important. Component-level studies are required to provide a better understanding of the tribological phenomena occurring in MEMS/NEMS. This chapter presents an overview of nanoscale adhesion, friction, and wear studies of materials and lubrication for MEMS/NEMS and bioMEMS/bioNEMS, and component-level studies of stiction phenomena in MEMS/NEMS devices.

38.1 **MEMS/NEMS Basics** 1332
38.1.1 Introduction to MEMS 1334
38.1.2 Introduction to NEMS 1335
38.1.3 Introduction to BioMEMS/BioNEMS 1336

38.2 **Nanotribology and Nanomechanics Issues in MEMS/NEMS and BioMEMS/BioNEMS** 1336
38.2.1 MEMS 1337
38.2.2 NEMS 1342
38.2.3 BioMEMS 1344
38.2.4 BioNEMS 1348
38.2.5 Microfabrication Processes 1351
38.2.6 Tribological Needs 1351

38.3 **Nanotribology and Nanomechanics Studies of Silicon and Related Materials** 1352
38.3.1 Virgin and Treated/Coated Silicon Samples 1352
38.3.2 Nanotribological and Nanomechanical Properties of Polysilicon Films and SiC Films 1357

38.4 **Lubrication Studies for MEMS/NEMS** 1359
38.4.1 Perfluoropolyether Lubricants 1359
38.4.2 Self-Assembled Monolayers (SAMs) 1361
38.4.3 Hard Diamond-Like Carbon (DLC) Coatings 1365

38.5 **Nanoscale Friction, Wear, and Mechanical Behavior of Nano-Objects** 1365
38.5.1 Single Nano-Object Friction 1368
38.5.2 Multiple Nano-Object Contact 1369
38.5.3 Summary 1375

38.6 **Nanotribological Studies of Biological Molecules on Polystyrene and Silicon Surfaces and Coated Polymer Surfaces** 1376
38.6.1 Nanoscale Adhesion, Friction, and Wear of Protein Layers on Polystyrene Surfaces 1376
38.6.2 Adhesion, Friction, and Wear of Biomolecules on Si-Based Surfaces . 1382
38.6.3 Adhesion of Coated Polymer Surfaces .. 1389

38.7 **Trajectory of Submicron Particles for Therapeutics and Diagnostics** 1390

38.8 **Component-Level Studies** 1392
38.8.1 Surface Roughness Studies of Micromotor Components 1392
38.8.2 Adhesion Measurements of Microstructures 1393
38.8.3 Microtriboapparatus for Adhesion, Friction, and Wear of Microcomponents 1394
38.8.4 Static Friction Force (Stiction) Measurements in MEMS 1397

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_38

38.8.5 Mechanisms Associated with Observed Stiction Phenomena in Digital Micromirror Devices (DMDs) and Nanomechanical Characterization 1400
38.9 **Conclusion** 1403
38.A **Appendix: Micro/Nanofabrication Techniques** 1404
38.A.1 Top-Down 1404
38.A.2 Bottom-Up Fabrication (Nanochemistry) 1407
References 1407

38.1 MEMS/NEMS Basics

Micro- & nanosystems include micro/nanoelectromechanical systems (MEMS/NEMS). The acronym MEMS originated in the US. The term commonly used in Europe is micro/nanosystem technology (MST/NST), and in Japan it is micro/nanomachines. Another term generally used is micro/nanodevices. Advances in silicon photolithographic process technology since the 1960s led to the development of MEMS in the early 1980s. MEMS combine electrical and mechanical components to create microscopic devices that have a characteristic length between 100 nm and 1 mm. In the early 2000s, with the development of nanofabrication techniques, NEMS were fabricated. NEMS combine electrical and mechanical components to create nanoscopic devices that have a characteristic length of less than 100 nm. In mesoscale devices, if the functional components are on the micro- or nanoscale, they may be referred to as MEMS or NEMS, respectively. These are intelligent miniaturized systems comprised of sensing, processing, and/or actuating functions that combine electrical and mechanical components [38.1]. MEMS/NEMS terms are also now used in a broad sense and include electrical, mechanical, fluidic, optical, and/or biological function. MEMS/NEMS for optical applications are referred to as micro/nanooptoelectromechanical systems (MOEMS/NOEMS). MEMS/NEMS for electronic applications are referred to as radio-frequency-MEMS/NEMS or RF-MEMS/RF-NEMS. MEMS/NEMS for biological applications are referred to as bioMEMS/bioNEMS.

To put the dimensions and weights of nanomaterials, MEMS/NEMS, and bioNEMS in perspective, see Fig. 38.1 and Table 38.1. Examples shown are a single-walled carbon nanotube (SWNT) chemical sensor (adapted from [38.2]), molecular dynamic simulations of carbon-nanotube-based gears (adapted from [38.3]), a quantum-dot transistor (adapted from [38.4]), and a digital micromirror device (adapted from [38.5]). Individual atoms are typically a fraction of a nanometer in diameter, DNA molecules are about 2.5 nm wide, biological cells are in the range of thousands of nm in diameter, and a human hair is about 75 000 nm (75 μm) in diameter. The smaller length of a bioNEMS shown in the figure is about 2 nm. NEMS shown range in size from 10 to 300 nm, and the size of MEMS is on the order of 12 000 nm (12 μm). The weight of a micromachined silicon structure can be as low as 1 nN, and NEMS can be built with weight as low as 10^{-20} N with cross-sections of about 10 nm. In comparison, the weight of a drop of water is about 10 μN, and the weight of an eyelash is about 100 nN.

These devices are fabricated by manufacturing at the nanoscale, referred to as nanomanufacturing. It is accomplished by using either a *bottom-up* or *top-down* approach to the production of nanomaterials, structures, devices, and systems [38.1, 7, 8].

In the bottom-up approach, nanoscale features are mainly built up from their elemental constituents. Elemental constituents are combined using various nanomanufacturing processes including self-assembly, chemical synthesis, molecular beam epitaxy, nanoimprint lithography, roll-to-roll processing, and dip-pen lithography [38.9]. Molecular self-assembly, the spontaneous self-assembly of molecular clusters, can occur from simple reagents in solution or from biological molecules (e.g., DNA) used as building blocks for the production of three-dimensional nanostructures. Chemical synthesis is carried out using gaseous precursors and solvents. These chemicals react to produce nanostructures. A variety of vacuum deposition and nonequilibrium plasma chemistry techniques are then used to produce layered nanocomposites and nanotubes. Molecular beam epitaxy is a method for deposition of thin films with a thickness of one atom on a surface. Nanoprint lithography is used to fabricate nanostructures by stamping or printing them on a surface. Roll-to-roll processing is a high-volume process used to produce nanodevices on a flexible substrate. Dip pen lithography uses an atomic force microscope tip in a fluid to produce nanoscale features on a surface.

The top-down approach uses lithographic and nonlithographic fabrication technologies. Lithographic technology is an underlying technology to make semi-

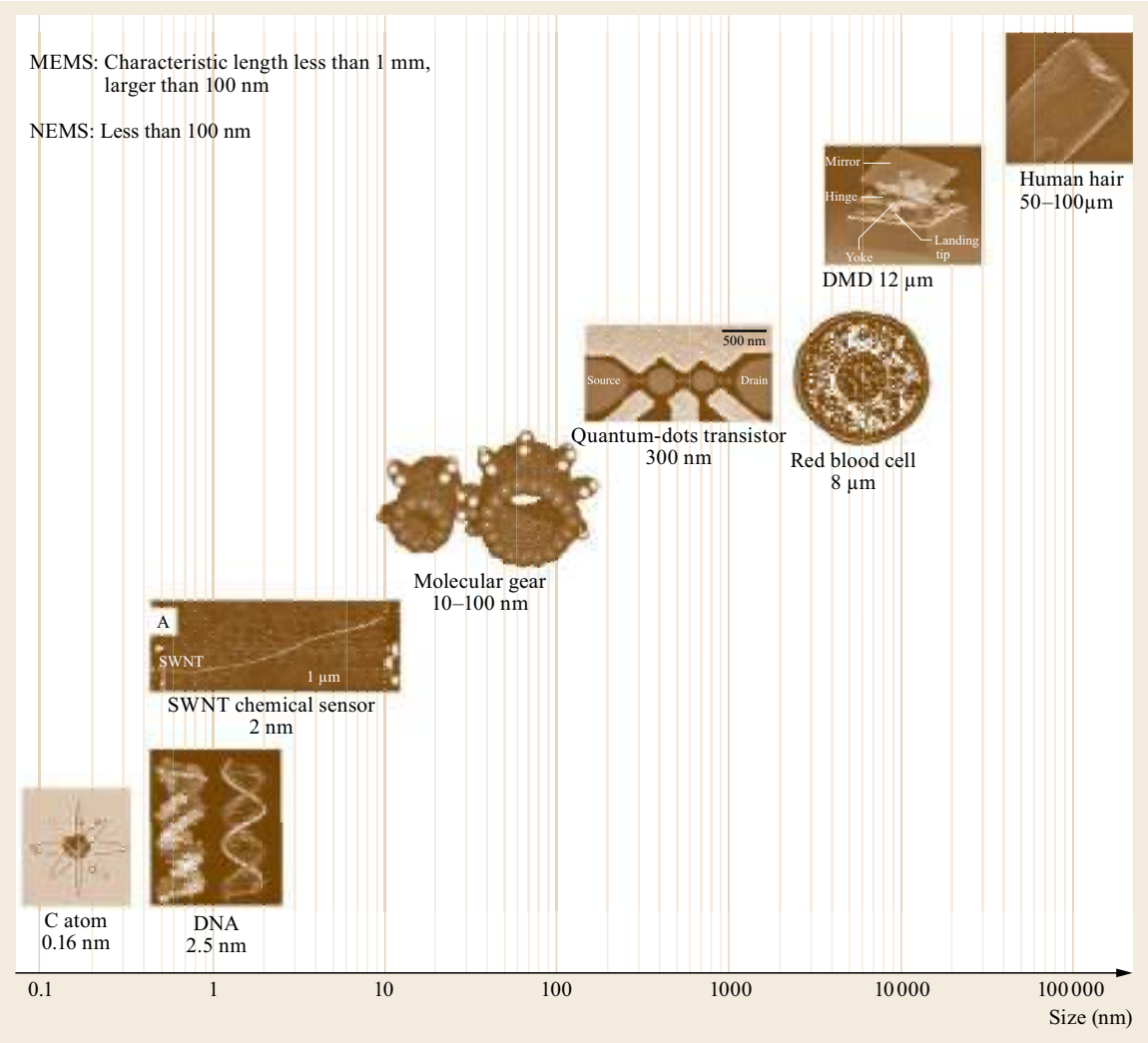


Fig. 38.1 Characteristic dimensions of nanomaterials, MEMS/NEMS, and bioNEMS in perspective (adapted from [38.6]). Examples shown are a single-walled carbon nanotube (SWNT) chemical sensor (adapted from [38.2]), molecular dynamic simulations of carbon-nanotube-based gears (adapted from [38.3]), a quantum-dot transistor (adapted from [38.4]), and a digital micromirror device (adapted from [38.5])

conductor chips and components. Continued improvements in lithography for use in the production of nanocomponents have resulted in line widths as small as 10 nm. The top-down approach is commonly used to fabricate micro- and nanosystem components, which range in size from micro- to nanometers.

MEMS/NEMS and bioMEMS/bioNEMS are beginning to have a major impact on our lives, comparable to that of semiconductor technology, information technology, or cellular and molecular biology in the 20th century. The integration of micro/nanodevices/systems into larger systems has provided breakthrough solutions to many current environmental, medical, and industrial problems including smart materials, nanomanufacturing, electronics, drug delivery, energy, water, biotechnology, information technology, and national security. bioMEMS/bioNEMS are increasingly used in commercial and defense applications. The revenues from nano-enabled products continue growing, with over $200 B in fiscal year 2012 in the US alone, and over $700 B worldwide [38.6]. MEMS/NEMS and bioMEMS/bioNEMS will have a profound impact on our society. They represent a megatrend, bringing disruptive innovation.

Table 38.1 Characteristic dimensions and weights in perspective

Characteristic dimensions in perspective	
NEMS characteristic length	< 100 nm
MEMS characteristic length	< 1 mm and > 100 nm
Single-walled carbon nanotube (SWNT) chemical sensor	≈ 2 nm
Molecular gear	≈ 10 nm
Quantum-dot transistor	300 nm
Digital micromirror	12 000 nm
Individual atoms	Typically a fraction of a nm in diameter
DNA molecules	≈ 2.5 nm wide
Biological cells	In the range of thousands of nm in diameter
Human hair	$\approx 75\,000$ nm in diameter
Weights in perspective	
NEMS built with cross-sections of ≈ 10 nm	As low as 10^{-20} N
Micromachines, silicon structure	As low as 1 nN
Eyelash	≈ 100 nN
Water droplet	$\approx 10\,\mu$N

38.1.1 Introduction to MEMS

The advances in silicon photolithographic process technology since the 1960s led to the development of MEMS in early 1980s. Later, lithographic processes were developed to process nonsilicon materials. The lithographic processes are complemented with nonlithographic processes for the fabrication of components or devices made from plastics or ceramics. Using these fabrication processes, researchers have fabricated a wide variety of devices with dimensions in the range of submicrons to a few thousand microns (see e.g., [38.7, 10–19]). MEMS for mechanical applications include acceleration, pressure, flow, and gas sensors, linear and rotary actuators, and other microstructures or microcomponents such as electric motors, gear chains, gas turbine engines, fluid pumps, fluid valves, switches, grippers, and tweezers. MEMS for chemical applications include chemical sensors and various analytical instruments. MOEMS are the devices that include optical components, such as micromirror arrays for displays, infrared image sensors, spectrometers, bar code readers, and optical switches. RF-MEMS include inductors, capacitors, antennas, and RF-switches. High-aspect-ratio MEMS (HARMEMS) have also been introduced.

A wide variety of MEMS devices have been produced and some are used commercially [38.7, 11, 13–16, 18, 19]. A variety of sensors are used in industrial, consumer, defense, and biomedical applications. The largest *killer* industrial applications include accelerometers, pressure sensors, thermal and piezoelectric-type inkjet printheads, and digital micromirror devices. Integrated capacitive-type silicon accelerometers have been used in airbag deployment in automobiles since 1991 [38.20, 21]. Some 90 million units were installed in vehicles in 2004. Accelerometer technology was over a billion-dollar-a-year industry in 2004, dominated by Analog Devices followed by Freescale Semiconductor (formerly Motorola) and Bosch. It is expected to grow with an annual growth exceeding 30%. Tri-axis accelerometers are needed to describe three-dimensional motion. Multi-axis accelerometers are being used for many other applications such as vehicle stability, rollover control, and gyro sensors for automotive applications. They are used also in various consumer applications including handheld devices, e.g., laptops for free-fall detection (2003), and cellular phones (2004) and tablets for menu navigation, gaming, image rotation, and free-fall detection. Silicon-based piezoresistive pressure sensors were launched in 1990 by GE NovaSensor for manifold absolute pressure (MAP) sensing for engines and for disposable blood pressure sensors, and their annual sales in 2004 were more than 30 million units and more than 25 million units respectively. MAP sensors measure the pressure in the intake manifold, which is fed to a computer that determines the optimum air/fuel mixture to maximize fuel economy. Most vehicles have these as part of the electronic engine control system. Capacitive pressure sensors for tire pressure measurements were launched by Freescale Semiconductor (formerly Motorola) in early 2000, and are also manufactured by Infineon/SensoNor and GE Novasensor (2003). Piezoresistive-type sensors are also used, manufactured by various companies such as EnTire Solutions (2003). The sensing module is located inside the rim of the wheel and relays the information via radio frequency to a central processing

unit (CPU) in order to display it to the driver. In 2005, about 9.2 million vehicles were equipped with sensors, which translated to about 37 million units. Their sales have grown rapidly as they are now required in automobiles in the US (starting in 2008), and it affects 17 million vehicles (one in each tire) sold every year. Pressure sensors can be used to detect altitude by measuring air pressure. For example, global positioning systems (GPS) used for navigation have good lateral resolution but poor vertical resolution, which creates problems in identifying the level in the case of multiple roads stacked up. A pressure sensor is needed to determine the level (altitude) by measuring the air pressure.

Thermal inkjet printers were developed independently by HP and Cannon and were commercialized in 1984 [38.22–25]. Today they are made by Canon, Epson, HP, Lexmark, Xerox, and others. They typically cost less initially than dry-toner laser printers and are the solution of choice for low-volume print runs. Annual sales of thermal inkjet printheads with microscale functional components were more than 500 million units in 2004.

Micromirror arrays are used for digital projection displays. Commercial digital light processing (DLP) equipment, using digital micromirror devices (DMDs), were launched in 1996 by Texas Instruments (TI) for digital projection displays in computer projectors, high-definition television (HDTV) sets, and movie projectors (DLP cinema) [38.5, 26, 27]. Several million projectors had been sold by 2004 (about $700 million revenue by TI in 2004). Electrostatically actuated, membrane-type or cantilever-type microswitches have been developed for direct current (DC), RF, and optical applications [38.28]. There exists two basic forms of RF microswitches: the metal-to-metal contact microswitch (ohmic) and the capacitive microswitch. RF microswitches can be used in a variety of RF applications including cellular phones, phase shifters, smart antennas, multiplexers for data acquisition, and more [38.29]. Optical microswitches are finding applications in optical networking, telecommunications, and wireless technologies [38.28, 30].

Other applications of MEMS devices include chemical, biological, and gas sensors [38.19, 31]; actuators, microresonators, infrared detectors, and focal plane arrays for earth observation, space science, and missile defense applications; picosatellites for space applications; fuel cells; and many hydraulic, pneumatic, and other consumer products. MEMS devices are also being pursued in magnetic storage systems [38.32], where they are being developed for super-compact and ultrahigh-recording-density magnetic disk drives. Several integrated head/suspension microdevices have been fabricated for contact recording applications [38.33]. High-bandwidth servo-controlled microactuators have been fabricated for ultrahigh-track-density applications, which serve as the fine-position control element of a two-stage, coarse/fine servo system, coupled with a conventional actuator [38.34, 35].

Micro/nano-instruments and micro/nanomanipulators are used to move, position, probe, pattern, and characterize nanoscale objects and nanoscale features [38.36]. Miniaturized analytical equipment includes gas chromatography and mass spectrometry. Other instruments include micro-STM, where STM stands for scanning tunneling microscope.

In some cases, MEMS devices are used primarily for their miniature size, while in others, as in the case of the air bags, because of their low-cost manufacturing techniques. This latter fact has been possible since semiconductor-processing costs have reduced drastically over the last decade, allowing the use of MEMS in many fields.

38.1.2 Introduction to NEMS

NEMS are produced by nanomachining in a typical top-down approach (from large to small) and bottom-up approach (from small to large) largely relying on nanochemistry (see e.g., [38.37–43]). The NEMS field, in addition to the fabrication of nanosystems, has provided the impetus to develop experimental and computation tools. Examples of NEMS include microcantilevers with integrated sharp nanotips for STM and atomic force microscopy (AFM) [38.44, 45], quantum corral formed using STM by placing atoms one by one [38.46], AFM cantilever arrays (Millipede) for data storage [38.47], STM and AFM tips for nanolithography, dip-pen nanolithography for printing molecules, nanowires, carbon nanotubes, quantum wires (QWRs), quantum boxes (QBs), quantum transistors [38.4], nanotube-based sensors [38.48, 49], biological (DNA) motors, molecular gears by attaching benzene molecules to the outer walls of carbon nanotubes [38.3], devices incorporating nm-thick films (e.g., in giant magnetoresistive (GMR) read/write magnetic heads and magnetic media for magnetic rigid disk and magnetic tape drives), nanopatterned magnetic rigid disks, and nanoparticles (e.g., nanoparticles in magnetic tape substrates and nanomagnetic particles in magnetic tape coatings) [38.32, 50]. More than two billion read/write magnetic heads were shipped for magnetic disk and tape drives in 2004.

Nanoelectronics can be used to build computer memory using individual molecules or nanotubes to store bits of information [38.51], molecular switches, molecular or nanotube transistors, nanotube flat-panel displays, nanotube integrated circuits, fast logic gates,

switches, nanoscopic lasers, and nanotubes as electrodes in fuel cells.

38.1.3 Introduction to BioMEMS/BioNEMS

The basic building block of all life is the cell [38.52]. Lipid bilayer membranes enclose all animal cells. The internal compartment is a hydrophilic space called cytosol, and is filled with the material necessary to sustain life. The human body consists of on the order of 30 trillion human cells, dominated by red blood cells, and on the order of 40 trillion bacterial cells, dominated by colonic bacteria. Bacteria help digest food, but they can also generate toxins [38.53]. Proteins are large molecules and consist of long chains of amino acids [38.54]. Complete proteins contain all eight essential amino acids. Proteins are a vital part of all living things. They are important to every cell operation, and perform a vast array of functions including catalyzing metabolic reactions, DNA replication, and stimuli response. However, their main function is to heal wounds, fight infection, and build muscle.

BioMEMS/bioNEMS are used for chemical and biochemical analyses (biosensors) in medical diagnostics (e.g., DNA, RNA, proteins, cells, blood pressure and assays, and bacterial pathogens or toxin identification) [38.55, 56], tissue engineering [38.57–59], and implantable pharmaceutical drug delivery [38.60–62]. They are increasingly used in commercial and defense applications (see e.g., [38.56, 63–69]).

Biosensors, also referred to as biochips, deal with liquids and gases. A large variety of biosensors are based on micro/nanofluidics [38.56, 70–73]. Micro/nanofluidic devices offer the ability to work with smaller reagent volumes and shorter reaction times, and perform multiple types of analyses at once. A second type of biosensors includes micro/nanoarrays, which perform one type of analysis thousands of times [38.74–80].

A chip with micro/nanofluidic technology embedded on the disk, called lab-on-a-CD, can test thousands of biological samples rapidly and automatically [38.70]. An entire laboratory can be integrated onto a single chip, called lab-on-a-chip [38.56, 71, 72]. Silicon-based disposable blood-pressure sensor chips were introduced in early 1990s by GE NovaSensor for blood pressure monitoring (about 25 million units in 2004). A blood sugar monitor, referred to as GlucoWatch, was introduced in 2002. It automatically checks blood sugar every 10 minutes by detecting glucose through the skin, without having to draw blood. If glucose is out of the acceptable range, it sounds an alarm so the diabetic can address the problem quickly. A variety of biosensors, many using plastic substrates, are manufactured by various companies including ACLARA, Agilent Technologies, Calipertech, and I-STAT.

Micro/nanoarrays are another type of biochip. They are a tool used in biotechnology research to analyze DNA or proteins to diagnose diseases or discover new drugs. Also called DNA arrays, they can identify thousands of genes simultaneously [38.66, 74]. They include a microarray of silicon nanowires, roughly a few nm in size, to selectively bind and detect even a single biological molecule such as DNA or protein by using nanoelectronics to detect the slight electrical charge caused by such binding, or a microarray of carbon nanotubes to electrically detect glucose.

After the tragedy of September 11, 2001, concern about biological and chemical warfare has led to the development of handheld units with bio- and chemical sensors for detection of biological toxins, chemical or nerve agents, mustard agents, and chemical precursors to protect subways, airports, water supply, and the population at large [38.81].

BioMEMS/bioNEMS are also being developed for minimally invasive surgery including endoscopic surgery, laser angioplasty, and microscopic surgery. Implantable artificial organs can also be produced. Other applications include implantable drug-delivery devices, e.g., micro/nanoparticles with drug molecules encapsulated in functionalized shells for site-specific targeting applications; a silicon capsule with a nanoporous membrane filled with drugs for long-term delivery [38.60, 82–84]; nanodevices for sequencing single molecules of DNA in the Human Genome Project [38.56]; cellular growth using carbon nanotubes for spinal cord repair; and nanotubes for nanostructured materials for various applications such as spinal fusion devices, organ growth, and growth of artificial tissues using nanofibers.

38.2 Nanotribology and Nanomechanics Issues in MEMS/NEMS and BioMEMS/BioNEMS

Tribological issues are important in MEMS/NEMS and bioMEMS/bioNEMS requiring intended and/or unintended relative motion. Nanomechanical properties are scale dependent [38.85]; therefore, these should be measured at relevant scales. In these devices, various forces associated with the device scale down as size decreases. When the length decreases from 1 mm to 1 μm, the surface area decreases by a factor of a million, and the volume decreases by a factor of a billion. As a result, surface forces such as adhesion, friction,

meniscus forces, and viscous forces that are proportional to surface area, become a thousand times larger than the forces proportional to the volume, such as inertial and electromagnetic forces. In addition to the consequences of large surface-to-volume ratios, the small tolerances that these devices are designed for make physical contacts more likely, which makes them particularly vulnerable to adhesion between adjacent components. Slight particulate or chemical contamination present at the interface can be detrimental. Further, the small startup forces and the torques available to overcome retarding forces are small, and the increase in resistive forces such as adhesion and friction become a serious tribological concern that limits the durability and reliability of MEMS/NEMS [38.1, 13, 44, 86–88]. A large lateral force required to initiate relative motion between two surfaces, large static friction, is referred to as *stiction*, which has been studied extensively in the tribology of magnetic storage systems [38.32, 44, 88–92]. The source of stiction is generally liquid-mediated adhesion with the source of liquid being process fluid or capillary condensation of the water vapor from the environment. Adhesion, friction/stiction, wear, and surface contamination affect MEMS/NEMS and bioMEMS/bioNEMS performance, and in some cases can even prevent the devices from working. Adhesion between a biological molecular layer and the substrate (referred to as *bioadhesion*), reduction of friction and wear of biological layers, biocompatibility, and biofouling in bioMEMS/bioNEMS can be important [38.75, 93]. Some examples of devices that experience nanotribological problems follow.

38.2.1 MEMS

Figure 38.2 shows examples of several microcomponents that can encounter tribological problems. The polysilicon electrostatic micromotor has 12 stators and a four-pole rotor, and is produced by surface micromachining. The rotor diameter is 120 μm, and the air gap between the rotor and stator is 2 μm [38.94]. It is capable of continuous rotation up to speeds of 100 000 RPM. The intermittent contact at the rotor–stator interface and physical contact at the rotor–hub flange interface result in wear issues, and high stiction between the contacting surfaces limits the repeatability of operation or may even prevent the operation altogether. Next, a bulk micromachined silicon stator/rotor pair is shown with bladed rotor and nozzle guide vanes on the stator with dimensions less than 1 mm [38.95, 96]. These are being developed for a high-temperature microgas turbine engine with rotor dimension of 4–6 mm in diameter, and an operating speed of up to one million RPM (with a sliding velocity in excess of 500 m/s, comparable to velocities of large turbines operating at high velocities) to achieve high specific power, up to a total of about 10 W. Erosion of blades and vanes and design of the microbearings required to operate at extremely high speeds used in the turbines are some of the concerns. Ultrashort, high-speed microhydrostatic gas journal bearings with a length-to-diameter ratio (L/D) of less than 0.1 are being developed for operation at surface speeds on the order of 500 m/s, which offer unique design challenges [38.97]. Microfabrica, Inc. in the US, is developing microturbines with an outer diameter as low as 0.9 mm to be used as power sources for medical devices. They use precision ball bearings.

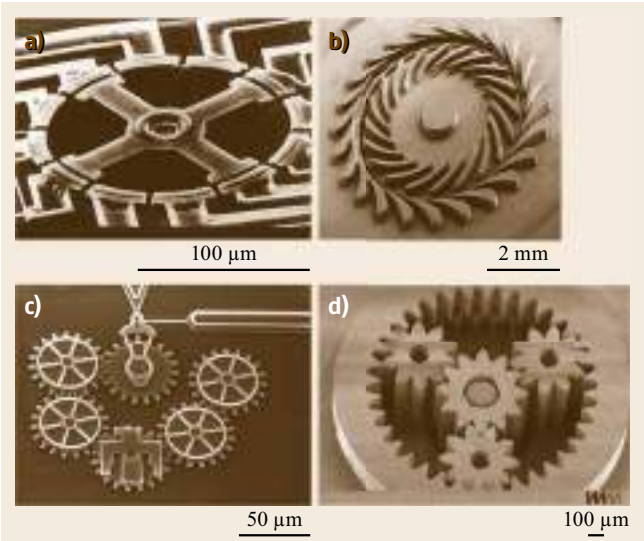


Fig. 38.2a–d Examples of MEMS components that experience tribological problems. **(a)** Electrostatic micromotor (after [38.94]). **(b)** Microturbine bladed rotor and nozzle vanes on the stator (after [38.95]). **(c)** Six-gear chain (Sandia). **(d)** Ni/Fe Wolfrom-type gear system produced by LIGA (after [38.98])

Next in Fig. 38.2 is a scanning electron microscopy (SEM) micrograph of a surface micromachined polysilicon six-gear chain from Sandia National Lab. (For more examples of early version, see [38.99].) As an example of nonsilicon components, a milligear system produced using the Lithographie, Galvanoformung, Abformung (LIGA) process for a DC brushless permanent magnet millimotor (diameter = 1.9 mm, length = 5.5 mm) with an integrated milligear box [38.98, 100, 101] is also shown. The gears are made of metal (electroplated Ni-Fe) but can also be made from injected polymer materials (e.g., polyoxy-methylene or POM) using the LIGA process (described in Appendix 38.A). Even though the torque transmitted at the gear teeth is small, on the order of a fraction of nN m, because of the small dimensions of gear teeth, the bending stresses are

large where the teeth mesh. Tooth breakage and wear at the contact of gear teeth is a concern.

Figure 38.3 shows an optical micrograph of a microengine driven by an electrostatically activated comb drive connected to the output gear by linkages. The microengine operates in the kHz frequency range, and can be used as a general drive and power source to drive micromechanisms [38.102]. Its parts are fabricated from polysilicon. A microgear unit is used to convert reciprocating motion from a linear actuator into circular motion. Another drive linkage oriented at 90° to the original linkage, driven by another linear actuator, allows maintenance of continuous motion. The linkages are connected to the output gear through pin joints that allow relative motion.

One inset shows a polysilicon, multiple microgear speed reduction unit and its components after laboratory wear tests conducted for 600 000 cycles at 1.8% relative humidity (RH) [38.103]. Wear of various components is clearly observed in the figure. Humidity was shown to be a strong factor in the wear of rubbing surfaces. In order to improve the wear characteristics of rubbing surfaces, a 20 nm-thick tungsten (W) coating was deposited at 450 °C using the chemical vapor deposition (CVD) technique [38.104]. Tungsten-coated microengines tested for reliability showed improved wear characteristics with longer lifetimes than polysilicon microengines. However, these coatings have poor yield. Instead, vapor deposited self-assembled monolayers of fluorinated (dimethylamino) silane have been used [38.105]. The second inset shows a comb drive with a deformed frame, which results in some fingers coming in contact. The contacting fingers can result in stiction.

Figure 38.4 shows a micromachined flow modulator. Several micromachined flow channels are integrated in series with electrostatically actuated microvalves [38.106]. The flow channels lead to a central gas outlet hole drilled in the glass substrate. Gas enters the device through a bulk micromachined gas inlet hole in the silicon cap. The gas, after passing through an open microvalve, flows parallel to the glass substrate through flow channels and exits the device through an outlet. The normally open valve structure consists of a freestanding double-end-clamped beam, which is positioned beneath the gas inlet orifice. When electrostatically deflected upwards, the beam seals against the inlet orifice and the valve is closed. In these microvalves used for flow control, the mating valve surfaces should be smooth enough to seal while maintaining a minimum roughness to ensure low adhesion [38.86, 88, 89, 91, 92, 107].

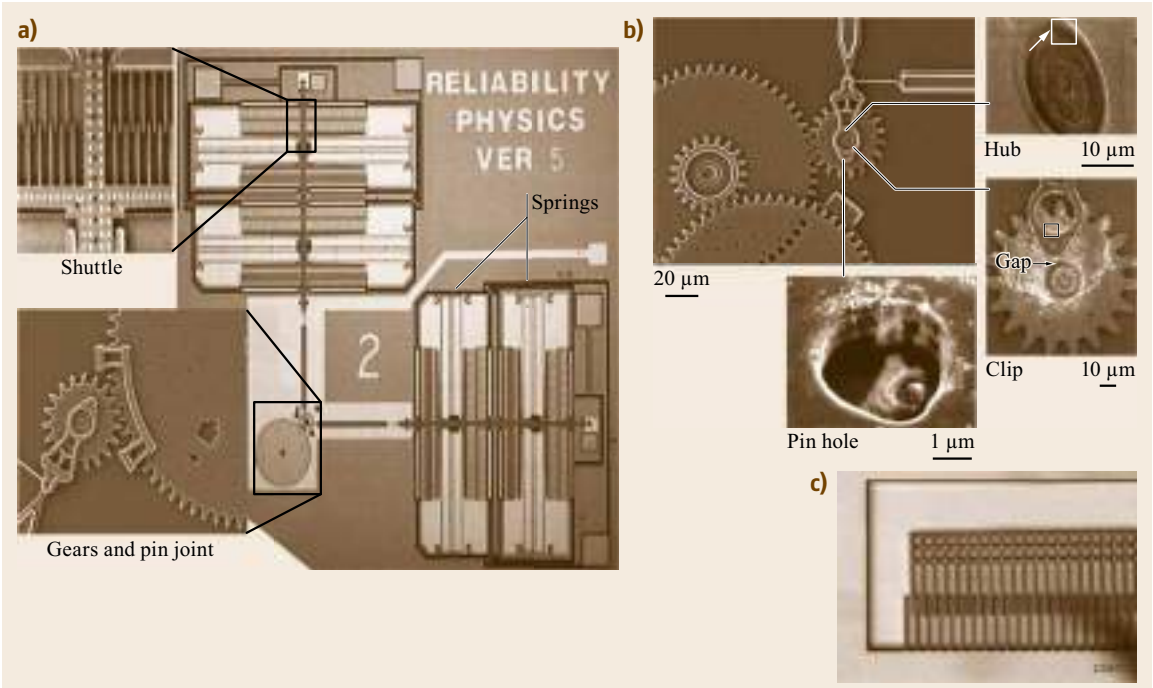


Fig. 38.3 **(a)** Optical micrograph of a microengine driven by an electrostatically actuated comb drive (microengine) fabricated by Sandia Summit Technologies (after [38.102]). **(b)** shows polysilicon microgear speed reduction unit after laboratory wear test for 600 000 cycles at 1.8% relative humidity (after [38.103]). **(c)** shows a stuck comb drive (CSEM)

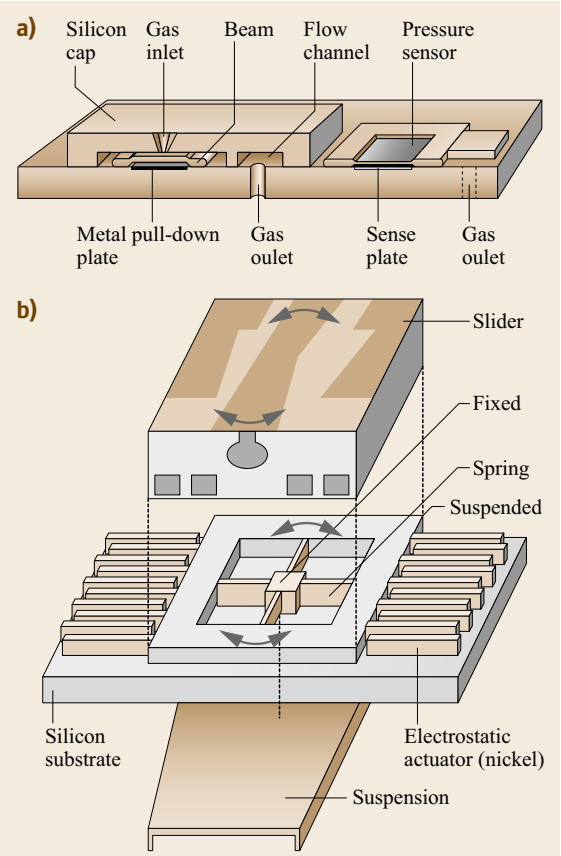


Fig. 38.4a,b Examples of MEMS devices that experience tribological problems. **(a)** Low-pressure flow modulator with electrostatically actuated microvalves (after [38.106]). **(b)** Electroplated-nickel rotary microactuator for magnetic disc drives (after [38.35])

A second MEMS device shown in Fig. 38.4 is an electrostatically driven rotary microactuator for a magnetic disk drive, surface micromachined by a multilayer electroplating method [38.35]. This high-bandwidth servo-controlled microactuator, located between a slider and a suspension, is being developed for ultrahigh-track-density applications, which serves as the fine-position and high-bandwidth control element of a two-stage, coarse/fine servo system when coupled with a conventional actuator [38.34, 35]. A slider is placed on top of the central block of a microactuator, which gives rotational motion to the slider. The bottom of the silicon substrate is attached to the suspension. The radial flexure beams in the central block give the rotational freedom of motion to the suspended mass (slider), and the electrostatic actuator drives the suspended mass. Actuation is accomplished via interdigitated, cantilevered electrode fingers, which are alternately attached to the central body of the moving part and to the stationary substrate to form pairs. A voltage applied across these electrodes results in an electrostatic force, which rotates the central block. The interelectrode gap width is about 2 μm. Any unintended contacts between the moving and stationary electroplated-nickel electrodes may result in wear and stiction.

Commercially available MEMS devices also exhibit tribological problems. Figure 38.5 shows an integrated capacitive-type silicon accelerometer fabricated using surface micromachining by Analog Devices, a couple of mm in dimension, which is used for the deployment of airbags in automobiles, and for various other devices in the consumer electronics market [38.20, 108]. The central suspended beam mass (about 0.7 μg) is supported on the four corners by spring structures. The central beam has interdigitated cantilevered electrode fingers (about 125 μm long and 3 μm thick) on all four sides that alternate with those of the stationary electrode fingers as shown, with about a 1.3 μm gap. Lateral motion of the central beam causes a change in the capacitance between these electrodes, which is used to measure the acceleration. Stiction between the adjacent electrodes as well as stiction of the beam structure with the underlying substrate, under isolated conditions, is detrimental to the operation of the sensor [38.20, 108]. Wear during unintended contacts of these polysilicon fingers is also a problem. A molecularly thick diphenyl siloxane lubricant film, resistant to high temperatures and oxidation, is applied by a vapor deposition process on the electrodes to reduce stiction and wear [38.109]. For deposition, a small amount of liquid is dispensed into each package before it is sealed. As the package is heated in the furnace, the liquid evaporates, which coats the sensor surface. As sensors are required to sense low g accelerations, they need to be more compliant and stiction becomes even a bigger concern.

Figure 38.5 also shows a cross-sectional view of a typical piezoresistive-type pressure sensor, which is used for various applications including manifold absolute pressure (MAP) and tire pressure measurements in automotive applications, and disposable blood pressure measurements. The sensing material is a diaphragm formed on a silicon substrate, which bends with applied pressure [38.110, 111]. The deformation causes a change in the band structure of the piezoresistors that are placed on the diaphragm, leading to a change in the resistivity of the material. The MAP sensors are subjected to drastic conditions – extreme temperatures, vibrations, sensing fluid, and thermal shock. Fluid under extreme conditions could cause corrosive wear, while fluid cavitation could cause erosive wear. The protective gel encapsulant generally used can react

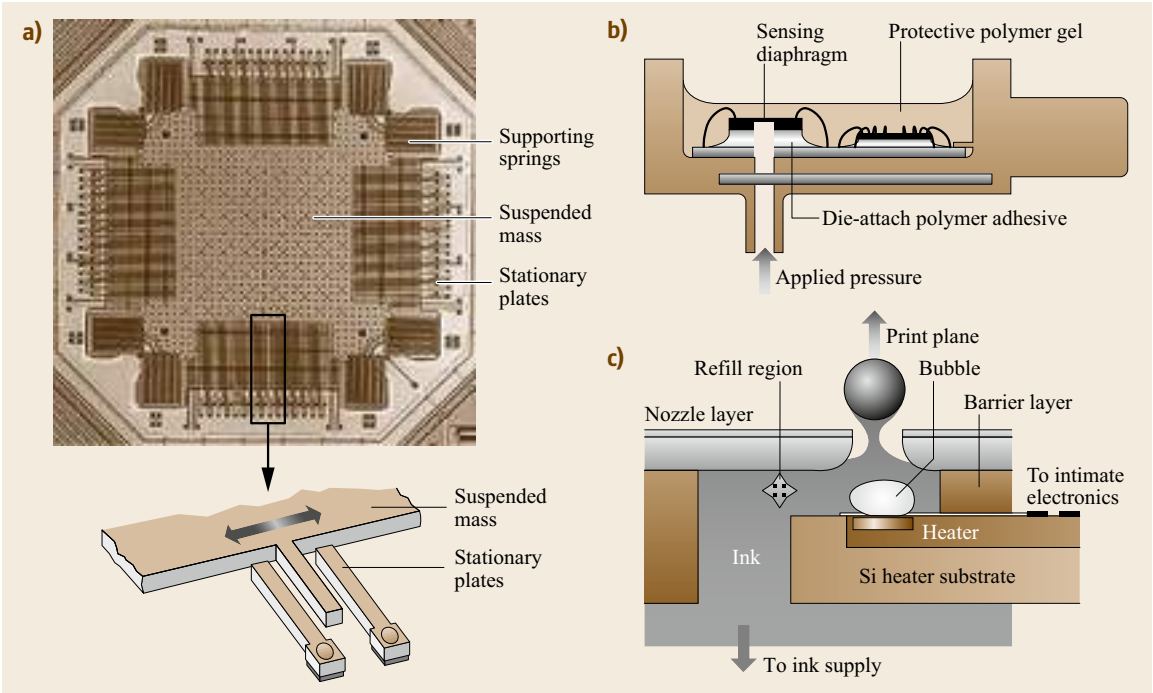


Fig. 38.5a–c Examples of commercial MEMS devices that experience tribological problems. **(a)** Capacitive-type silicon accelerometer for automotive sensory applications (after [38.108]). **(b)** Piezoresistive-type pressure sensor (after [38.110]). **(c)** Thermal inkjet printhead (after [38.24])

with sensing fluid and result in swelling or dissolution of the gel. Silicon cannot deform plastically, therefore any pressure spikes leading to deformation past its elastic limit will result in fracture and crack propagation. Pressure spikes could also cause the diaphragm to delaminate from the support substrate. Finally, cyclic loading of the diaphragm during use can lead to fatigue and wear of the silicon diaphragm or delamination.

The bottom schematic in Fig. 38.5 shows a cross-section of a thermal printhead chip (on the order of 10–50 cm^3 in volume) used in inkjet printers [38.24]. They comprise of a supply of ink and an array of elements with microscopic heating resistors on a substrate mated to a matching array of ink-injection orifices or nozzles (about 70 μm in diameter) [38.22, 23, 25]. In each element, a small chamber is heated by the resistor where a brief electrical impulse vaporizes part of the ink and creates a tiny bubble. The heaters operate at several kHz and are therefore capable of high-speed printing. As the bubble expands, some of the ink is pushed out of the nozzle onto the paper. When the bubble pops, a vacuum is created and this causes more ink from the cartridge to move into the printhead. Clogged ink ports are the major failure mode. There are various tribological concerns [38.22]. The surface of the printhead where the ink is shot out towards the paper can get scratched and damaged as a result of countless trips back and forth across the pages, which are somewhat rough. As a result of repeated heating and cooling, the heated resistors expand and contract. Over time, these elements will experience fatigue and may eventually fail. Bubble formation in the ink reservoir can lead to cavitation erosion of the chamber, which occurs when bubbles formed in the fluid become unstable and implode against the surface of the solid and impose impact energy on that surface. Fluid flow through nozzles may cause erosion and ink particles may also cause abrasive wear. Corrosion of the ink reservoir surfaces can also occur as a result of exposure of ink at high temperatures as well as due to ink pH. The substrate of the chip consists of silicon with a thermal barrier layer followed by a thin film of resistive material and then conducting material. The conductor and resistor layers are generally protected by an overcoat layer of a plasma-enhanced chemical vapor deposition (PECVD) α-SiC:H layer, 200–500 nm thick [38.112].

Figure 38.6 shows two DMD pixels used in DLP technology for digital projection displays in computer projectors, HDTV sets, and movie projectors [38.5, 26, 27]. The entire array (chip set) consists of a large number of oscillating aluminum alloy micromirrors as digital light switches, which are fabricated on top of

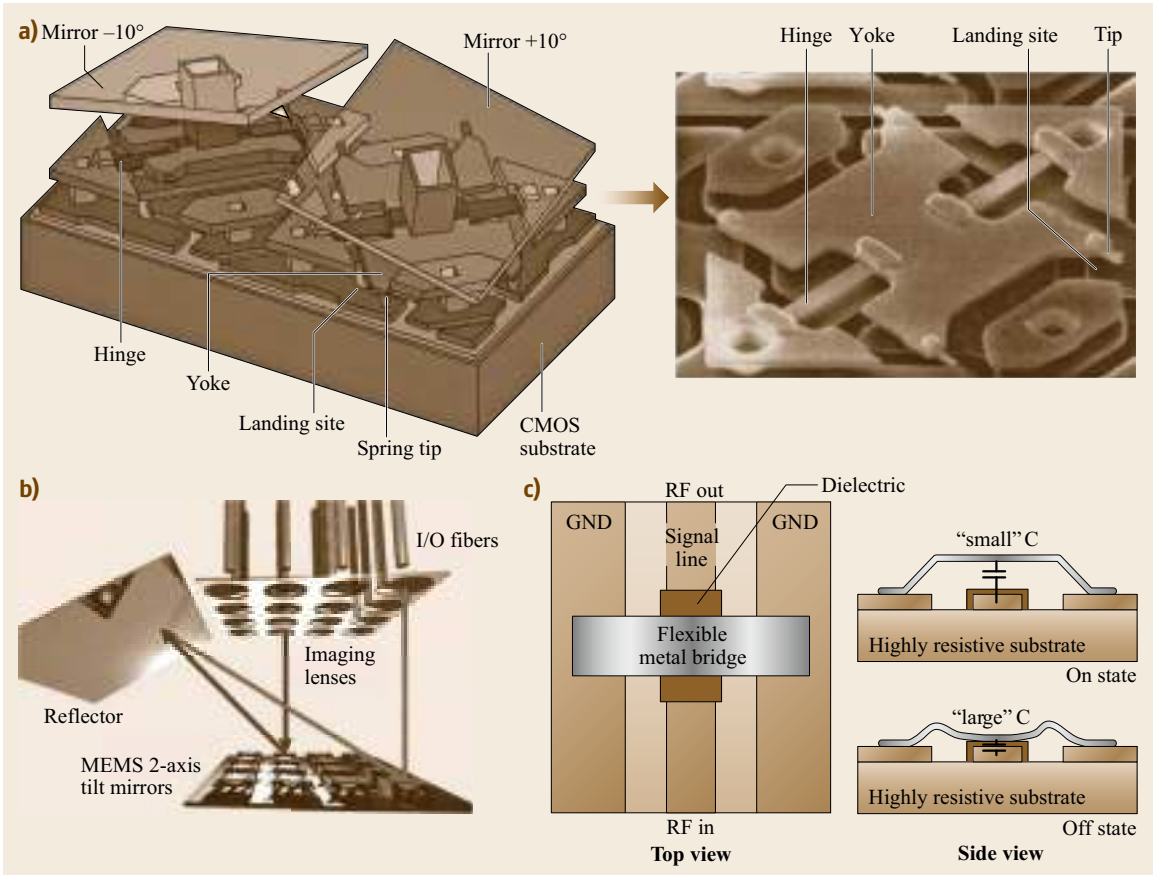


Fig. 38.6a–c Examples of two commercial MOEMS (**a**,**b**) and one RF-MEMS device (**c**) that experience tribological problems. (**a**) Digital micromirror device for displays (after [38.5]). (**b**) Tilt mirror arrays for switching optical signal input and output fiber arrays in optical cross-connect for telecommunications (after [38.113]). (**c**) RF microswitch (© IMEC, Belgium)

a complementary metal–oxide–semiconductor (CMOS) static random access memory integrated circuit. The surface micromachined array consists of half a million to more than two million of these independently controlled reflective micromirrors. Each micromirror is about 12 μm square and with 13 μm pitch, and flip backward and forward at a frequency on the order of 5–7 kHz as a result of electrostatic attraction between the micromirror structure and the underlying electrodes. For binary operation, the micromirror/yoke structure mounted on torsional hinges is oscillated ±10° (with respect to the plane on the chip set), and is limited by a mechanical stop. Contact between cantilevered spring tips at the end of the yoke (four present on each yoke) with the underlying stationary landing sites is required for true digital (binary) operation. Stiction and wear during contact between aluminum alloy spring tips and landing sites, hinge memory (metal creep at high operating temperatures), hinge fatigue, shock and vibration failure, sensitivity to particles in the chip package, and operating environment are some of the important issues affecting the operation reliability of a micromirror device [38.114–118]. A vapor-phase deposited self-assembled monolayer of the fatty acid perfluorodecanoic acid (PFDA) on the surfaces of tip and landing sites is used to reduce stiction and wear [38.119, 120]. However, these films are susceptible to moisture, and to keep moisture out and create a background pressure of PFDA, a hermetic chip package is used. The spring tip uses the stored spring energy to pop up the tip during pull-off. A lifetime estimate of over 100 000 operating hours with no degradation in image quality is typical. At a mirror modulation frequency of 7 kHz, each micromirror element needs to switch about 2.5 trillion cycles.

Figure 38.6 also shows a schematic of a 256×256-port large optical cross-connect, introduced in 2000 by Glimmerglass, Hayward, California, for optical

telecommunication networks in order to be able to rapidly manipulate a larger number of optical signals [38.113]. This optical microswitch uses 256 or more movable mirrors on a chip for switching a light beam from an input fiber to a few output fibers. The mirrors are made of gold-coated polysilicon and are about 500 μm in diameter. Reliability concerns are the same as those described for DMDs. To minimize stiction, the chip set is hermetically sealed in dry nitrogen (90% N_2, 10% He).

Figure 38.6 also shows a schematic of an electrostatically actuated, capacitive-type RF microswitch for switching of RF signals at microwave and low frequencies [38.121]. It is a membrane-type microswitch, and consists of a flexible metal (Al) bridge that spans the RF transmission line in the center of a coplanar waveguide. When the bridge is up, the capacitance between the bridge and RF transmission line is small, and the RF signal passes without much loss. When a DC voltage is applied between the RF transmission line and the bridge, the latter is pulled down until it touches a dielectric isolation layer. Thus, the large capacitance created shorts the RF signal to the ground. The failure modes include creep in the metal bridge, fatigue of the bridge, charging and degradation of the dielectric insulator, and stiction of the bridge to the insulator [38.28, 121–123]. Stiction occurs due to capillary condensation of water vapor from the environment, van der Waals forces, and/or charging effects [38.90–92]. If the restoring force in the bridge of the switch is not large enough to pull the bridge up again after the actuation voltage has been removed, the device fails due to stiction. Humidity-induced stiction can be avoided by hermetically sealing the microswitch. Some roughness of the surfaces reduces the probability of stiction. Selected actuation waveforms can be used to minimize charging effects.

38.2.2 NEMS

Probe-based data recording technologies have been explored for ultrahigh areal density recording where the probe tip (with a radius of about 5 nm) is expected to be scanned at velocities up to 100 mm/s. Major techniques include thermomechanical [38.47], phase change [38.124], and ferroelectric recording [38.125, 126]. We discuss the tribological issues with two of the widely pursued techniques [38.127].

Figure 38.7a shows the thermomechanical recording system that uses arrays of 1024 silicon microcantilevers (*Millipede*) and playback on a ≈ 40 nm-thick polymer medium with a harder Si substrate [38.47]. The cantilevers consist of integrated tip heaters with tips of nanoscale dimensions. Sharp tips themselves are also examples of NEMS. Thermomechanical recording is a combination of applying a local force to the polymer layer and softening it by local heating. The tip, heated to about 400 °C, is brought in contact with the polymer

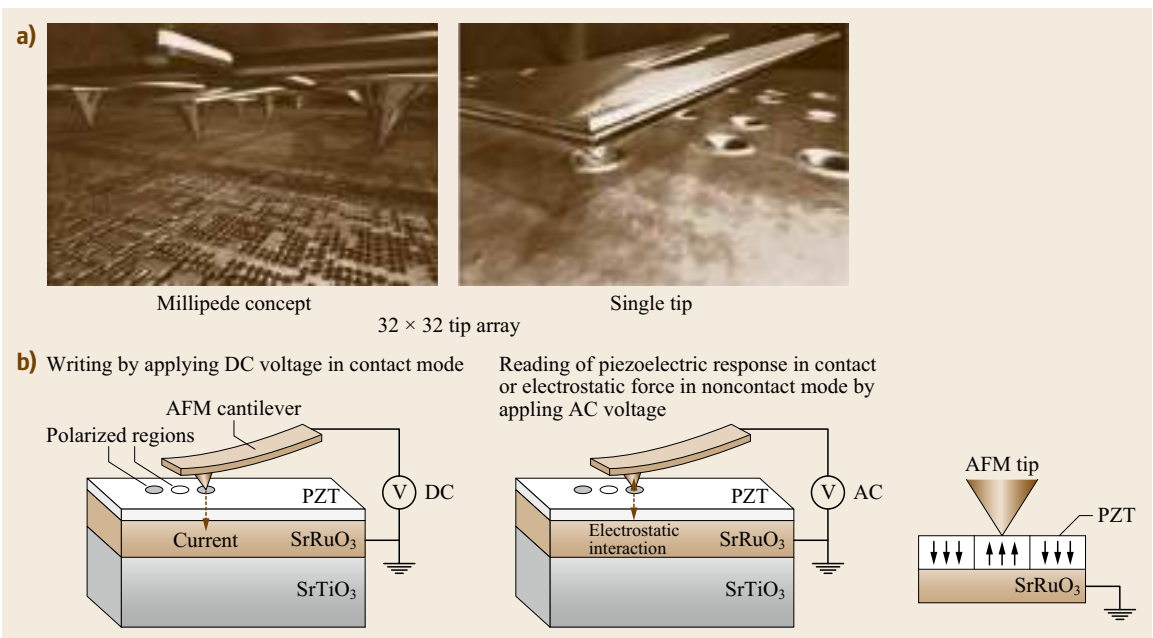


Fig. 38.7a,b Two example of NEMS devices – **(a)** thermomechanical recording, and **(b)** ferroelectric recording – that experience tribological problems

for recording. Readings are done using the heater cantilever, originally used for recording, as a thermal readback sensor by exploiting its temperature-dependent resistance. The principle of thermal sensing is based on the fact that the thermal conductivity between the heater and the storage substrate changes according to the spacing between them. When the spacing between the heater and sample is reduced as the tip moves into a bit, the heater's temperature and hence its resistance will decrease. Thus, changes in temperature of the continuously heated resistor are monitored while the cantilever is scanned over data bits, providing a means of detecting the bits. Erasing for subsequent rewriting is carried out by thermal reflow of the storage field by heating the medium to 150 °C for a few seconds. The smoothness of the reflown medium allows multiple rewriting of the same storage field. Bit sizes ranging between 10 and 50 nm have been achieved by using a 32×32 (1024) array write/read chip (3 mm × 3 mm). It has been reported that tip wear occurs due to contact between the tip and the Si substrate during writing. Tip wear is a major concern for device reliability.

Figure 38.7b shows the schematic of domain writing and reading in a ferroelectric film. The electrically conductive AFM tips are placed in contact with a piezoelectric (lead zirconate titanate, PZT), film-coated medium [38.125, 126]. Ferroelectric domains on PZT film are polarized by applying short voltage pulses (≈ 10 V, $\approx 100\,\mu$s) that exceed the coercive field of the PZT layer, resulting in local, nonvolatile changes in the electronic properties of the underlying film. The temperature rise during recording is expected to be on the order of 80 °C. Reading out of the polarization states in the ferroelectric film can be carried out using two different methods. In one method, the static surface charge, proportional to the normal component of polarization, can be detected by electrostatic force microscopy in the noncontact mode. In the second method, an AFM is operated in the contact mode and piezoresponse force is measured by applying an alternating-current (AC) voltage. Wear of the conducting tip and the PZT layer at high scanning velocities is a major concern for device reliability. Various lubricant films are being developed to minimize wear [38.124–130].

In magnetic data storage, magnetic recording is accomplished by relative motion between the magnetic head slider and a magnetic rigid disk [38.32]. Magnetic rigid disks and heads used today for magnetic data storage consist of nanostructured films a few nm thick. Figure 38.8a shows the sectional view of a conventional magnetic rigid disk. The superparamagnetic effect poses a serious challenge for the ever-increasing areal density of disk drives. One of the promising methods to circumvent the density limitations imposed by this effect is the use of a nanopatterned disk (Fig. 38.8b). In a conventional disk, the thin magnetic layer forms a random mosaic of nanometer-scale grains, and each recorded bit consists of many tens of these random grains. In a patterned disk, the magnetic layer is created as an ordered array of highly uniform islands, each island capable of storing an individual bit. These islands may be one or a few grains, rather than a collection of random decoupled grains. This increases the density by a couple of orders of magnitude. Figure 38.8c shows a schematic of an inductive write GMR read head structure. These are

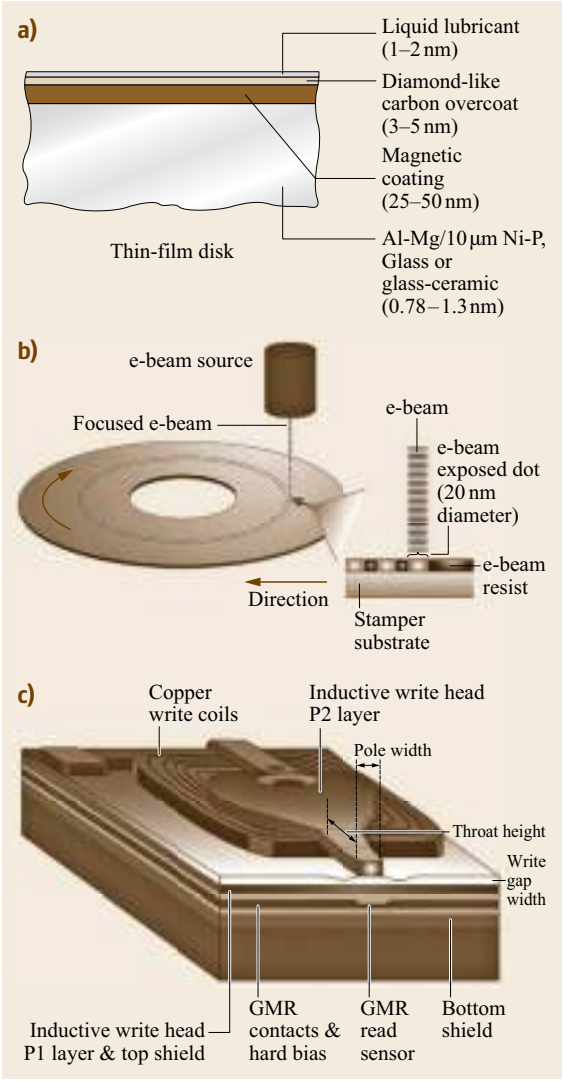


Fig. 38.8a–c Schematic of (a) sectional view of a conventional magnetic rigid disk, (b) nanopatterned magnetic rigid disk, and (c) an inductive write, GMR read magnetic head structure for magnetic data storage (Hitachi)

constructed from a variety of materials: magnetic alloys, metal conductors, ceramic, and polymer insulators in a complex three-dimensional structure. The multilayered thin-film structures used to construct the sensor and individual films are only a few nm thick. The head slider surface, which flies on the disk surface, is coated with about 3 nm thick of diamond-like carbon coatings to protect the thin-film structure from electrostatic discharge. Any isolated contacts between the disk and sensor and lubricant pickup pose tribological concerns [38.32].

Nano-objects are used in various applications where they come into sliding contact with each other and the surfaces where they are used, which can lead to nano-object deformation [38.8]. Some examples of these applications include drug delivery for cancer treatment (to be presented in the bioNEMS section), oil detection, contaminant removal, catalysis, and tribology on the macro- to nanoscale. For oil detection applications, as in the example shown in Fig. 38.9a, oxidized carbon black nanoparticles with a polyvinyl alcohol shell are coated with an oil-detecting agent. This agent is released on contact with hydrocarbons in oil, and its absence on recovery and analysis of the nano-objects is an indication of the presence of oil [38.131]. Agglomeration of nano-objects can prevent flow through porous media, which results in losses and prevents eventual recovery of the nano-objects [38.132]. Friction on surfaces and during flow in liquids can also hinder recovery. Low friction of nano-objects over surfaces and within liquids is of interest.

In contaminant removal, magnetic shell cross-linked knedel-like nanoparticles (MSCKs) are of interest. These are composites of polymer layers on iron oxide nanoparticles. In Fig. 38.9b, the MSCKs are added to polluted water (left), where they selectively absorb oil (center) and are finally removed by introducing a neodymium magnet [38.133]. Reducing agglomeration and fluid friction is important for efficient manipulation of nano-objects through liquids to contaminated sites.

In applications involving catalysis for hydrogen evolution recovery (HER), nanowires have been used to aid in the catalytic activity, as shown in Fig. 38.9c. The nanowires consist of an inner molybdenum trioxide (MoO_3) core and outer molybdenum disulfide (MoS_2) shell. The core is used for transporting electrons to edge sites of the outer MoS_2 shell, which is catalytically active. This is preferred over bulk MoS_2, which has a limited number of edge sites for the same volume of material compared to the nanowires. At these sites, the transfer of electrons necessary for the recovery of hydrogen from water is enhanced by the MoS_2, which improves HER efficiency [38.134]. Low friction of water moving over the nanowires is important for efficient flow over edge sites and conversion to hydrogen.

Nano-objects have demonstrated the ability to reduce friction and wear in applications on the macroscale. For applications on the micro/nanoscale, increasing the lifetime and efficiency of individual components of systems is crucial to the commercialization of micro/nanosystems. Adhesive and friction forces can hinder device operation and reliability [38.88, 91, 92]. In these environments, nano-objects can be deformed locally or the entire nano-object can be compressed. Knowledge of the mechanical behavior during loading and unloading in these environments is essential for sustained use. Studies in friction, wear, and deformation behavior are critical to successful integration of nano-objects for the various applications.

38.2.3 BioMEMS

Biosensors are devices typically used to detect target biomolecules such as proteins, enzymes, or nucleic acids, as well as chemical contaminants. Biosensors can be classified into two broad categories: microarray type, which typically consist of (1) microfluidic and nanofluidic sensors that usually involve manipulation of small fluidic volumes (microliters to nanoliters), leading to optical methods for detection, and (2) cantilever or field-effect devices with adsorption of target analytes to sensing elements as the main transduction mechanism.

Advances in micro- and nanofabrication have led to the development of integrated micro- and nanofluidic biosensors that bring together several operations, e.g., sample extraction, purification, sensing, etc., on a simple platform. These sensors are designed either to detect a single biochemical or a class of (bio)chemicals, or for system-level analytical capabilities for a broad range of (bio)chemical species, known as a micro-total analysis system (μTAS). Detection is generally performed using optical methods in micro- and nanofluidic biosensors and mechanical or electrical methods in microarray-type biosensors [38.73]. Optical methods include fluorescence, optical cavity resonator, and surface plasmon resonance. Various methods are used for driving the fluid flow in micro- and nanochannels. Mechanical methods include mechanical changes induced by adsorption of target molecules or analytes to cantilevers or nanowires. Electrical methods include changes to inherent biomolecular charge in the case of field-effect transistors or electrochemical changes. In micro- and nanofluidic biosensors, two methods are commonly used to drive the fluid flow – pressure-driven and electric field-driven or electrokinetic.

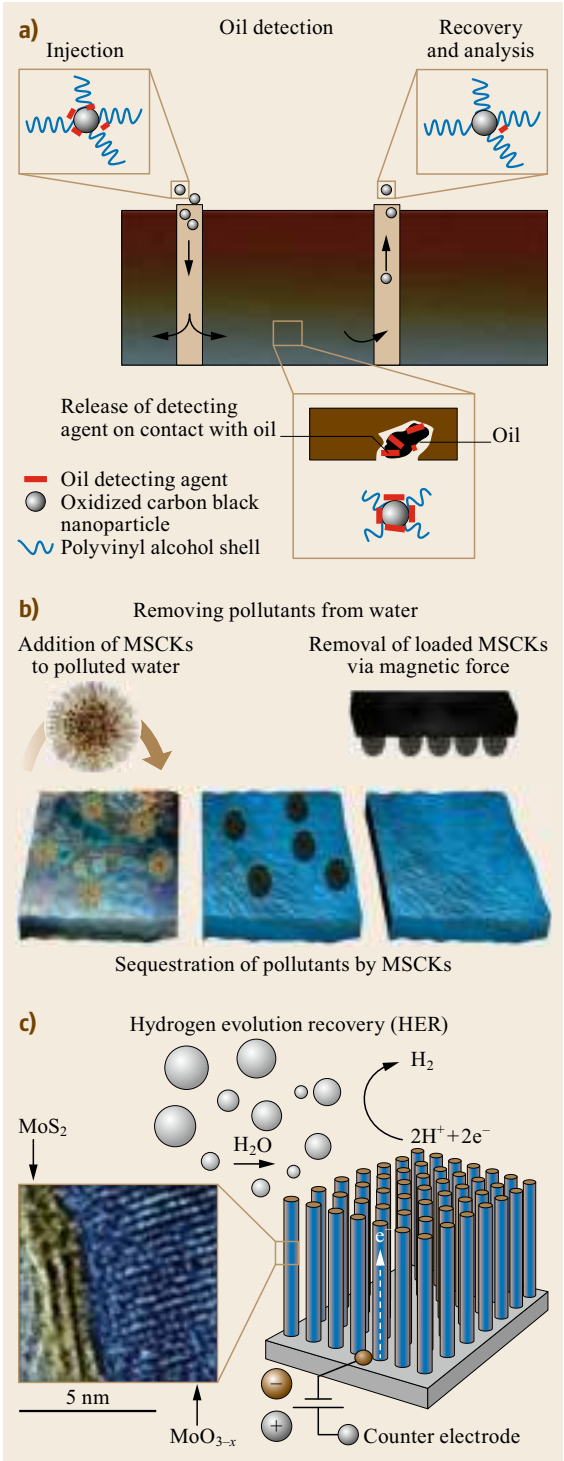


Fig. 38.9a–c Examples of nano-objects in applications in: **(a)** Oil detection, where carbon black nano-objects are coated with an oil detecting agent. After injection into the ground, the agent is released on contact with hydrocarbons, and this is used as an indication of the presence of oil on recovery and analysis of the nanoparticle. Reprinted from [38.131] with permission of The Royal Society of Chemistry. **(b)** Contaminant removal with magnetic shell cross-linked knedel-like nanoparticles (MSCKs). The MSCKs are added to the polluted water (*left*), where they selectively absorb the oil (*center*) and are finally removed by introduction of a neodymium magnet. Reprinted with permission from [38.133]. Copyright (2013) American Chemical Society. **(c)** Catalysis for hydrogen evolution recovery (HER), using nanowires. The nanowires consist of an inner MoO_3 core and outer MoS_2 shell, which is catalytically active. At these sites, the transfer of electrons necessary for recovery of hydrogen from water is enhanced by the MoS_2, which improves the efficiency of the HER. Reprinted with permission from [38.134]. Copyright American Chemical Society. In these applications, low friction of nano-objects in liquids is desirable ◄

An example of a wristwatch-type biosensor based on microfluidics referred to as lab-on-a-chip system is shown in Fig. 38.10a [38.56, 71]. The chip relies on microfluidics and involves manipulation of tiny amounts of fluids in microchannels using microvalves. The test fluid is injected into the chip generally using an external pump or syringe for analysis. Some chips have been designed with an integrated electrostatically actuated diaphragm-type micropump. The sample, which can have volume measured in nanoliters, flows through microfluidic channels via an electric potential and capillary action using microvalves (having various designs including membrane type) for various analyses. The fluid is preprocessed and then analyzed using a biosensor. Another example of a biosensor is the cassette-type biosensor used for human genomic DNA analysis, and integrated biological sample preparation is shown in Fig. 38.10b [38.72]. The implementation of micropumps and microvalves allows for fluid manipulation and multiple sample processing steps in a single cassette. Blood or other aqueous solutions can be pumped into the system where various processes are performed.

Microvalves, which are found in most microfluidic components of biosensors, can be classified in two categories: active microvalves (with an actuator) for flow regulation in microchannels, and passive microvalves integrated with micropumps. Active microvalves consist of a valve seat and a diaphragm actuated by an external actuator [38.55, 135, 136]. Different types of actuators are based on piezoelectric, electrostatic, thermopneumatic, electromagnetic, and bimetallic materials, shaped memory alloys, and solenoid plungers. An example of an electrostatic cantilever-type active microvalve is shown in Fig. 38.11 [38.135]. Passive

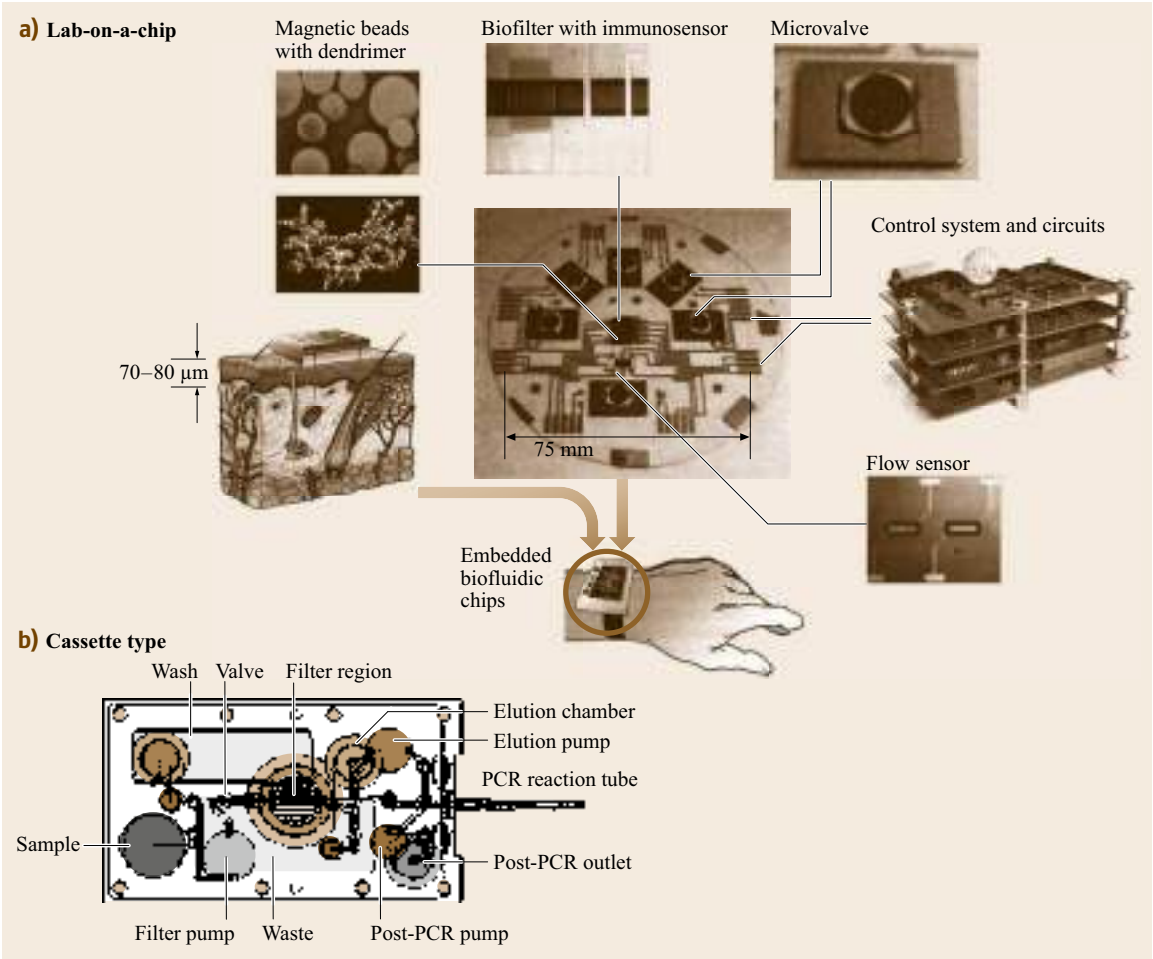


Fig. 38.10 **(a)** MEMS-based biofluidic chip, commonly known as a lab-on-a-chip, that can be worn like a wristwatch (after [38.71]). **(b)** Cassette-type biosensor used for human genomic DNA analysis (after [38.72]) (PCR: polymerase chain reaction)

microvalves used in micropumps include mechanical check valves and a diffuser/nozzle [38.55, 136–139]. Check valves consist of a flap or membrane that is capable of opening and closing with changes in pressure; see Fig. 38.11 for schematics. A diffuser/nozzle uses an entirely different principle and only works with the presence of a reciprocating diaphragm. When one convergent channel works simultaneously with another convergent channel oriented in a specific direction, a change in pressure is possible.

There are four main types of mechanical micropumps: a diaphragm micropump involving mechanical check valves; electrostatic micropumps; valveless rectification pumps with or without diffuser/nozzle type valves; and rotary micropumps [38.55, 136–139]. A diaphragm micropump consists of a reciprocating diaphragm that can be piezoelectrically driven, working in synchronization with two check valves (Fig. 38.11) [38.138]. Electrostatic micropumps have a diaphragm as well, but are driven using two electrodes (Fig. 38.11) [38.140]. Valveless micropumps also consist of a diaphragm that is piezoelectrically driven, but does not incorporate passive mechanical valves. Instead, these pumps use an elastic buffer mechanism or a variable gap mechanism. Finally, a rotary micropump has a rotating rotor that simply adds momentum to the fluid by the fast moving action of the blades (Fig. 38.11) [38.141, 142]. Rotary micropumps can be driven using an integrated electromagnetic motor or by the presence of an external electric field. All of these micropumps can be made of silicon or a polymer material.

During the operation of the microvalves and micropumps discussed above, in which contacts occur due

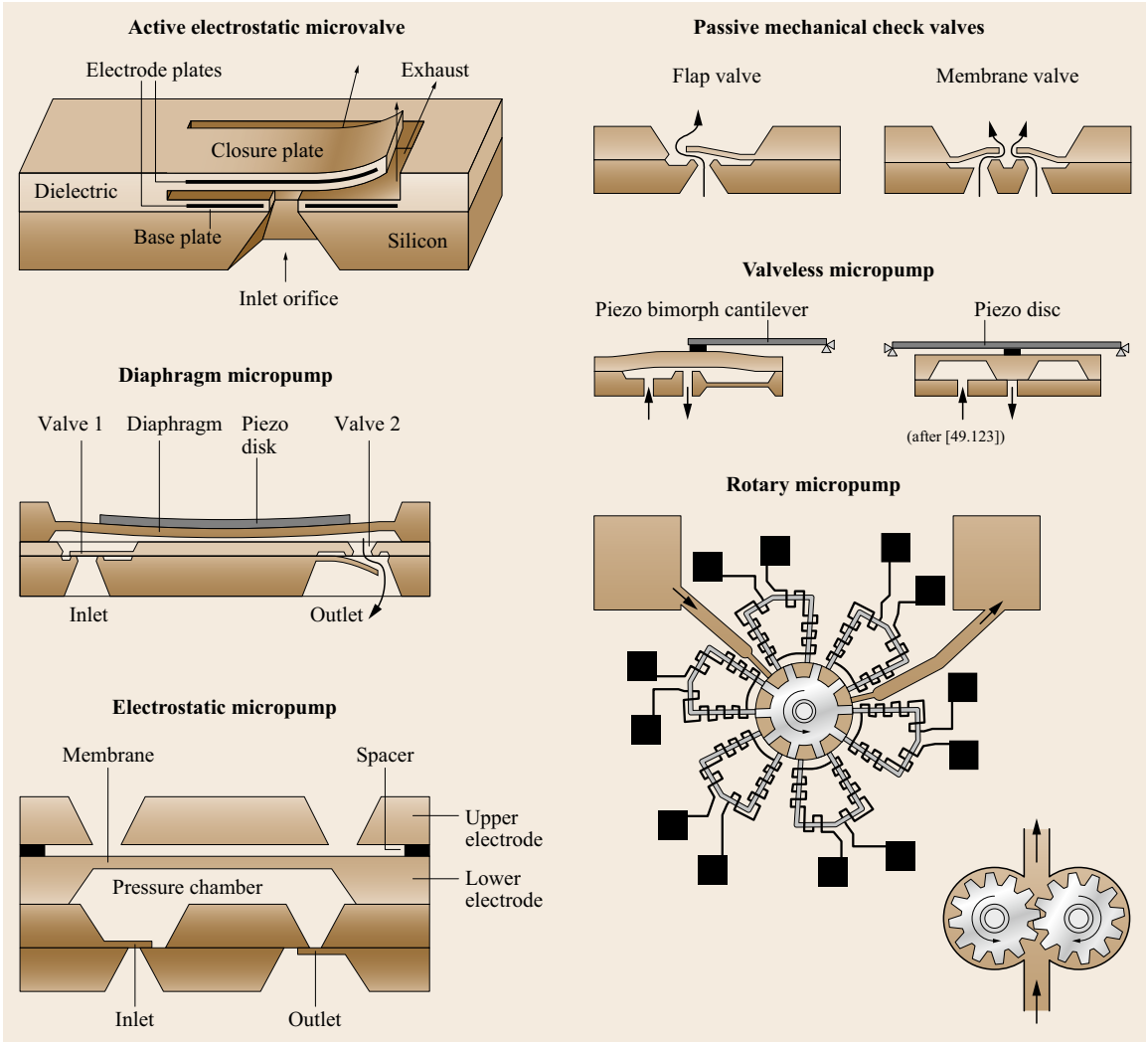


Fig. 38.11 Multiple examples of valves and pumps found in microfluidic devices. Mechanical check valves, diaphragm micropump, valveless micropump, and rotary micropump

to relative motion, adhesion and friction properties become important. During operation, active mechanical microvalves have an externally actuated diaphragm that comes into contact with a valve seat to restrict the fluid flow. Adhesion between the diaphragm and valve seat will affect the operation of the microvalve. In the diaphragm micropumps, two passive mechanical check valves are incorporated into the design. Passive mechanical check valves also exhibit adhesion when the flap or membrane comes into contact with the valve seat when fluid flow is removed. Adhesion also occurs during the operation of valveless micropumps when the diaphragm, which is piezoelectrically driven, comes into contact with the rigid outlet. Finally, adhesion and friction can also be seen during the operation of rotary micropumps when the gears rotate and come into contact and rub against one another.

If the adhesion between the microchannel surface and the biofluid is high, the biomolecules will stick to the microchannel surface and restrict flow. In order to facilitate flow, microchannel surfaces with low bioadhesion are required. Fluid flow in polymer channels can produce triboelectric surface potential, which may affect the flow. Polymers are known to generate surface potential, and the magnitude of the potential varies from one polymer to another [38.143–145]. Conductive surface layers on the polymer channels can be deposited to reduce triboelectric effects.

As mentioned earlier, the microfluidic biosensor shown in Fig. 38.10a required the use of micropumps

and microvalves. As an example, a microdevice with 1000 channels requires 1000 micropumps and 2000 microvalves, which makes it bulky and poses reliability concerns. To overcome this problem, two methods can be used for driving the flow of fluids in microchannels: pressure-driven and electrokinetic. Electrokinetic flow is based on the movement of molecules in an electric field due to their charges. There are two components to electrokinetic flow: electrophoresis, which results from the accelerating force due to the charge of a molecule in an electric field, and electroosmosis, which uses electrically controlled surface tension to drive the uniform liquid flow. Biosensors based on electrokinetic flow have also been developed. In so-called *digital-based microfluidics*, based on the electroosmosis process, electrically controlled surface tension is used to drive liquid droplets, thus eliminating the need for valves and pumps [38.146, 147]. These microdevices consist of a rectangular grid of gold nanoelectrodes instead of micro/nanochannels. An externally applied electric field enables manipulation of few-nanoliter samples through the capillary circuitry.

An example of a microarray-type biosensor, based on a field-effect transistor (FET), is shown in Fig. 38.12a [38.75, 78–80, 148]. FETs are sensitive to the electrical field produced by the charge at the surface of the gate insulator. In this sensor, the gate metal electrode of a metal–oxide–semiconductor field-effect transistor (MOSFET) is removed and replaced with a protein (receptor layer) whose cognate is the analyte (e.g., virus or bacteria) that is meant to be sensed. Various proteins may have 1–25 (positive or negative) charges per molecule. The binding of the receptor layer with the analyte produces a change in the effective charge, creating a change in the electrical field. This electrical field change may produce a measurable change in the current flow through the device. Adhesion between the protein layer and silica substrate affects the reliability of the biosensor. In the case of implanted biosensors, the biosensors come in contact with the exterior environment, such as tissues and fluids, and any relative motion of the sensor surface with respect to that exterior environment may result in surface damage. A schematic of the generation of friction and wear points when an implanted biosensor surface comes in contact with a living tissue is shown in Fig. 38.12b [38.93]. The friction, wear, and adhesion of the biosensor surface may be critical in these applications [38.75, 93, 149, 150].

Polymer bioMEMS have been designed to measure cellular surfaces. For two examples, see Fig. 38.13 [38.151]. The device on the left has cantilevers anchored at the periphery of the circular structure, while the device on the right has cantilevers

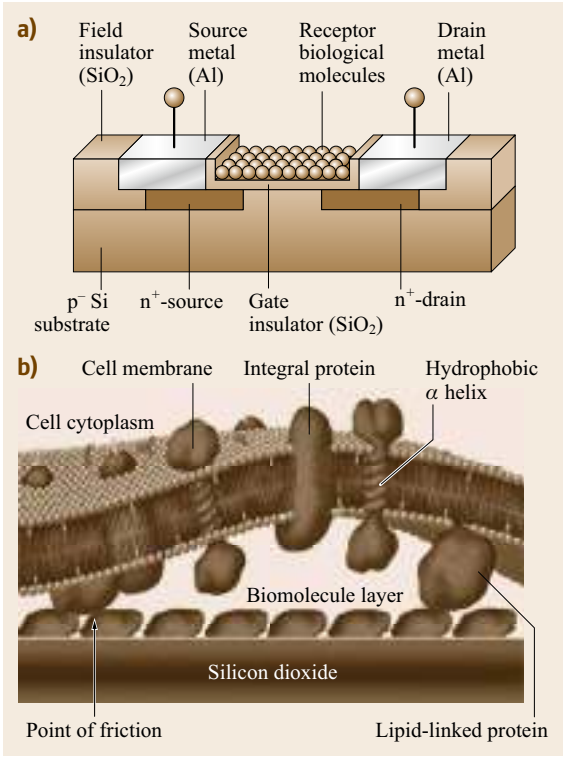


Fig. 38.12 **(a)** Schematic of a bioFET sensor (after [38.75]), and **(b)** schematic showing the generation of friction and wear points due to interaction of an implanted biomolecule layer on a biosensor with living tissue (after [38.93])

anchored at the two corners on the top and the bottom. The cell adheres to the center of the structure, and the contractile forces generated in the cell's cytoskeleton causes the cantilever to deflect. The deflection of the compliant polymer cantilevers is measured optically and related to the magnitude of the forces generated by the cell. Adhesion between cells and the polymer beam is desirable. In order to design the sensors, micro- and nanoscale mechanical properties of polymer structures are needed.

38.2.4 BioNEMS

Micro/nanofluidic devices provide a powerful platform for electrophoretic separations for a variety of chemical and biochemical analyses. Electrophoresis is a versatile analytical method that is used for separation of small ions, neutral molecules, and large biomolecules. Figure 38.14 shows an interdigitated micro/nanofluidic silicon array with nanochannels for separation processes. Figure 38.15 shows a schematic of an implantable, immunoisolation submicroscopic biocapsule

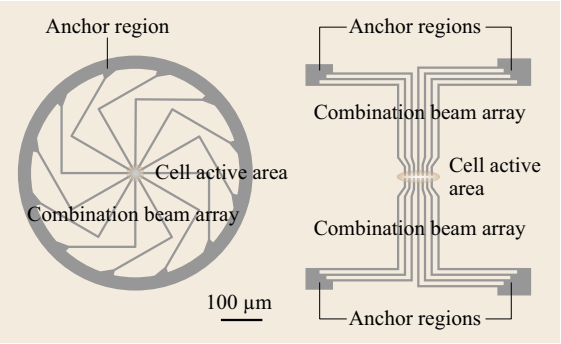


Fig. 38.13 Schematic of two designs for polymer bioMEMS structures to measure cellular forces (after [38.151])

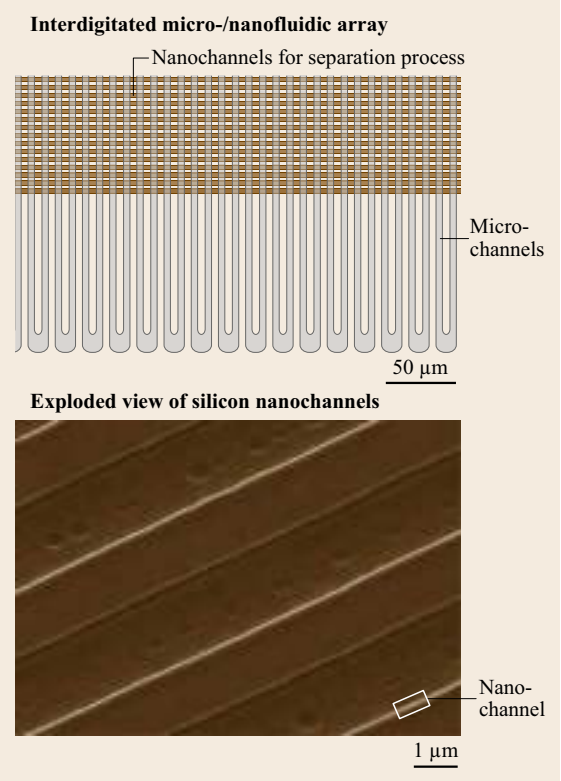


Fig. 38.14 Interdigitated micro/nanofluidic silicon array for separation processes (after [38.83])

for drug delivery treating significant medical conditions such as Type I diabetes [38.82, 83]. The purpose of the immunoisolation biocapsule is to create an implantable device capable of supporting foreign living cells that can be transplanted into humans. It is a silicon capsule consisting of two nanofabricated membranes bonded together with the drug (e.g., encapsulated insulin-producing islet cells) contained within the

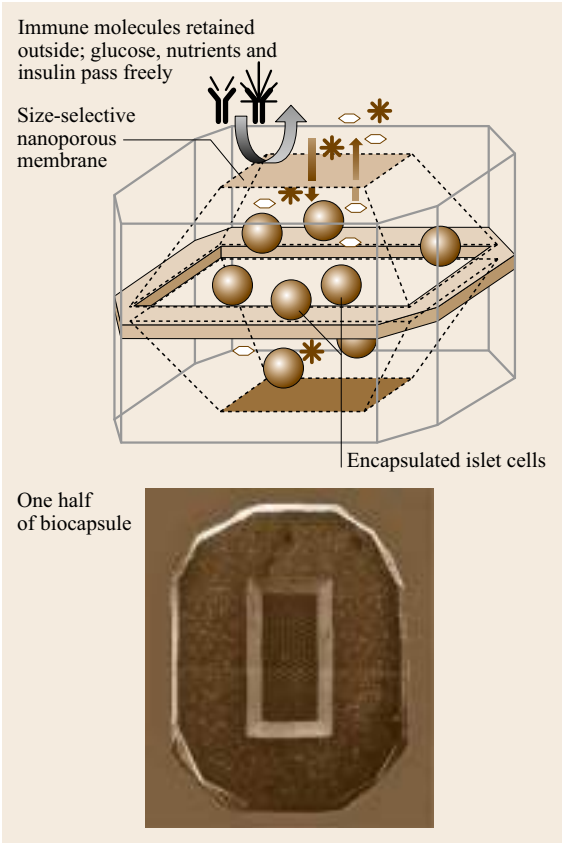


Fig. 38.15 Schematics of implantable, immunoisolation submicroscopic biocapsules (drug-delivery devices) (after [38.83])

cavities for long-term delivery. The pores or nanochannels in a semipermeable membrane as small as 6 nm are used as flux regulators for the long-term release of drugs. The nanomembrane also protects therapeutic substances from attack by the body's immune system. The pores are large enough to provide the flow of nutrients (e.g., glucose molecules) and drugs (e.g., insulin). However, they are small enough to block natural antibodies, which can penetrate any orifice larger than 18 nm. Pores in silicon can be etched using sacrificial layer lithography, described in Appendix 38.A [38.83].

The main reliability concerns in micro/nanofluidic silicon arrays and implantable biocapsules are biocompatibility and potential biofouling (undesirable accumulation of microorganisms) of the channels/membrane by means of protein and cell adsorption from biological fluids. Biofouling can also result in the clogging of the nanochannels/nanopores, which potentially could render the device ineffective. The adhesion of proteins and cells to an implanted device can also cause detrimental results such as inflammation and excessive

fibrosis. Deposition of the self-assembled monolayers of selected organic molecules on the channels implants, which makes them hydrophobic, presents an innovative solution to combat the adverse effects of the biological fluids [38.152–155].

Figure 38.16 shows a conceptual model of a targeted intravascular drug-delivery device using nanoparticles to search and destroy disease (tumor) cells [38.84]. The tumor cells have a 1–2 orders of magnitude higher density of receptors than the normal cells and a lower pH. Some receptors are only expressed on tumor cells. Nanoparticles made of various materials including gold (Au), graphene, iron oxide, polymer, and silica have been studied in targeted drug delivery for cancer treatment [38.156–162]. With lateral dimensions of 1 μm or less, the particles are smaller than any blood cells. These particles can be injected into the bloodstream and travel freely through the circulatory system. In order to direct these drug-delivery nanoparticles to cancer sites, their external surfaces are chemically modified to carry molecules that have lock-and-key binding specificity with molecules that support a growing cancer mass. As the particles come in close proximity with the diseased cells, the ligands on the particle surfaces attach to the receptors on the cells. As soon as the particles dock on the cells, a compound is released that forms a pore on the membrane of the cells, which leads to cell death and, ultimately, death of the cancer mass that was being nourished by the blood vessel. The adhesive interactions are regulated by specific (ligand-receptor binding) and nonspecific (short-range van der Waals, electrostatic, steric) interactions [38.163]. Adhesion between nanoparticles and disease cells is required. Furthermore, the particles should travel close to the endothelium lining of vascular arteries to facilitate the interaction between the particles and diseased cells. Human capillaries can have radii as small as 4–5 μm. *Decuzzi* et al. [38.163] analyzed the margination of a particle circulating in the bloodstream and calculated the speed and time for margination (drifting of particles towards the blood vessel walls) as a function of density and diameter of particle, based on various forces present between the circulating particle and the endothelium lining. They reported that the particles used for drug delivery should have a radius smaller than the critical value (in the range of 100 nm). Studies show that a lateral force on the particles assists them in faster margination towards the endothelium walls. Thus, nonspherical particles are more desirable. In addition, reducing agglomeration and friction of nano-objects moving through a liquid and in contact with surfaces is important for efficient transport to diseased sites.

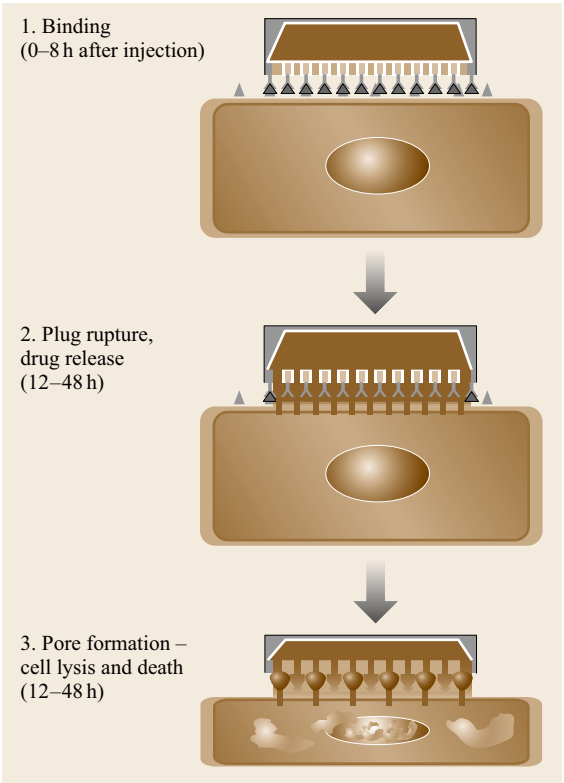


Fig. 38.16 Intravascular nanoparticles to search and destroy diseased blood cells. Life cycles include (1) target binding, (2) drug release, and (3) cell death (after [38.84])

Because of their unique mechanical and electrical properties, single-walled and multiwalled carbon nanotubes (SWNT and MWNT) are being used for thermal management of high-power devices, reinforced composites and super-strong fiber and sheets, chemical and biological sensors, electromechanical devices, field emission devices, molecular electronics and computing [38.49, 164]. Figure 38.17 shows a SWNT biosensor [38.2]. The conductance of carbon nanotube (CNT) devices changes when proteins adsorb on the surface. The change in electrical resistance is a measure of protein adsorption. For high performance, adhesion should be strong between the adsorbent and SWNT.

In summary, adhesion, stiction/friction, and wear clearly limit the lifetimes and compromise the performance and reliability of MEMS/NEMS and bioMEMS/bioNEMS. Figure 38.18 summarizes tribological problems encountered in some of the MEMS, MOEMS, RF-MEMS, and bioMEMS devices just discussed.

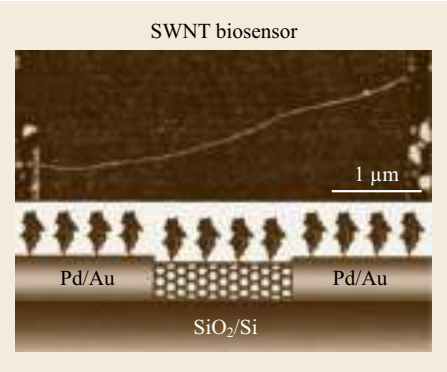


Fig. 38.17 SEM micrograph of SWNT biosensor; *bottom schematic* shows the adsorption of protein molecules to the SWNT (after [38.2])

38.2.5 Microfabrication Processes

In addition to in-use stiction, stiction issues are also present in some processes used for the fabrication of MEMS/NEMS. For example, the last step in surface micromachining involves the removal of sacrificial layer(s) called release, since the microstructures are released from the surrounding sacrificial layer(s). The release is accomplished by an aqueous chemical etching, rinsing, and drying processes. Due to meniscus effects as a result of wet processes, the suspended structures can sometimes collapse and permanently adhere to the underlying substrate, as shown in Fig. 38.19 [38.165]. Adhesion is caused by water molecules adsorbed on the adhering surfaces and/or because of formation of adhesive bonds by silica residues that remain on the surfaces after the water has evaporated. This so-called release stiction is overcome by using dry release methods, such as CO_2 critical point drying or sublimation drying [38.166]. CO_2 at high pressure is in a supercritical state and becomes liquid. Liquid CO_2 is used to remove wet etchant, and then it is converted back to the gas phase.

38.2.6 Tribological Needs

Various MEMS/NEMS are designed to perform expected functions in the millisecond to picosecond range. The life expectancy of devices for high-speed contacts can vary from a few hundred thousand to many billions of cycles, e.g., over a hundred billion cycles for DMDs, which puts stringent requirements on materials [38.13, 103, 167–170]. Adhesion between a biological molecular layer and the substrate (referred to as *bioadhesion*), reduction of friction and wear of biological layers, biocompatibility, and biofouling for

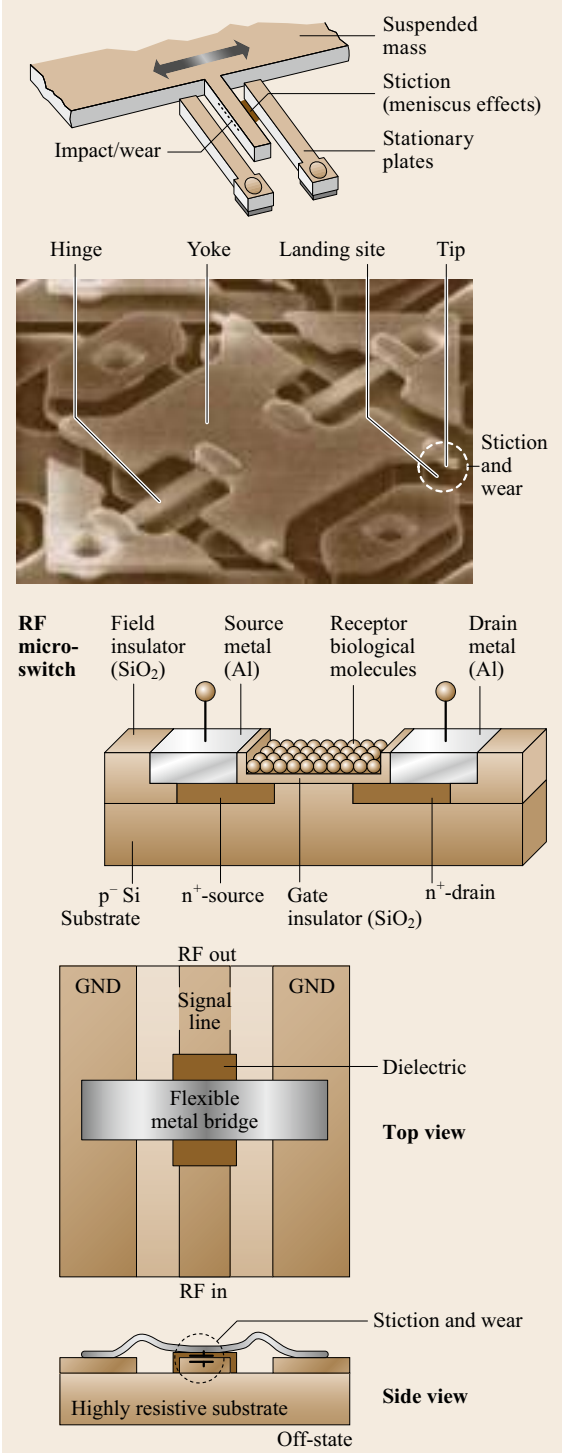


Fig. 38.18 Summary of tribological issues in MEMS, MOEMS, RF-MEMS and bioMEMS device operation (after [38.75, 152])

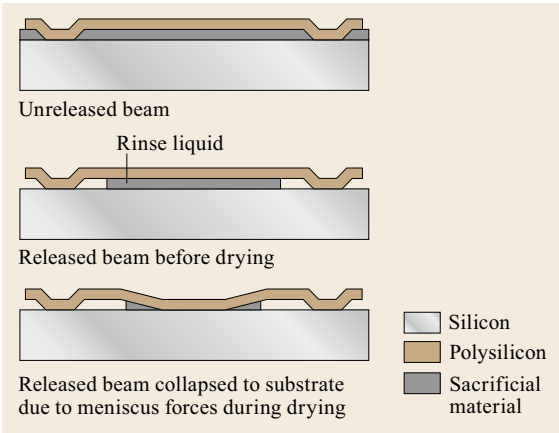


Fig. 38.19 Tribological issues in microfabrication via surface micromachining

bioMEMS/bioNEMS are important [38.75, 93]. Most mechanical properties are known to be scale dependent [38.85]. Therefore, the properties of nanoscale structures need to be measured [38.171]. There is a need to develop a fundamental understanding of adhesion, friction/stiction, wear, and the role of surface contamination, and environment [38.13]. MEMS/NEMS materials need to possess good mechanical and tribological properties on the micro/nanoscale. There is a need to develop lubricants and identify lubrication methods that are suitable for MEMS/NEMS devices and applications. Methods need to be developed to enhance adhesion between biomolecules and the device substrate (bioadhesion). Component-level studies are required to provide a better understanding of the tribological phenomena occurring in MEMS/NEMS.

The fields of nanotribology and nanomechanics, and atomic force microscopy-based techniques, have provided researchers with a viable approach to address these problems [38.44, 45, 86–88, 172, 173]. The following is an overview of nanoscale adhesion, friction and wear studies of materials and lubrication for MEMS/NEMS and bioMEMS/bioNEMS, and component-level studies of stiction phenomena, in MEMS/NEMS devices.

38.3 Nanotribology and Nanomechanics Studies of Silicon and Related Materials

The materials of most interest for planar fabrication processes using silicon as the structural material are undoped and boron-doped (p^+-type) single-crystal silicon for bulk micromachining, and doped and undoped phosphorus (n^+-type) low-pressure chemical vapor deposition (LPCVD) polysilicon films for surface micromachining. Since silicon-based devices lack high-temperature capabilities with respect to both mechanical and electrical properties, SiC is being developed as a structural material for high-temperature microsensor and microactuator applications [38.174–177]. SiC can also be desirable for high-frequency micromechanical resonators, in the GHz range, because of its high modulus of elasticity-to-density ratio and, consequently, high resonance frequency. Table 38.2 compares selected bulk properties of SiC and Si(100). Researchers have found low-cost techniques for producing single-crystalline 3C-SiC (cubic or β-SiC) films via epitaxial growth on large-area silicon substrates for bulk micromachining [38.178] and polycrystalline 3C-SiC films on polysilicon and silicon dioxide layers for surface micromachining of SiC [38.179]. Single-crystalline 3C-SiC piezoresistive pressure sensors have been fabricated using bulk micromachining [38.176, 177]. Surface-micromachined polycrystalline SiC micromotors have been fabricated and reported to provide satisfactory operation at high temperatures [38.180]. Because of their importance to MEMS/NEMS applications, the tribology of these materials must be examined and optimized.

In this section, both macroscale and microscale tribological properties of virgin and treated/coated silicon, polysilicon films, and SiC are presented. As will be shown, bare silicon exhibits inadequate tribological performance and needs to be coated with a solid and/or liquid overcoat or be surface treated (e.g., oxidation and ion implantation, commonly used in semiconductor manufacturing), which reduces friction and wear. SiC films exhibit good tribological performance.

38.3.1 Virgin and Treated/Coated Silicon Samples

Nanotribological and Nanomechanical Properties of Silicon and Effect of Ion Implantation

Friction and wear of single-crystalline and polycrystalline silicon samples have been studied, and the effect of ion implantation with various doses of C^+, B^+, N_2^+, and Ar^+ ion species at 200 keV energy to improve their friction and wear properties has been studied [38.183–185]. The coefficient of macroscale friction and the

Table 38.2 Selected bulk properties[a] of 3C (β- or cubic) SiC and Si(100)

Sample	Density (kg/m^3)	Hardness (GPa)	Elastic modulus (GPa)	Fracture toughness ($MPa\,m^{1/2}$)	Thermal conductivity[b] ($W/(m\,K)$)	Coefficient of thermal expansion[b] ($\times 10^{-6}\,K^{-1}$)	Melting point (°C)	Bandgap (eV)
β-SiC	3210	23.5–26.5	440	4.6	85–260	4.5–6	2830	2.3
Si(100)	2330	9–10	130	0.95	155	2–4.5	1410	1.1

[a] Unless otherwise stated, data shown were obtained from [38.181]
[b] Obtained from [38.182]

wear factor of virgin single-crystal silicon and C^+-implanted silicon samples as a function of ion dose are presented in Fig. 38.20 [38.183]. The macroscale friction and wear tests were conducted using a ball-on-flat tribometer. Each data bar represents the average value of four to six measurements. The coefficient of friction and wear factor for bare silicon are very high and decrease drastically with ion dose. Silicon samples bombarded above the ion dose of $10^{17}\,C^+\,cm^{-2}$ exhibit extremely low values of coefficients of friction (typically 0.03–0.06 in air) and wear factors (reduced by as much as four orders of magnitude). *Gupta* et al. [38.183] reported that a decrease in the coefficient of friction and wear factor of silicon as a result of C^+ ion bombardment occurred because of formation of silicon carbide rather than amorphization of silicon. *Gupta* et al. [38.184] also reported an improvement in friction and wear with B^+ ion implantation.

Microscale friction measurements were performed using an atomic force/friction force microscope (AFM/FFM) [38.44, 45, 86, 88, 173]. Table 38.3 shows values of surface roughness and coefficients of macroscale and microscale friction for virgin and doped silicon. There is a decrease in the coefficients of microscale and macroscale friction values as a result of ion implantation. When measured for small contact areas and very low loads used in microscale studies, indentation hardness and elastic modulus are higher than those at the macroscale. This, added to the effect of the small apparent area of contact reducing the number of trapped particles on the interface, results in less plowing contribution and lower friction in the case of microscale friction measurements. Figures 38.21a and 38.21b show results of microscale wear resistance studies of ion-implanted silicon samples studied using a diamond tip in an AFM [38.186]. For tests conducted at various loads on Si(111) and C^+-implanted Si(111), it is noted that the wear resistance of the implanted sample is slightly poorer than that of virgin silicon up to about 80 μN. Above 80 μN, the wear resistance of implanted Si improves. As one continues to run tests at 40 μN for a larger number of cycles, the implanted sample, which forms hard and tough silicon carbide, exhibits higher wear resistance than the unimplanted sample. Damage from the implantation in the top layer results in poorer wear resistance; however, the implanted zone at the subsurface is more wear resistant than the virgin silicon.

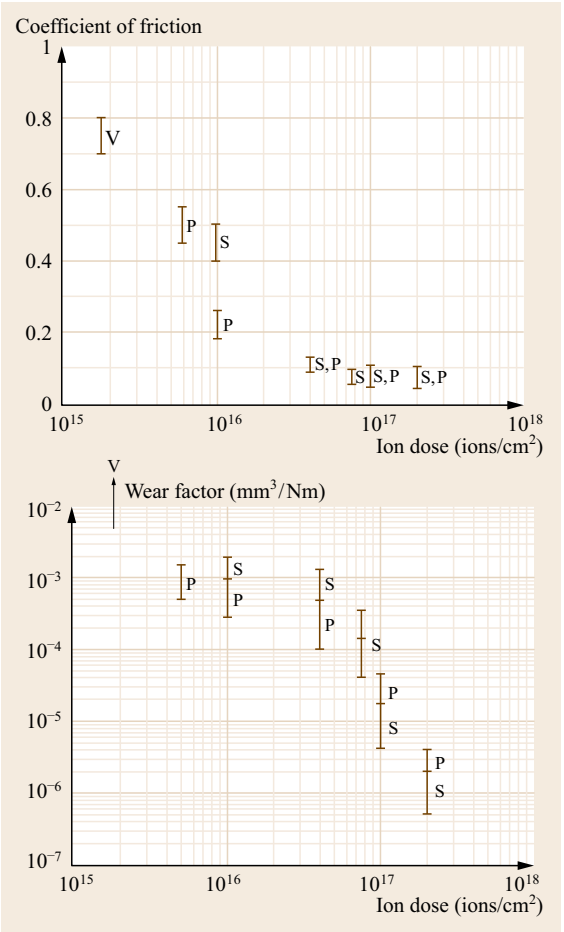


Fig. 38.20 Influence of ion doses on the coefficient of friction and wear factor on C^+ ion bombarded single-crystal and polycrystalline silicon slid against an alumina ball. V corresponds to virgin single-crystal silicon, while S and P denote tests that correspond to doped single- and polycrystalline silicon respectively (after [38.183])

Table 38.3 Surface roughness and micro- and macroscale coefficients of friction of selected samples

Material	RMS roughness (nm)	Coefficient of microscale friction[a]	Coefficient of macroscale friction[b]
Si(111)	0.11	0.03	0.33
C^+-implanted Si(111)	0.33	0.02	0.18

[a] Versus Si_3N_4 tip, tip radius of 50 nm in the load range of 10–150 nN (2.5–6.1 GPa) at a scanning speed of 5 μm/s over a scan area of 1 μm × 1 μm in an AFM
[b] Versus Si_3N_4 ball, ball radius of 3 mm at a normal load of 0.1 N (0.3 GPa) at an average sliding speed of 0.8 mm/s using a tribometer

Hardness values of virgin and C^+-implanted Si(111) at various indentation depths (normal loads) are presented in Fig. 38.21c [38.186]. The hardness at a small indentation depth of 2.5 nm is 16.6 GPa, and it drops to a value of 11.7 GPa at a depth of 7 nm and a normal load of 100 μN. Higher hardness values obtained in low-load indentation may arise from the observed pressure-induced phase transformation during the nanoindentation [38.187, 188]. Additional increases in the hardness at an even lower indentation depth of 2.5 nm reported here may arise from the contribution of complex chemical films (not from native oxide films) present on the silicon surface. At small volumes there is a lower probability of encountering material defects (dislocations, etc.). Furthermore, according to the strain gradient plasticity theory advanced by *Fleck* et al. [38.189], large strain gradients inherent in small indentations lead to accumulation of geometrically necessary dislocations that cause enhanced hardening. These are some of the plausible explanations for an increase in hardness at smaller volumes. If the silicon material were to be used at very light loads, such as in microsystems, the high hardness of surface films would protect the surface until it is worn.

From Fig. 38.21c, the hardness value of C^+-implanted Si(111) at a normal load of 50 μN is 20.0 GPa with an indentation depth of about 2 nm, which is comparable to the hardness value of 19.5 GPa at 70 μN; whereas the measured hardness value for virgin silicon at an indentation depth of about 7 nm (normal load of 100 μN) is only about 11.7 GPa. Thus, ion implantation with C^+ results in an increase in hardness in silicon. Note that the surface layer of the implanted zone is much harder compared with the subsurface, and may be brittle, leading to higher wear on the surface. The subsurface of the implanted zone (SiC) is harder than the virgin silicon, resulting in higher wear resistance, which is also observed in the results of the macroscale tests conducted at high loads.

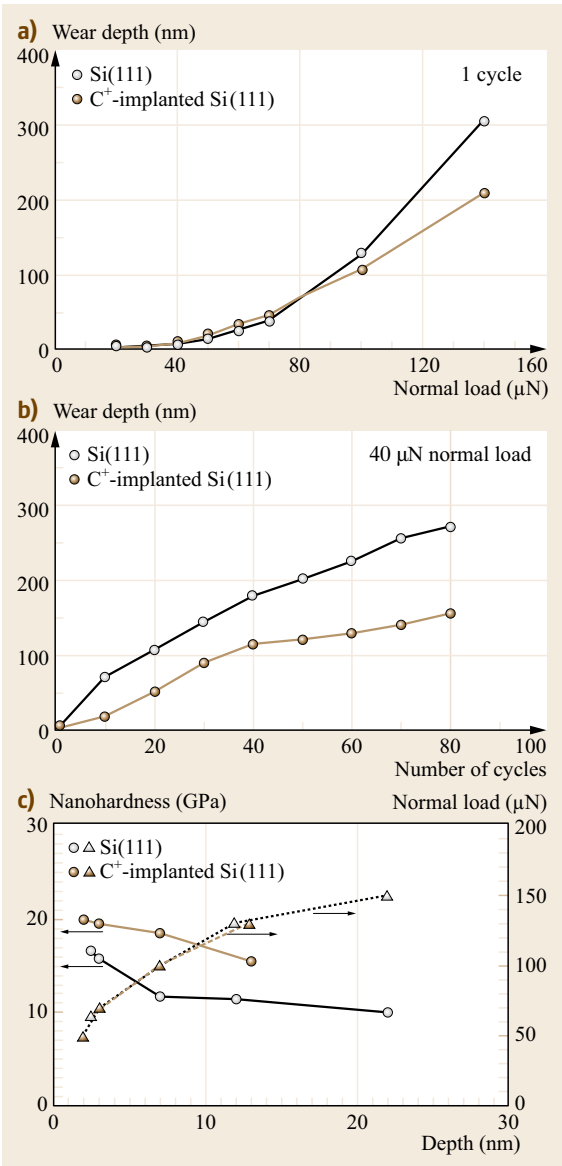


Fig. 38.21a–c Wear depth as a function of **(a)** load (after one cycle), and **(b)** cycles (normal load = 40 mN) for Si(111) and C^+-implanted Si(111). **(c)** Nanohardness and normal load as function of indentation depth for virgin and C^+-implanted Si(111) (after [38.186]) ◀

Effect of Oxide Films on Nanotribological Properties of Silicon

Macroscale friction and wear experiments have been performed using a magnetic disk drive with bare, oxidized, and implanted pins sliding against amorphous-carbon-coated magnetic disks lubricated with a thin layer of perfluoropolyether lubricant [38.190–193]. Representative profiles of the variation of the coefficient of friction with a number of sliding cycles for an Al_2O_3-TiC slider and bare and dry-oxidized silicon pins are shown in Fig. 38.22. For bare Si(111), after an initial increase in the coefficient of friction, it drops to a steady state value of 0.1, as seen in Fig. 38.22. The rise in the coefficient of friction for the Si(111) pin is associated with the transfer of amorphous carbon from the disk to the pin, and oxidation-enhanced fracture of pin material followed by tribochemical oxidation of the transfer film, while the drop is associated with the formation of a transfer coating on the pin. Dry-oxidized Si(111) exhibits excellent characteristics and no significant increase was observed over 50 000 cycles (Fig. 38.22). This behavior has been attributed to the chemical passivity of the oxide and lack of transfer of diamond-like carbon (DLC) from the disk to the pin. The behavior of PECVD-oxide (data are not presented here) was comparable to that of dry oxide, but for the wet oxide there was some variation in the coefficient of friction (0.26–0.4). The difference between the dry and wet oxide was attributed to increased porosity of the wet oxide [38.190]. Since tribochemical oxidation was determined to be a significant factor, experiments were conducted in dry nitrogen [38.191, 192]. The variation of the coefficient of friction for a silicon pin sliding against a thin-film disk in dry nitrogen is shown in Fig. 38.22. It is seen that in a dry nitrogen environment, the coefficient of friction of Si(111) sliding against a disk decreased from an initial value of about 0.2 to 0.05 with continued sliding. Based on SEM and chemical analysis, this behavior has been attributed to the formation of a smooth amorphous-carbon/lubricant transfer patch and suppression of oxidation in a dry nitrogen environment. Based on macroscale tests using disk drives, it was found that the friction and wear performance of bare silicon is not adequate. With dry-oxidized or PECVD SiO_2-coated silicon, no significant friction increase or interfacial degradation was observed in ambient air.

Table 38.4 and Fig. 38.23 show surface roughness, microscale friction and scratch data, and nanoindentation hardness for the various silicon samples [38.186]. Scratch experiments were performed using a diamond tip in an AFM. Results on polysilicon samples are also shown for comparison. Coefficients of microscale friction values for all the samples are about the same.

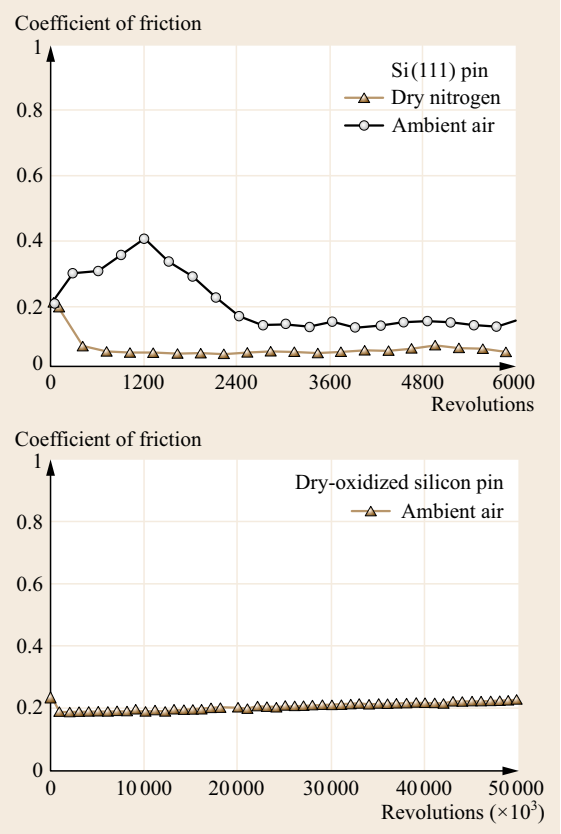


Fig. 38.22 Coefficient of friction as a function of number of sliding revolutions for Si(111) pin in ambient air and dry nitrogen and dry-oxidized silicon pin in ambient air (after [38.190])

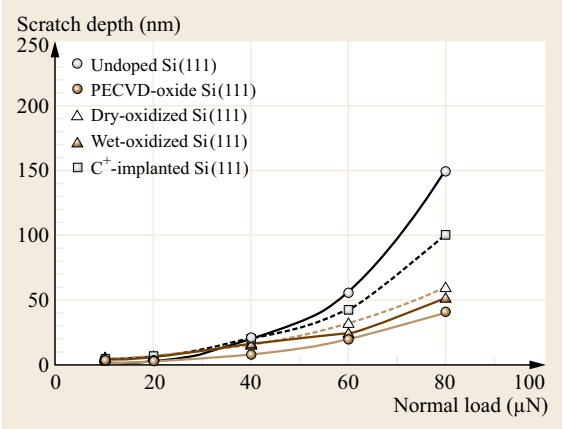


Fig. 38.23 Scratch depth as a function of normal load after 10 cycles for various virgin, treated, and coated silicon samples (after [38.186])

Table 38.4 RMS, microfriction, microscratching/microwear and nanoindentation hardness data for various virgin, coated and treated silicon samples

Material	RMS roughness[a] (nm)	Coefficient of microscale friction[b]	Scratch depth[c] at 40 μN (nm)	Wear depth[c] at 40 μN (nm)	Nanohardness[c] at 100 μN (GPa)
Si(111)	0.11	0.03	20	27	11.7
Si(110)	0.09	0.04	20	–	–
Si(100)	0.12	0.03	25	–	–
Polysilicon	1.07	0.04	18	–	–
Polysilicon (lapped)	0.16	0.05	18	25	12.5
PECVD-oxide-coated Si(111)	1.50	0.01	8	5	18.0
Dry-oxidized Si(111)	0.11	0.04	16	14	17.0
Wet-oxidized Si(111)	0.25	0.04	17	18	14.4
C^+-implanted Si(111)	0.33	0.02	20	23	18.6

[a] Scan size of 500 nm × 500 nm using AFM
[b] Versus Si_3N_4 tip in AFM/FFM, radius 50 nm; at 1 μm × 1 μm scan size
[c] Measured using an AFM with a diamond tip of radius of 100 nm

Table 38.5 Summary of micro/nanotribological properties of the sample materials

Sample	RMS roughness[a] (nm)	P–V distance[a] (nm)	Coefficient of friction, micro[b]	Coefficient of friction, macro[c]	Scratch depth[d] (nm)	Wear depth[e] (nm)	Nano-hardness[f] (GPa)	Young's modulus[f] (GPa)	Fracture toughness[g], K_{IC} (MPa $m^{1/2}$)
Undoped Si(100)	0.09	0.9	0.06	0.33	89	84	12	168	0.75
Undoped polysilicon film (as deposited)	46	340	0.05	–	–	–	–	–	–
Undoped polysilicon film (polished)	0.86	6	0.04	0.46	99	140	12	175	1.11
n^+-Type polysilicon film (as deposited)	12	91	0.07	–	–	–	–	–	–
n^+-Type polysilicon film (polished)	1.0	7	0.02	0.23	61	51	9	95	0.89
SiC film (as deposited)	25	150	0.03	–	–	–	–	–	–
SiC film (polished)	0.89	6	0.02	0.20	6	16	25	395	0.78

[a] Measured using AFM over a scan size of 10 μm × 10 μm
[b] Measured using AFM/FFM over a scan size of 10 μm × 10 μm
[c] Obtained using a 3 mm-diameter sapphire ball in a reciprocating mode at a normal load of 10 mN and average sliding speed of 1 mm/s after 4 m sliding distance
[d] Measured using AFM at a normal load of 40 μN for 10 cycles, scan length of 5 μm
[e] Measured using AFM at normal load of 40 μN for one cycle, wear area of 2 μm × 2 μm
[f] Measured using a nanoindenter at a peak indentation depth of 20 nm
[g] Measured using a microindenter with Vickers indenter at a normal load of 0.5 N

These samples showed wear scars from scratching with a 10 μN load. Scratch depth increased with normal load. Crystalline orientation of silicon has little influence on scratch resistance because natural oxidation of silicon in ambient air masks the expected effect of crystallographic orientation. PECVD-oxide samples showed the best scratch resistance, followed by dry-oxidized, wet-oxidized, and ion-implanted samples. Ion implantation with C^+ does not appear to improve scratch resistance.

Wear data on the silicon samples are also presented in Table 38.4 [38.186]. PECVD-oxide samples showed superior wear resistance followed by dry-oxidized, wet-oxidized, and ion-implanted samples. This agrees with the trends seen in scratch resistance. In PECVD, ion bombardment during the deposition improves the coating properties such as suppression of columnar growth, freedom from pinhole, decrease in crystalline size, and increase in density, hardness, and substrate-coating adhesion. These effects may help in improving the mechanical integrity of the sample surface. Coatings and treatments improved the nanohardness of silicon. Note that dry-oxidized and PECVD films are harder than wet-oxidized films as these films may be porous. High hardness of oxidized films may be responsible for measured high scratch/wear resistance.

38.3.2 Nanotribological and Nanomechanical Properties of Polysilicon Films and SiC Films

Studies have also been conducted on undoped polysilicon film, heavily doped (n^+-type) polysilicon film, heavily doped (p^+-type) single-crystal Si(100), and 3C-SiC (cubic or β-SiC) film [38.194–196]. The polysilicon films studied here are different from the ones discussed previously.

Table 38.5 presents a summary of the tribological studies conducted on polysilicon and SiC films. Values for single-crystal silicon are also shown for comparison. Polishing of the as-deposited polysilicon and SiC films drastically affects the roughness, as the values reduce by two orders of magnitude. Si(100) appears to be the smoothest followed by polished undoped polysilicon and SiC films, which have comparable roughness. The doped polysilicon film shows higher roughness than the undoped sample, which is attributed to the doping process. The polished SiC film shows the lowest friction followed by polished and undoped polysilicon film, which strongly supports the candidacy of SiC films for use in MEMS/NEMS devices. Macroscale friction measurements indicate that SiC film exhibits one of the lowest friction values as compared to the other samples. The doped polysilicon sample shows low friction on the macroscale as compared to the undoped polysilicon sample possibly due to the doping effect.

Figure 38.24a shows a plot of scratch depth versus normal load for various samples [38.194, 195]. Scratch depth increases with increasing normal load. Figure 38.25 shows AFM three-dimensional (3-D) maps and averaged two-dimensional (2-D) profiles of the scratch marks on the various samples. It is observed that scratch depth increases almost linearly with the normal load. Si(100) and the doped and undoped polysilicon film show similar scratch resistance. From the data, it is clear that the SiC film is much more scratch resistant than the other samples. Figure 38.24b shows results from microscale wear tests on the various films. For all the materials, the wear depth increases almost linearly with increasing number of cycles. This suggests that the material is removed layer by layer in all the materials. Here also, SiC film exhibits lower wear depths than the other samples. Doped polysilicon film wears less than the undoped film. Higher fracture toughness and higher hardness of SiC as compared to Si(100) is

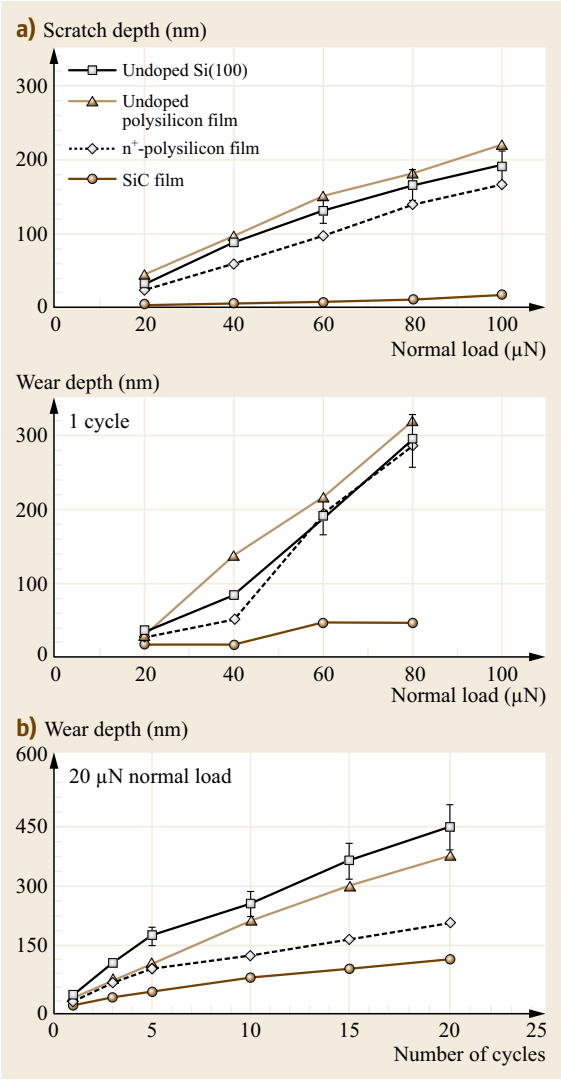


Fig. 38.24 **(a)** Scratch depths for 10 cycles as a function of normal load and **(b)** wear depths as a function of normal load and as a function of number of cycles for various samples (after [38.194])

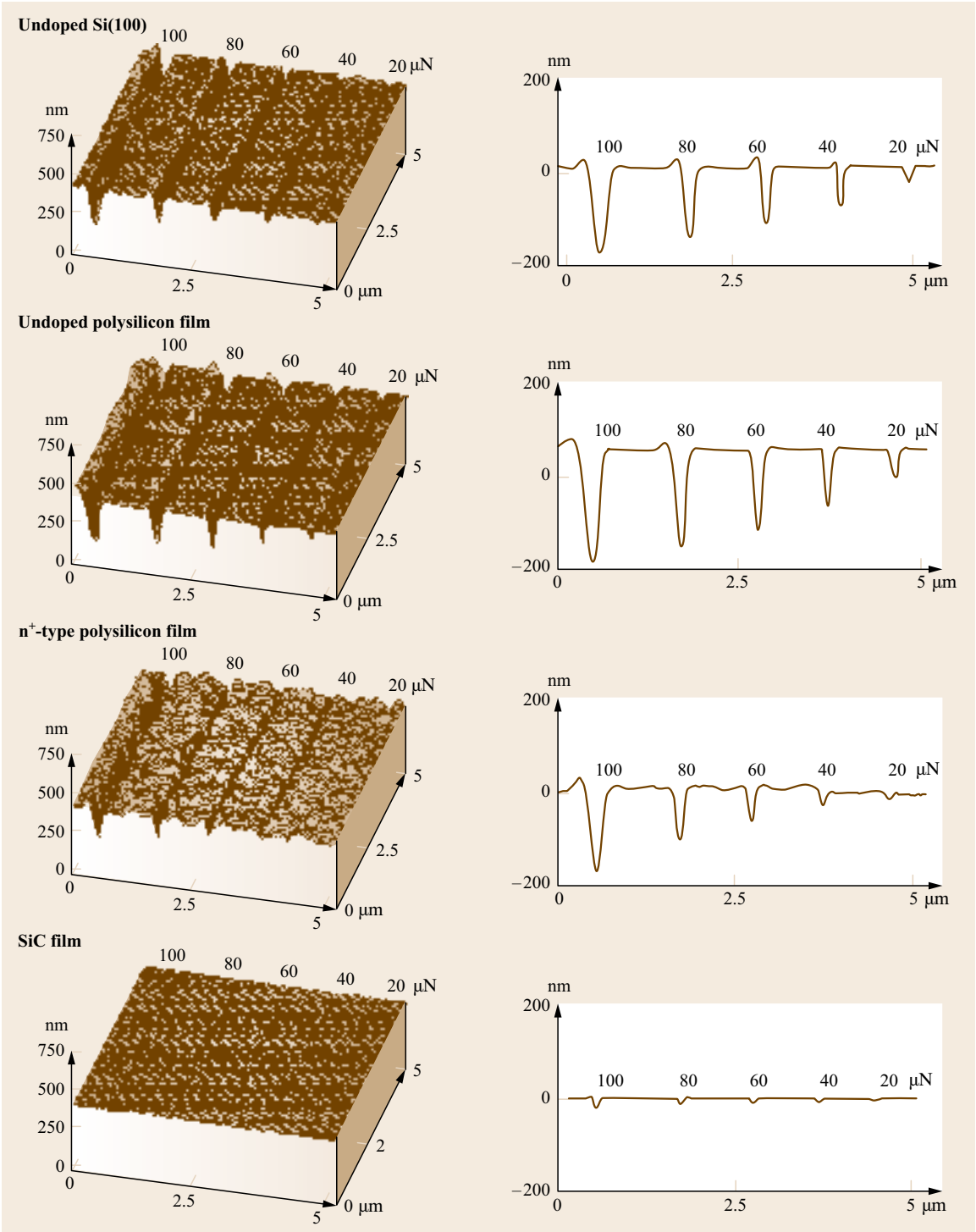


Fig. 38.25 AFM 3-D maps and averaged 2-D profiles of scratch marks on various samples (after [38.194])

responsible for its lower wear. Also, the higher thermal conductivity of SiC (Table 38.2) as compared to the other materials leads to lower interface temperatures, which generally results in less degradation of the surface [38.32, 91, 92]. Doping of the polysilicon does not affect the scratch/wear resistance and hardness much. The measurements made on the doped sample are affected by the presence of grain boundaries. These studies indicate that SiC film exhibits desirable tribological properties for use in MEMS devices.

38.4 Lubrication Studies for MEMS/NEMS

MEMS/NEMS devices may benefit from various lubricants. Several studies of liquid perfluoropolyether (PFPE) lubricant films, self-assembled monolayers (SAMs), and hard diamond-like carbon (DLC) coatings have been carried out for the purpose of minimizing adhesion, friction, and wear [38.44, 45, 86–92, 172, 173, 193]. Many variations of these films are hydrophobic (low surface tension and high contact angle) and have low shear strength, which provides low adhesion, friction, and wear. Relevant details are presented here.

38.4.1 Perfluoropolyether Lubricants

The classical approach to lubrication uses freely supported multimolecular layers of liquid lubricants [38.44, 88, 91, 92]. The liquid lubricants are sometimes chemically bonded to improve their wear resistance. Partially chemically bonded, molecularly thick PFPE lubricants are widely used for lubrication of magnetic storage media because of their thermal stability and extremely low vapor pressure [38.32], and are found to be suitable for MEMS/NEMS devices.

Adhesion, friction, and durability experiments have been performed on virgin Si(100) surfaces and silicon surfaces lubricated with various PFPE lubricants [38.44, 45, 197–203]. Results of two PFPE lubricants will be presented here: Z-15 (with $-CF_3$ nonpolar end groups),

$$CF_3{-}O{-}(CF_2{-}CF_2{-}O)_m{-}(CF_2{-}O)_n{-}CF_3$$

($m/n \approx 2/3$), and Z-DOL (with $-OH$ polar end groups),

$$HO{-}CH_2{-}CF_2{-}O{-}(CF_2{-}CF_2{-}O)_m{-}(CF_2{-}O)_n{-}CF_2{-}CH_2{-}OH$$

($m/n \approx 2/3$). The Z-DOL film was thermally bonded at 150 °C for 30 min, and the unbonded fraction was removed by a solvent (fully bonded) [38.32]. The thicknesses of Z-15 and the Z-DOL (fully bonded) films were 2.8 and 2.3 nm respectively. Nanoscale measurements were made using an AFM. The adhesive forces of Si(100), Z-15, and Z-DOL (fully bonded) measured by force calibration plot, and friction force versus normal load plot are summarized in Fig. 38.26. The results measured by these two methods are in good agreement. Figure 38.26 shows that the presence of mobile Z-15 lubricant film increases the adhesive force as compared to that of Si(100) by meniscus formation [38.91, 92, 204]; whereas, the presence of solid-phase Z-DOL (fully bonded) film reduces the adhesive force as compared that of Si(100) because of the absence of mobile liquid. The schematic (bottom) in Fig. 38.26 shows relative size and sources of meniscus. It is well known that the native oxide layer (SiO_2) on the top of the Si(100) wafer exhibits hydrophilic properties, and some water molecules can be adsorbed onto this surface. The con-

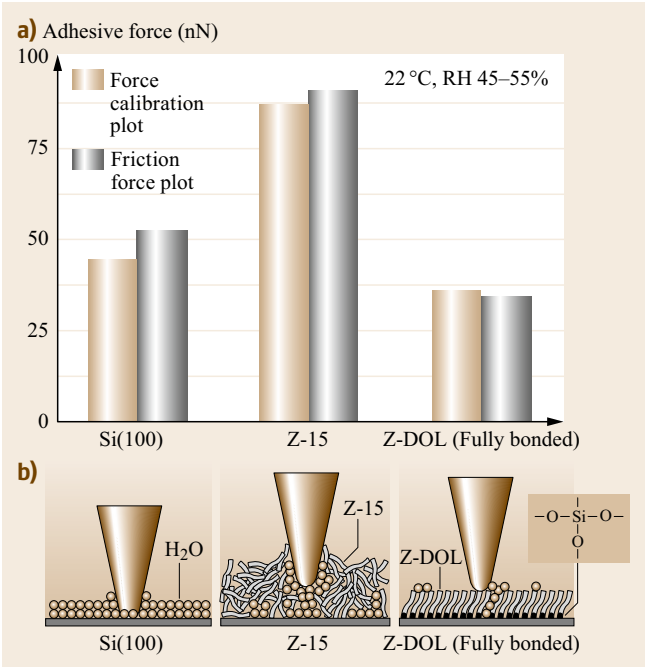


Fig. 38.26 **(a)** Summary of the adhesive forces of Si(100) and Z-15 and Z-DOL (fully bonded) films measured by force calibration plots and friction force versus normal load plots in ambient air. **(b)** Schematic showing the effect of the meniscus, formed between AFM tip and the surface sample, on the adhesive and friction forces (after [38.199])

densed water will form a meniscus as the tip approaches the sample surface. The larger adhesive force in Z-15 is not just caused by the Z-15 meniscus; the nonpolarized Z-15 liquid does not have good wettability and strong bonding with Si(100). Consequently, in the ambient environment, the condensed water molecules from the environment permeate through the liquid Z-15 lubricant film and compete with the lubricant molecules presented on the substrate. The interaction of the liquid lubricant with the substrate is weakened, and a boundary layer of the liquid lubricant forms puddles [38.197, 198]. This dewetting allows water molecules to be adsorbed on the Si(100) surface as aggregates along with Z-15 molecules, and both of them can form a meniscus while the tip approaches the surface. Thus, the dewetting of the liquid Z-15 film results in higher adhesive force and poorer lubrication performance. In addition, as the Z-15 film is soft compared to the solid Si(100) surface, penetration of the tip in the film occurs while pushing the tip down. This leads to a large area of the tip involved in forming the meniscus at the tip–liquid (mixture of Z-15 and water) interface. It should also be noted that Z-15 has a higher viscosity compared to water, therefore the Z-15 film provides higher resistance to lateral motion and coefficient of friction. In the case of the Z-DOL (fully bonded) film, the active groups of Z-DOL molecules are mostly bonded on the Si(100) substrate, thus the Z-DOL (fully bonded) film has low free surface energy and cannot be displaced readily by water molecules or readily adsorb water molecules. Thus, the use of Z-DOL (fully bonded) can reduce the adhesive force.

To study the relative humidity effect on friction and adhesion, the variation of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (fully bonded) as a function of relative humidity are shown in Fig. 38.27. It shows that for Si(100) and Z-15 films, the friction force increases with a relative humidity increase up to 45%, and then it shows a slight decrease with further increase in the relative humidity. Z-DOL (fully bonded) has a smaller friction force than Si(100) and Z-15 in the whole testing range, and its friction force shows a relative apparent increase when the relative humidity is higher than 45%. For Si(100), Z-15, and Z-DOL (fully bonded), their adhesive forces increase with relative humidity, and their coefficients of friction increase with a relative humidity up to 45%, after which they decrease with further increasing of the relative humidity. It is also observed that the humidity effect on Si(100) really depends on the history of the Si(100) sample. As the surface of the Si(100) wafer readily adsorbs water in air, without any pretreatment the Si(100) used in our study almost reaches its saturated stage of adsorbed water, and is responsible for less of an effect during increasing relative humidity. However, once the Si(100) wafer was thermally treated by baking at 150 °C for 1 h, a greater effect was observed.

The schematic in Fig. 38.27b shows that Si(100), because of its high free surface energy, can adsorb more water molecules while increasing relative humidity. As discussed earlier, for the Z-15 film in the humid environment, the condensed water competes with the lubricant film present on the sample surface, and interaction of the liquid lubricant film with the silicon substrate is weakened, and a boundary layer of the liquid lubricant forms puddles. This dewetting allows water molecules to be adsorbed on the Si(100) substrate mixed with Z-15 molecules [38.197, 198]. Obviously, more water molecules can be adsorbed on the Z-15 surface with increasing relative humidity. The more adsorbed water molecules in the case of Si(100), along with lubricant molecules in the Z-15 film case, form a bigger water meniscus, which leads to an increase of friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 with humidity. But at a very high humidity of 70%, large quantities of adsorbed water can form a continuous water layer that separates the tip and sample surface, acting as a kind of lubricant, which causes a decrease in the friction force and coefficient of friction. For Z-DOL (fully bonded) film, because of their hydrophobic surface properties, water molecules can be adsorbed at a humidity higher than 45%, and cause an increase in the adhesive force and friction force.

To study the durability of lubricant films at the nanoscale, the friction forces of Si(100), Z-15, and Z-DOL (fully bonded) as a function of the number of scanning cycles are shown in Fig. 38.28. As observed earlier, the friction force of Z-15 is higher than that of Si(100), with the lowest values for Z-DOL (fully bonded). During cycling, the friction force of Si(100) shows a slight decrease during the initial few cycles then remains constant. This is related to the removal of the top adsorbed layer. In the case of the Z-15 film, the friction force shows an increase during the initial few cycles and then approaches higher and stable values. This is believed to be caused by the attachment of the Z-15 molecules onto the tip. The molecular interaction between these attached molecules to the tip and molecules on the film surface is responsible for an increase in friction. However, after several scans, this molecular interaction reaches equilibrium, and after that, friction force and the coefficient of friction remain constant. In the case of the Z-DOL (fully bonded) film, the friction force and coefficient of friction start out low and remain low during the entire test for 100 cycles. It suggests that Z-DOL (fully bonded) molecules do not get attached or displaced as readily as Z-15.

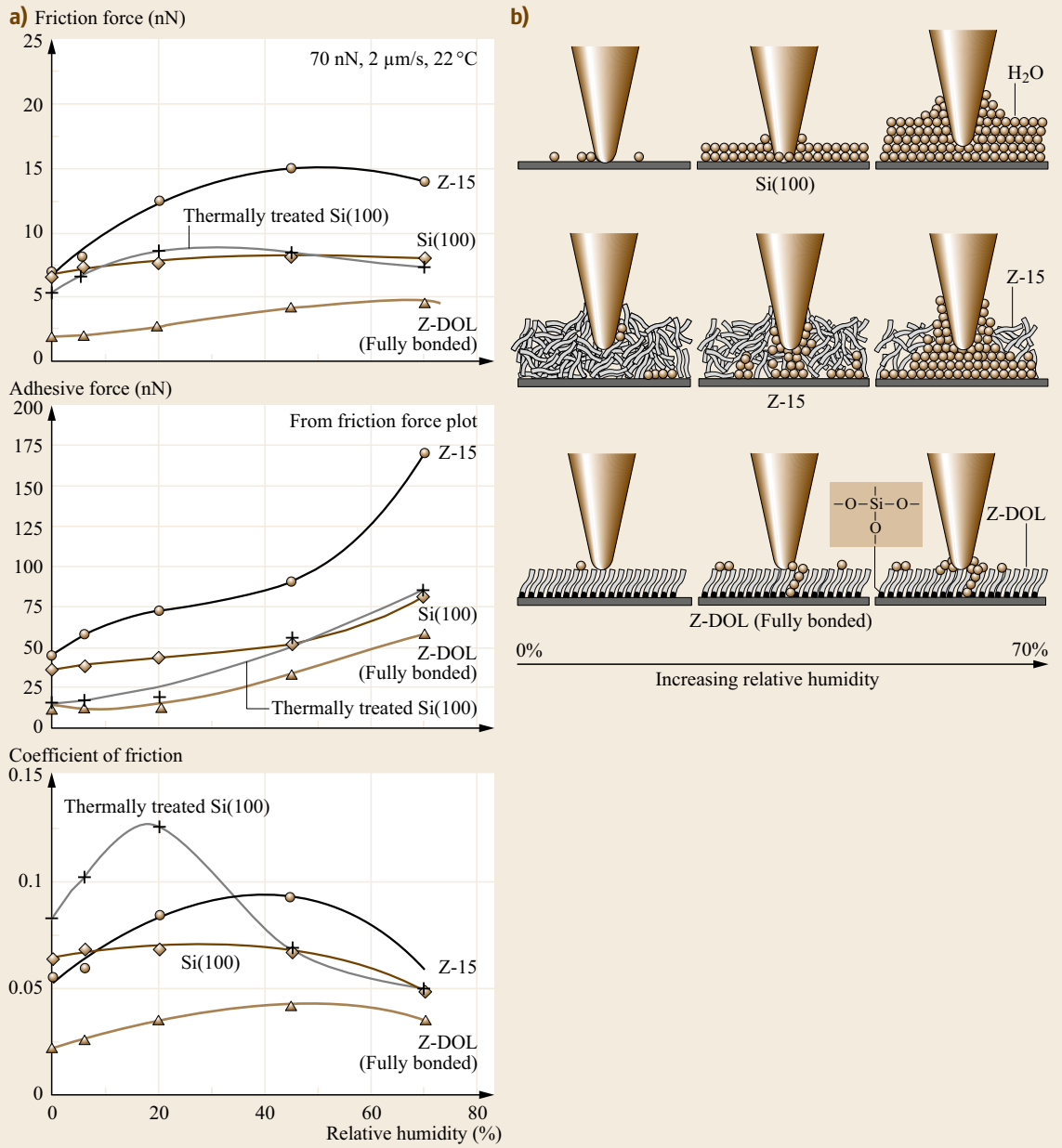


Fig. 38.27 **(a)** The influence of relative humidity of the friction force, adhesive force, and coefficient of friction of Si(100), Z-15 and Z-DOL (fully bonded) films at 70 nN, 2 µm/s, and in 22 °C air. Schematic **(b)** shows the change of meniscus while increasing the relative humidity. In this figure, the thermally treated Si(100) represents the Si(100) wafer that was baked at 150 °C for 1 h in an oven (in order to remove the adsorbed water) just before it was placed in the 0% RH chamber (after [38.199])

38.4.2 Self-Assembled Monolayers (SAMs)

For lubrication of MEMS/NEMS, another effective approach involves the deposition of organized and dense molecular layers of long-chain molecules. Two common methods to produce monolayers and thin films are the Langmuir–Blodgett (L–B) deposition and self-assembled monolayers (SAMs) by chemical grafting of molecules. L–B films are physically bonded to the substrate by a weak van der Waals attraction, while

SAMs are chemically bonded via covalent bonds to the substrate. Because of the choice of chain length and terminal linking group that SAMs offer, they hold great promise for boundary lubrication of MEMS/NEMS.

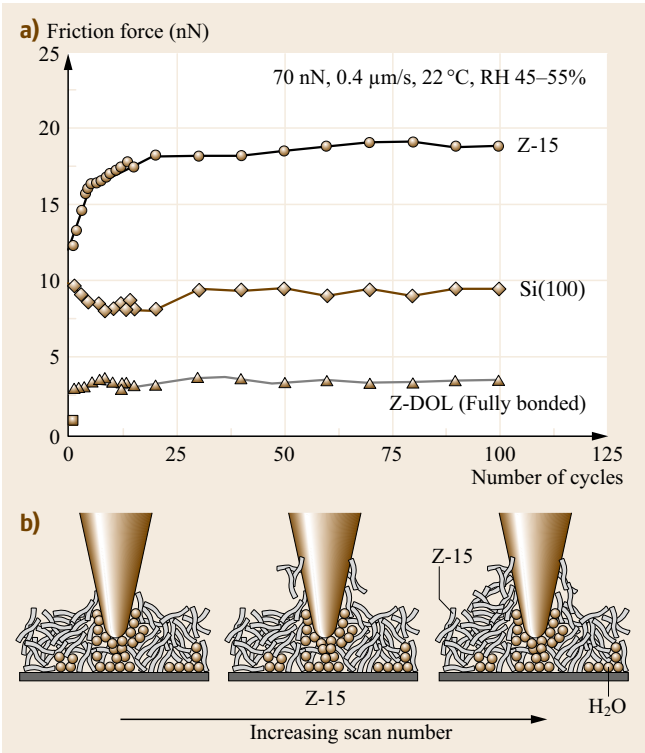


Fig. 38.28 (a) Friction force versus number of sliding cycles for Si(100), Z-15 and Z-DOL (fully bonded) films at 70 nN, 0.4 μm/s, and in ambient air. Schematic (b) shows that some liquid Z-15 molecules can be attached onto the tip. The molecular interaction between the attached molecules onto the tip with the Z-15 molecules in the film results in an increase of the friction force with multiscanning (after [38.199])

A number of studies have been conducted on tribological properties of various SAMs [38.152–155, 205–211].

Bhushan and *Liu* [38.206] studied the effect of film compliance on adhesion and friction. Based on friction and stiffness measurements, SAMs with high-compliance long-carbon chains exhibit low friction; chain compliance is desirable for low friction. The friction mechanism of SAMs is explained by a so-called *molecular spring* model. According to this model, the chemically adsorbed self-assembled molecules on a substrate are just like assembled molecular springs anchored to the substrate. An asperity sliding on the surface of SAMs is like a tip sliding on the top of *molecular springs or brush*. The molecular spring assembly has compliant features and can experience orientation and compression under load. The orientation of the molecular springs or brush under normal load reduces the shearing force at the interface, which in turn reduces the friction force. The orientation is determined by the spring constant of a single molecule as well as the interaction between the neighboring molecules, which can be reflected by packing density or packing energy. It should be noted that the orientation can lead to conformational defects along the molecular chains, which leads to energy dissipation. The SAMs with high-compliance long-carbon chains also exhibit the best wear resistance [38.206, 208]. In wear experiments, the wear depth as a function of normal load curves show a critical normal load. Below the critical normal load, SAMs undergo orientation. At the critical load, SAMs wear away from the substrate due to weak interface bond strengths, while above the critical normal load, severe wear takes place on the substrate.

Bhushan et al. [38.152, 209], *Kasai* et al. [38.154], *Tambe* and *Bhushan* [38.210], and *Tao* and *Bhushan* [38.211] studied various SAMs that were vapor-phase deposited on Si and Al substrates with native oxide films, Fig. 38.29. Perfluorodecyltricholorosi-

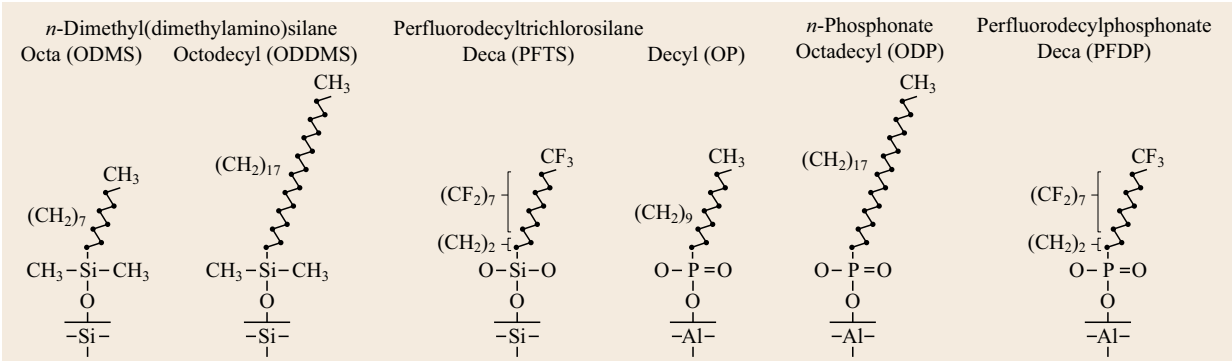


Fig. 38.29 Schematics of the structures of perfluoroalkylsilane and alkylsilane SAMs on Si with native oxide substrates, and perfluoroalkylphosphonate and alkylphosphonate SAMs on Al with native oxide substrates

lane (PFTS), *n*-octyldimethyl (dimethylamino) silane (ODMS) ($n = 7$), and *n*-octadecylmethyl (dimethylamino) silane (ODDMS) ($n = 17$) were deposited on Si substrates. Perfluorodecylphosphonate (PFDP), *n*-decylphosphonate (DP) ($n = 7$) and *n*-octadecylphosphonate (ODP) ($n = 17$) were deposited on Al substrates. Figure 38.30 shows the static contact angles, adhesive forces, friction forces, and coefficients of friction of the two substrates with various SAMs under ambient conditions [38.152, 209]. Based on the data, all films exhibit higher contact angles and lower adhesive forces and coefficients of friction as compared to their corresponding substrates. Among various films, PFTS/Si exhibits a higher contact angle and lower adhesive force as compared to that of ODMS/Si or ODDMS/Si. Longer-chain film ODDMS/Si has superior performance than shorter-chain film ODMS. Trends for films on Al substrates are similar to that on Si substrates. Thus, substrate had little effect. The coefficients of friction for various SAMs are comparable.

The effect of relative humidity on adhesion and friction for various SAMs on Si substrate was studied. Adhesive force, friction force at 5 nN of normal load, coefficient of friction, and microwear data are presented in Fig. 38.31 [38.154, 209]. Adhesive force for silicon showed an increase with relative humidity. This is expected since the surface of silicon is hydrophilic as shown in Fig. 38.30. More condensation of water at the tip–sample interface at higher humidity increases the adhesive force due to the capillary effect. On the other hand, the adhesive force for the SAMs showed a very weak dependency on the change in humidity. This occurs since the surface of the SAMs is hydrophobic. The adhesive force of ODMS/Si and ODDMS/Si showed a slight increase from 75 to 90% RH. Such an increase was absent for PFTS/Si, possibly because of the hydrophobicity of PFTS/Si. The friction force of silicon showed an increase with relative humidity up to about 75% RH and a slight decrease beyond this point. The initial increase possibly results from the increase in adhesive force. The decrease in friction force at higher humidity could be attributed to the lubricating effect of the water layer. This effect is more pronounced in the coefficient of friction. Since the adhesive force increased and the coefficient of friction decreased in this range, those effects cancel out each other, and the resulting friction force showed only slight changes. On the other hand, the friction force and coefficient of friction of SAMs showed very small changes with relative humidity, like that found for adhesive force. This suggests that the adsorbed water layer on the surface maintained a similar thickness throughout the relative humidity range tested. The differences among the SAM types were small, within the measurement error. However a closer look at the coefficient of friction for ODMS/Si showed a slight increase from 75 to 90% RH as compared to PFTS/Si, possibly for the same reason as the adhesive force

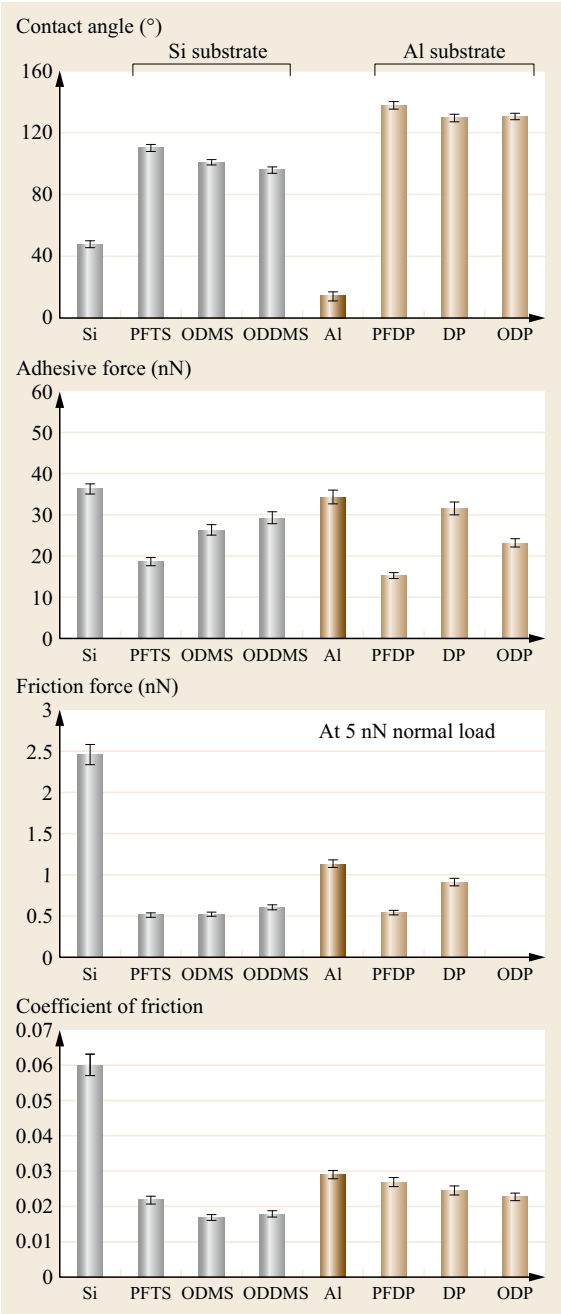


Fig. 38.30 The static contact angle, adhesive force, friction force, and coefficient of friction measured using an AFM for various SAMs on Si and Al substrates (after [38.152, 209])

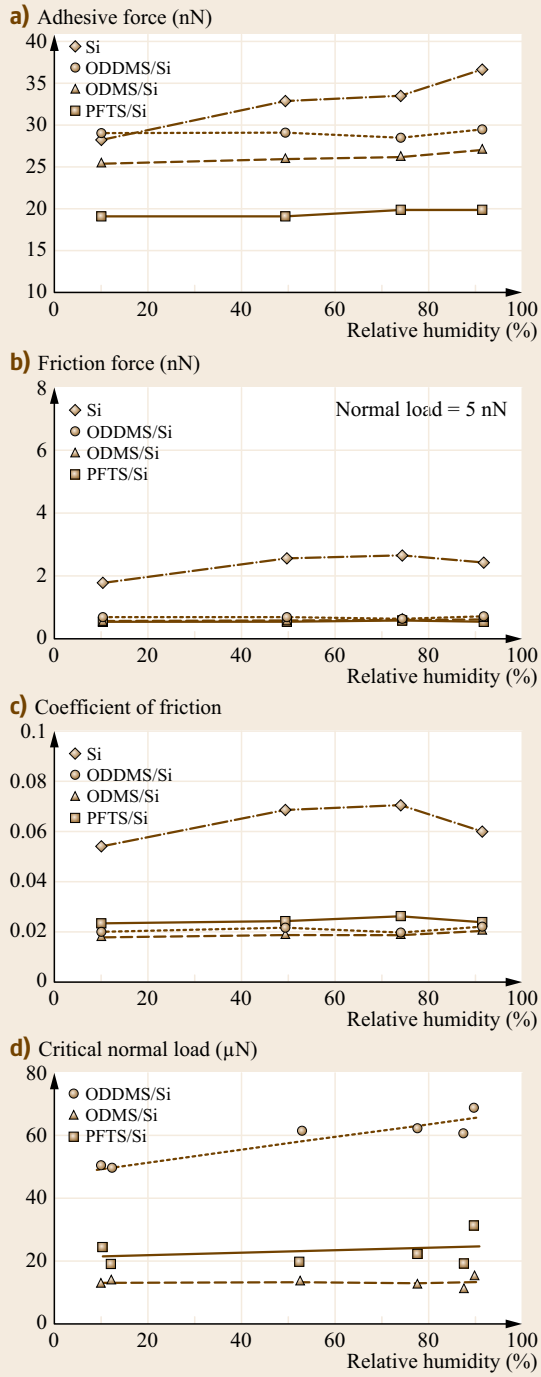


Fig. 38.31a–d Relative humidity effect on (a) adhesive force, (b) friction force, (c) coefficient of friction, and (d) microwear for various SAMs on Si substrates (after [38.154, 209])

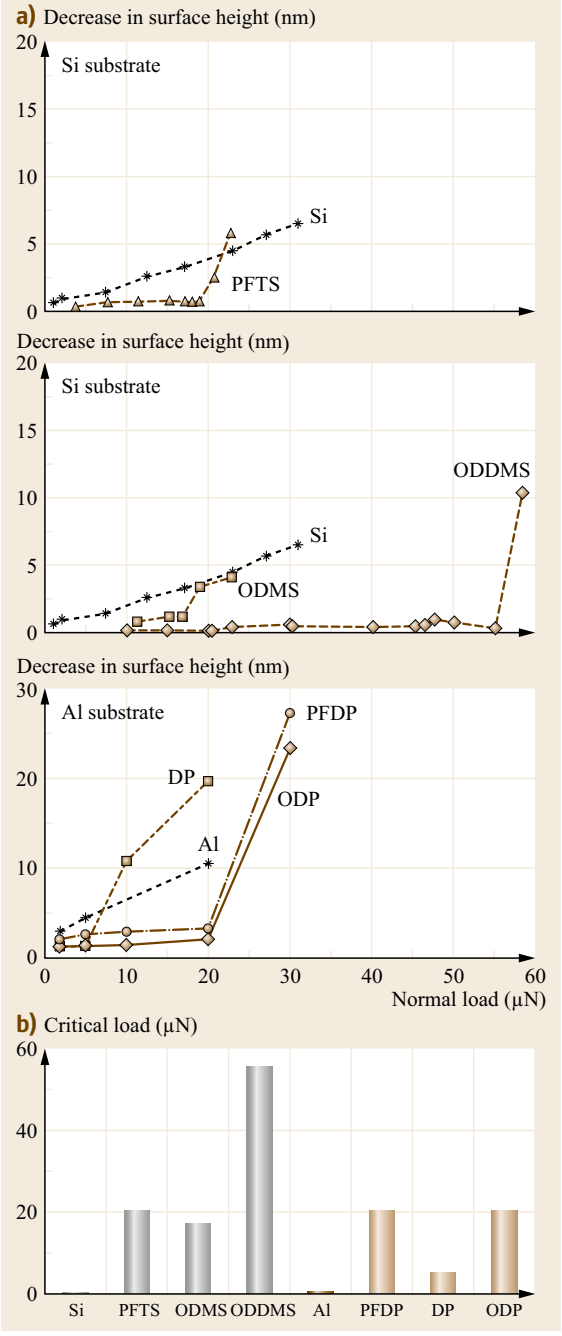


Fig. 38.32 (a) Decrease in surface height as a function of normal load after one scan cycle for various SAMs on Si and Al substrates, and (b) comparison of critical loads for failure during wear tests for various SAMs (after [38.154, 209])

changes. The inherent hydrophobicity of SAMs means that they did not show much relative humidity dependence.

Figure 38.32a shows the relationship between the decrease in surface height as a function of the normal load during wear tests for various SAMs on Si and Al substrates [38.154, 209]. As shown in the figure, the SAMs exhibit a critical normal load beyond the point of which the surface height drastically decreases. Figure 38.32a also shows the wear behavior of the Al and Si substrates. Unlike the SAMs, the substrates show a monotonic decrease in surface height with increasing normal load, with wear initiating from the very beginning, i. e., even for low normal loads. Si (Young's modulus of elasticity, $E = 130$ GPa [38.212], hardness, $H = 11$ GPa [38.193]) is relatively hard in comparison to Al ($E = 77$ GPa, $H = 0.41$ GPa). Hence the decrease in surface height for Al is much larger than that for Si for similar normal loads.

The critical loads corresponding to the sudden failure of SAMs are shown in Fig. 38.32b. Amongst all the SAMs, ODDMS shows the best performance in the wear tests, believed to be due to the longer chain length effect. Fluorinated SAMs – PFTS and PFDP show higher critical load as compared to ODMS and DP with similar chain lengths. ODP shows higher critical load as compared to DP because of its longer chain length. The mechanism of failure of compliant SAMs during wear tests was presented earlier. It is believed that the SAMs fail mostly due to shearing of the molecule at the head group; that is, by means of shearing of the molecules off the substrate.

To study the effect of relative humidity, wear tests were performed at various humidities. The bottom of Fig. 38.31 shows critical normal load as a function of relative humidity. The critical normal load showed weak dependency on relative humidity for ODMS/Si and PFTS/Si, and was larger for ODMS/Si than PFTS/Si throughout the humidity range. This suggests that water molecules could penetrate into the ODDMS, which might work as lubricant [38.208]. This effect was absent for PFTS/Si and ODMS/Si.

Bhushan et al. [38.153] and *Lee* et al. [38.155] studied various fluoropolymer multilayers and fluorosilane monolayers on Si and a selected fluorosilane on poly(dimethylsiloxane) (PDMS) surfaces. For nanoscale devices such as nanochannels, monolayers are preferred. They reported that all fluorosilane films increased the contact angle. The fluorosilane monolayer – 1H, 1H, 2H, 2H – perfluorodecyltriethoxysilane (PFDTES) resulted in a contact angle of about 100°.

Based on these studies, a perfluoro-SAM with a compliant layer should have optimized tribological performance for MEMS/NEMS and bioMEMS/bioNEMS applications.

38.4.3 Hard Diamond-Like Carbon (DLC) Coatings

Coating of hard amorphous carbon (a-C), commonly known as diamond-like carbon (DLC) (implying high hardness), are deposited by a variety of deposition techniques, including filtered cathodic arc (FCA), ion beam, electron cyclotron resonance chemical vapor deposition (ECR-CVD), plasma-enhanced chemical vapor deposition (PECVD), and sputtering [38.181, 193]. These coatings are used in a wide range of applications, including tribological, optical, electronic, and biomedical applications. Ultrathin coatings (3–10 nm thick) are employed to protect against wear and corrosion in magnetic storage applications – thin-film rigid disks, metal evaporated tapes, and thin-film read/write heads, Gillette Mach 3 razor blades, glass windows, and sunglasses. The coatings exhibit low friction, high hardness and wear resistance, chemical inertness to both acids and alkalis, lack of magnetic response, and optical bandgaps ranging from zero to a few eV, depending upon the deposition technique and its conditions. Selected data on DLC coatings relevant for MEMS/NEMS applications is presented in a later section on microtriboapparatus.

38.5 Nanoscale Friction, Wear, and Mechanical Behavior of Nano-Objects

Nano-objects are described as a single material or composite having at least one dimension that is between 1 and 100 nm. They come in a variety of discrete geometries that include spheres, tubes, rods, wires, and pillars. Compared to their bulk material counterparts, many nano-objects exhibit enhanced mechanical, electrical, magnetic, and chemical properties. This makes them attractive for use in many applications on the macro- to nanoscale. Many nano-objects also exhibit low friction and wear properties. These are used in various sliding applications in which they come into sliding contact with each other and the mating surfaces. Contacts can lead to nano-object deformation. For these situations, friction, wear, and nano-object deformation are of interest.

Maharaj and *Bhushan* [38.213] studied Au nanoparticles (NPs) and nanorods (NRs), MoS_2, and WS_2 multiwalled nanotubes (MWNTs), and carbon nanohorns (CNHs). These nano-objects were chosen since they have either been used in applications (discussed earlier), or have been of interest. They also provide an opportunity to explore the relationship between their friction and wear-reduction ability in relation to their geometry, mechanical properties, and deformation mechanisms under different loading conditions. Figure 38.33 shows transmission electron microscopy (TEM) images of spherical Au nanoparticles approximately 50 nm in diameter, and Au nanorods 50 nm in diameter and 200 nm in length. Figure 38.34 shows representative TEM images of the MoS_2 and WS_2 MWNTs and CNHs. Table 38.6 shows the water contact angle of various thin films. Tables 38.7 and 38.8 show nanomechanical properties of various nanoparticles, MWNTs, films, and bulk materials.

Au nano-objects have been used in applications requiring targeted drug delivery due to their chemical inertness and low to zero toxic effects. In tribological studies, their small contact area and mobility due to their spherical shape allows them to roll and slide. This results in reduced friction and wear on the macro- to nanoscale in dry and submerged-in-liquid environments [38.181, 224–228]. Au thin films also have been studied to investigate further scale effects on the nanoscale. Their flat surfaces prevent inaccuracies in measuring mechanical properties due to curvature effects of increasingly smaller nano-objects.

Table 38.6 Water contact angle for thin films of Au, MoS_2, WS_2, and graphene in contact with water (after [38.224])

Films	Water contact angle (°)
Au	50[a]
MoS_2	75[b]
WS_2	93[c]
Graphene	127[d]

[a] [38.214]; [b] [38.215]; [c] [38.216]; [d] [38.217]

MoS_2 and WS_2 have a layered lattice structure with weak bonding between layers, which provides low friction, and these are known solid lubricants [38.91, 92, 181]. MoS_2 and WS_2 MWNTs have also been studied in tribology on the macro- to nanoscale in dry and liquid environments and increasingly in other applications such as catalysis in HER. MoS_2 and WS_2 MWNTs have been shown to reduce friction and wear similar to Au nano-objects. In addition, the exfoliation of the outer layers of the nanotubes and subsequent formation of an easily shearable film aids in friction and wear reduction [38.181, 227, 229].

A single CNH has a structure similar to a single-walled carbon nanotube with one end closed in a cone-shaped cap (horn). These horns agglomerate through van der Waals attractions to form a spherical nano-object with unique morphology. CNHs have been the subject of myriad studies due to their unique surface morphology and the wide-ranging properties of graphene, which include chemical stability, low surface energy [38.230], high thermal [38.231] and electrical conductivity [38.232], and low friction and wear [38.224].

For nano-objects submerged in liquid environments, water, dodecane, and glycerol have been used [38.213]. They have been utilized in friction and wear studies as they are simple liquids without additives, and their properties are well known.

Maharaj and *Bhushan* [38.213] carried out single nano-object contact and multiple nano-object contact studies. In single nano-object contact, Au NPs and NRs,

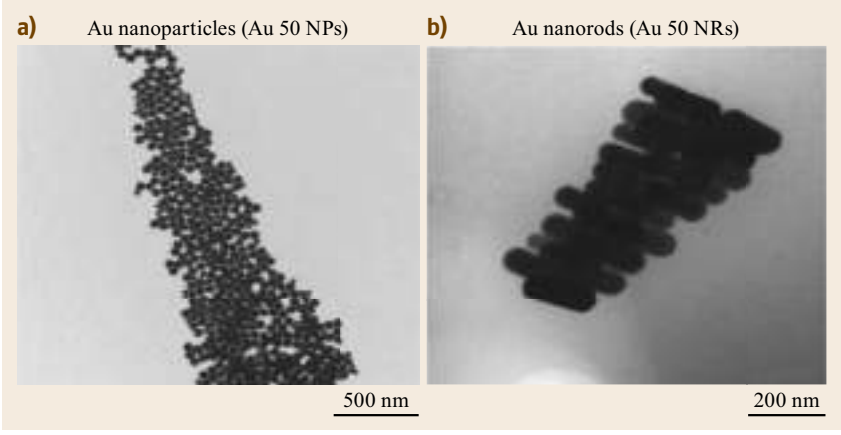


Fig. 38.33a,b TEM images of **(a)** spherical Au nanoparticles approximately 50 nm in diameter, and **(b)** Au nanorods 50 nm in diameter and 200 nm in length. Copyright Nanopartz Inc., USA

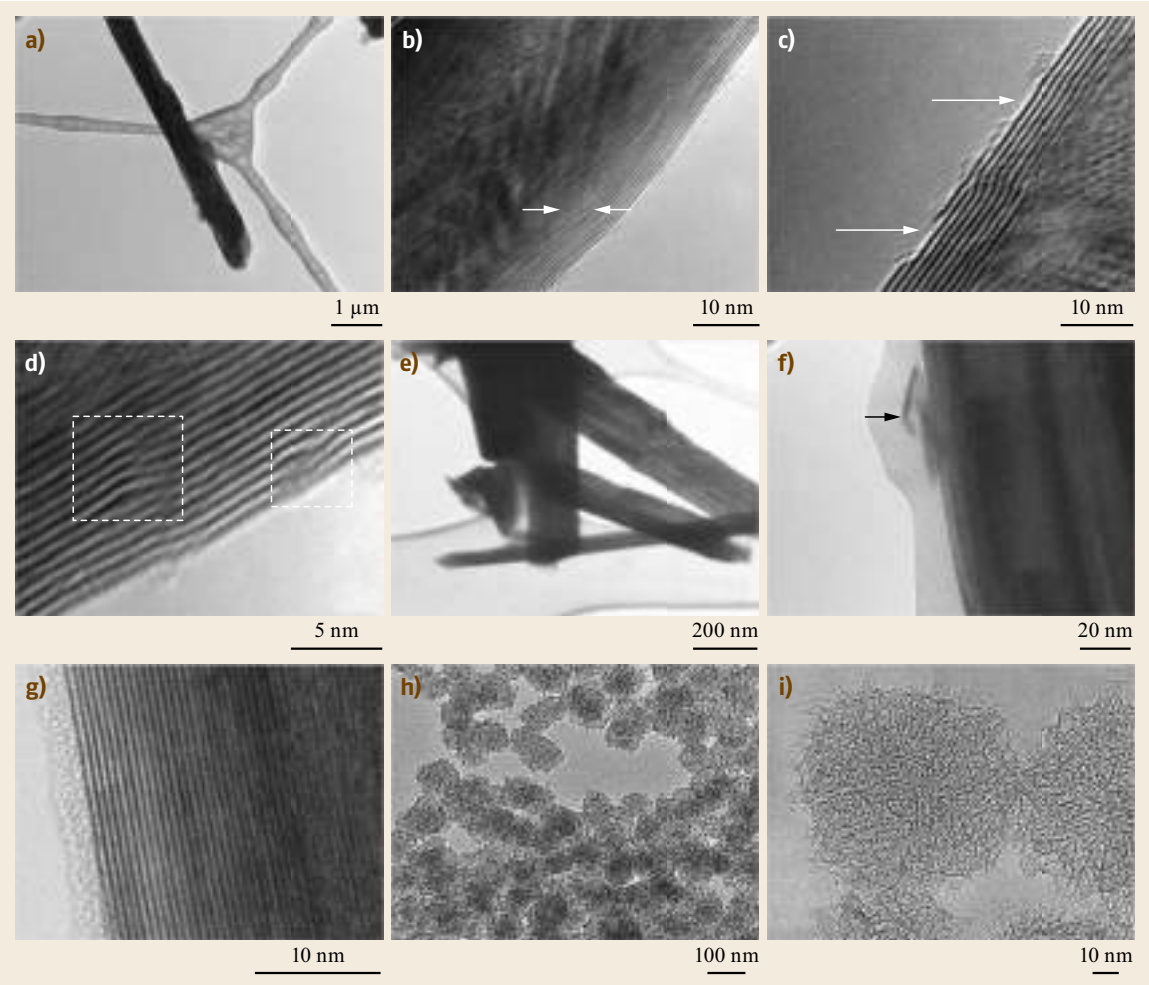


Fig. 38.34a–i TEM images of **(a)** MoS_2 MWNT, a magnified image showing separated layers between the *arrows* **(b)**, magnified image showing broken surface layers **(c)**, and a magnified image of a nanotube showing stacking faults highlighted by the *dashed lines* **(d)**; agglomerated WS_2 MWNTS **(e)**, a magnified image with an *arrow* pointing to exfoliated layers **(f)**, and a magnified image of a nanotube showing ordered layers **(g)**; spherical CNHs **(h)** with magnified image of CNHs **(i)**. **(a–g)** from [38.220], **(h,i)** reprinted from [38.233] with permission from Elsevier

Table 38.7 Nanomechanical properties of Au nanoparticles and thin films at a maximum load of 80 µN compared to bulk Au (after [38.218])

	Nanoparticle diameter/ film thickness	Contact depth (nm)	Hardness (GPa)	Young's modulus of elasticity (GPa)	Explanation for enhanced hardness
Au 500[a]	≈ 500 nm	45±4	0.7±0.1	70±11	Hall–Petch, strain gradient plasticity
Au thin film (100 nm)[a]	≈ 100 nm	15±1	1±0.1	76±11	Hall–Petch, strain gradient plasticity
Bulk Au[b]	5 mm		0.22	78	

[a] [38.218]
[b] [38.219]

MoS_2 and WS_2 MWNTs, and CNHs have been studied with an AFM as illustrated in Fig. 38.35 where they are manipulated by lateral pushing. Friction forces were examined in dry and submerged-in-liquid environments. Friction was found to be dependent on size, shape, material, and roughness. Friction was found to be lower in submerged-in-liquid environments compared to dry conditions. These studies provide an understanding of the nature of the friction mechanism of nano-objects at the initiation of sliding.

In multiple nano-object contact studies, the effect of nano-objects on friction and wear reduction was determined. Sliding takes place over several nano-objects, as illustrated in Fig. 38.35. Real area of contact, rolling, sliding, and exfoliation and tribofilm formation were studied for their effect on friction and wear. Results of studies on the macroscale were compared to the nanoscale to understand scale effects of friction.

An overview of friction and wear in dry and liquid environments is presented next, and various mechanisms are presented to explain the friction and wear behavior of nano-objects.

38.5.1 Single Nano-Object Friction

For nano-objects that are spherical in nature, manipulation can involve rolling, sliding, and rotation about the vertical axis. Manipulation can also involve rolling, sliding, and in-plane rotation in the case of the MoS_2 and WS_2 MWNTs. In-plane rotation can occur when the nanotubes are not pushed directly at the center of the MWNT's length.

Figure 38.36a presents the results for the average friction forces for Au 90 NPs, CNHs and MoS_2 and WS_2 MWNTs in dry and submerged-in-liquid environments. The friction force presented is the result of adhesion between the nano-objects and the silicon substrate. The adhesive force can include both van der Waals and meniscus forces for dry environments. However, under submerged-in-liquid environments, meniscus forces are eliminated.

Friction forces for the CNHs and MoS_2 and WS_2 MWNTs compared to Au 90 NPs are lower due to a material effect. For the MWNTs it is not believed that exfoliation and tribofilm formation are respon-

Table 38.8 Nanomechanical properties of MoS_2 MWNTs obtained from the nanoindentation experiments normal to the basal plane. For comparison, mechanical properties of naturally occurring MoS_2 (normal to the basal planes) and single-layer graphene are given (after [38.220])

Mechanical properties	MoS_2 MWNTs[a] (500 nm)	MoS_2[b,c] (2 mm)	Graphene[d,e] (single layer)
Hardness (GPa)	0.6 ± 0.2	0.6	950
Young's modulus (GPa)	12.5 ± 4	52	1000

[a] [38.220]; [b] [38.181]; [c] [38.221]; [d] [38.222]; [e] [38.223]

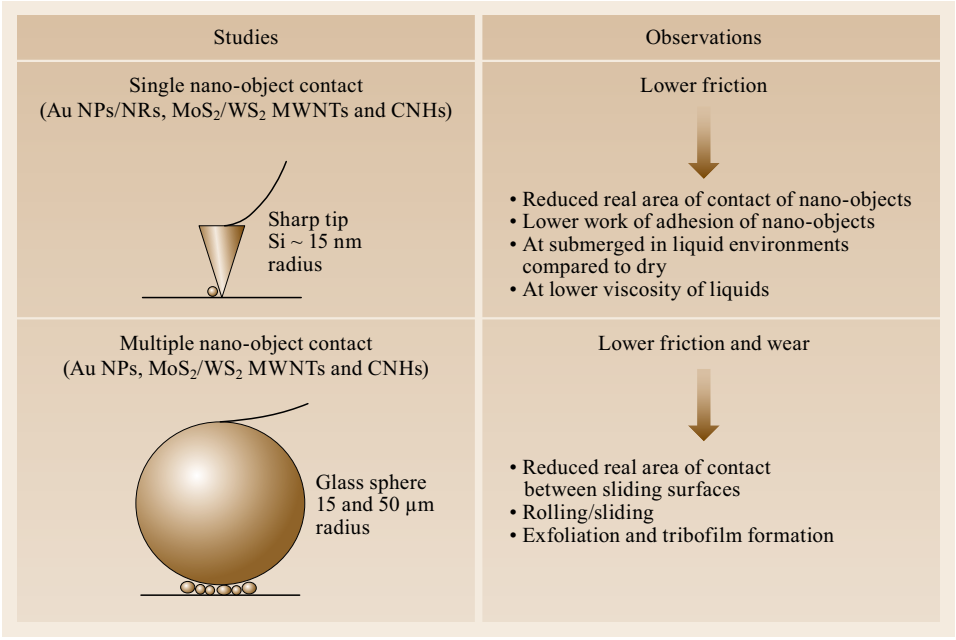


Fig. 38.35 Illustration of the studies in friction and wear of nano-objects. The studies examined were single and multiple nano-object contact in dry and liquid environments. Observations for each study are given in the column on the right

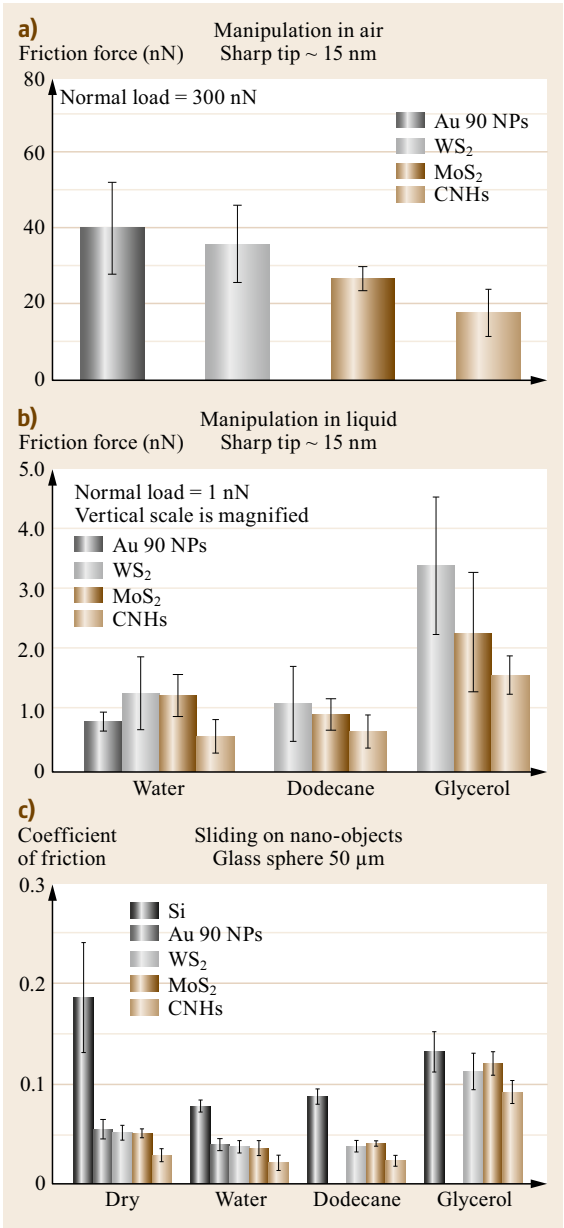


Fig. 38.36a–c Friction force for Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs on the silicon substrate during manipulation, at normal loads of **(a)** 300 nN in air and **(b)** 1 nN in liquids. The vertical scale is magnified for data in liquids. **(c)** Coefficients of friction for both dry and submerged-in-liquid environments, with and without nano-objects. Friction forces and coefficient of friction are lower in submerged-in-water environments (after [38.224])

sible for the lower values, as discussed by *Maharaj* et al. [38.224], since there is no normal load acting on them.

Figure 38.36b presents the friction forces for the nanoparticles and nanotubes submerged in liquid environments. For Au NPs that were already suspended in DI water, experiments were not carried out in dodecane and glycerol. In dodecane and glycerol, similar to water, adhesive forces are due to van der Waals interactions since meniscus forces are eliminated. The lower friction forces in submerged-in-liquid environments compared to the dry environments can be attributed to both elimination of meniscus forces and nano-objects sliding on a low shear-strength surface. The lowest friction forces occur with the CNHs due to the low contact area provided by the roughness of the nanohorns. Here, the material effects due to the work of adhesion do not contribute to friction. The highest friction force occurs in glycerol, which has the highest viscosity and therefore the greatest contribution to viscous drag force. Similar viscosity of dodecane and water accounts for the similarity in friction forces observed for both liquids with and without nano-objects.

The characteristics of the nano-objects discussed in single nano-object contact studies will also be investigated in multiple nano-object contact studies in the context of reducing friction and wear under load. As will be discussed in the next section, new nano-object behavior not previously seen will contribute to friction and wear reduction due to the application of a load.

38.5.2 Multiple Nano-Object Contact

Nanoscale Friction and Wear

In multiple nano-object contact, the effect of normal load acting on the Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs between two surfaces is studied to determine the effects on friction and wear. Figure 38.36c summarizes the coefficient of friction for dry and submerged-in-liquid environments. Sliding in multiple nano-object contact results in lower friction forces in all conditions compared to sliding on the bare silicon substrate as discussed by *Maharaj* et al. [38.224]. For Au nano-objects that were already suspended in deionized (DI) water, experiments were not carried out in dodecane and glycerol. In general, the friction forces were lower for sliding in submerged-in-liquid environments as compared to sliding in dry environments for nano-object-coated and uncoated surfaces.

Sliding on multiple asperities on nanopatterned surfaces resulted in lower friction. With the immobile asperities, friction reduction occurs as a result of the reduced contact area. For Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs, the friction force reduction can be attributed to nano-object mobility in addition to the reduced contact area. The resulting friction reduction mechanism can be attributed to the reduced contact

area, the sliding over deformed nano-objects, and the individual nano-objects being dragged with the glass sphere as shown in Fig. 38.37b,c. It is also possible that as the glass sphere encounters a greater number of nano-objects, the contact pressure is reduced, allowing undeformed nano-objects to roll between the surfaces as shown in Fig. 38.37a. Exfoliation and tribofilm formation as shown in Fig. 38.37d,e are believed to be possible in dry environments for MoS_2 and WS_2 MWNTs as there is a general trend towards lower friction compared to Au 90 NPs. However this is not very significant due to the limited number of MWNTs that would be able to fit under the 50 μm sphere. In dry environments, the coefficient of friction data for the CNHs and the MoS_2 and WS_2 MWNTs are generally lower than the Au 90 NPs. As previously discussed, with the Au 90 NPs there is a higher meniscus force contribution to friction compared to the CNHs and MoS_2 and WS_2 MWNTs. With the CNHs, the reduced contact area due to the roughness of the nanohorns and low meniscus force contribution are responsible for the low friction forces [38.224].

For nano-objects submerged in liquid environments, the coefficients of friction were lower compared to dry. For water, as the meniscus forces are eliminated, the coefficient of friction data shows very little difference for Au 90 NPs and MoS_2 and WS_2 MWNTs. The CNHs have the lowest coefficient of friction in all liquid environments due to the nanohorn roughness contributing to lower real area of contact. Similar to single nano-object contact, the highest friction force and coefficient of friction occurs in glycerol, which has the highest viscosity and results in a greater contribution from viscous drag force followed by dodecane. The elimination of the meniscus forces, coupled with sliding on a surface of low shear strength results in lower coefficients of friction [38.234].

The nano-objects were also investigated in wear conditions to determine their effectiveness in protecting the underlying surface, which is a crucial feature for a potential lubricant. Figure 38.38 summarizes the wear data for sliding on Si, and Si coated with Au 90 NPs, CNHs, MoS_2 and WS_2 MWNTs, for 1, 10, and 100 cycles at a normal load of 20 μN under dry environments. A 20 μm × 20 μm area is imaged to show wear scars, which are created over a 10 μm × 10 μm area. Topography maps along with corresponding height profiles are also shown.

The surface of the uncoated silicon, after one cycle, appears rough during the early stages of wear

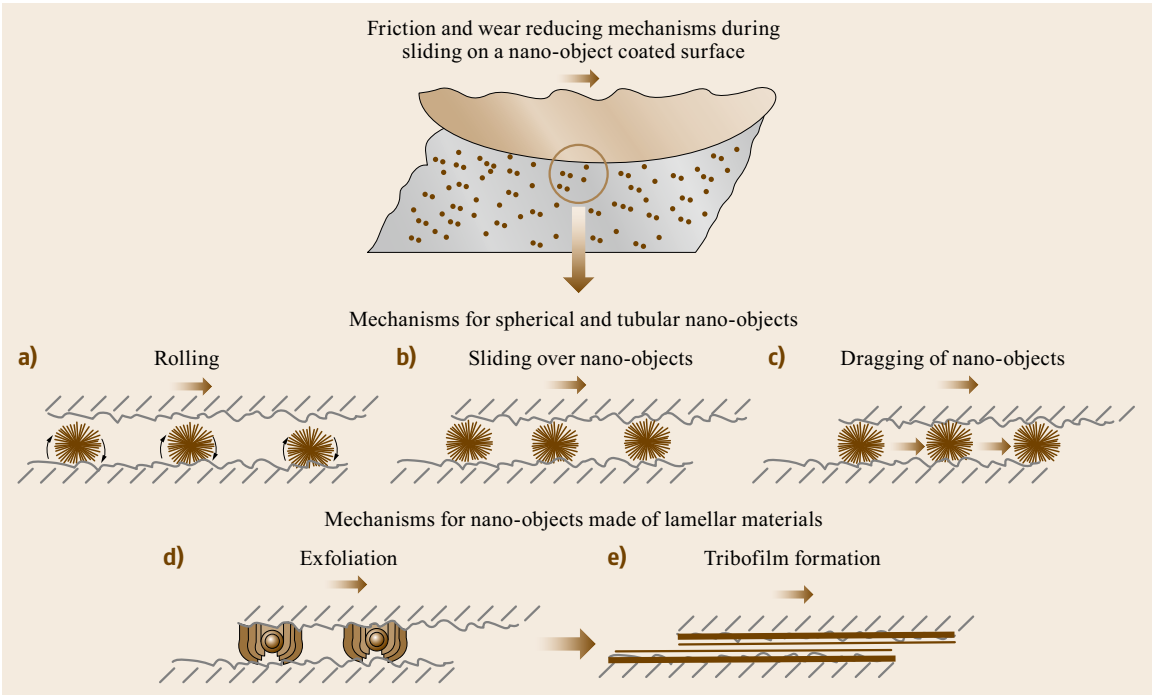


Fig. 38.37a–e Schematic of friction- and wear-reducing mechanism of nano-object-coated surface by **(a)** rolling, **(b)** sliding over nano-objects, **(c)** dragging of nano-objects, **(d)** exfoliation, and **(e)** tribofilm formation. CNHs are used as an example for mechanisms **(a)** through **(c)** and WS_2 and MoS_2 MWNTs are used as an example for mechanisms **(d)** through **(e)** (after [38.224])

as it becomes damaged. For the Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs coated surfaces, the nano-objects are just beginning to be pushed out of the wear area after one cycle. After 10 cycles, the topography profile of the uncoated silicon surface appears smoother than after one cycle. This is due to polishing of the surface. The nano-objects are pushed out of the wear area and agglomerate around the edges of the wear scar. After 100 cycles, for the uncoated silicon substrate, a small amount of material is removed, which gives a wear depth of approximately 0.25 nm, as seen in the height profile. For the coated surfaces, a few nano-objects remain in the wear area for MoS_2 and WS_2 MWNTs while they are completely removed for CNHs and Au 90 NPs.

Figure 38.39 shows SEM micrographs of the wear scars in dry environments after 100 cycles (first row). Vertical arrows point to magnified micrographs of the areas within the squares that are shown in the sec-

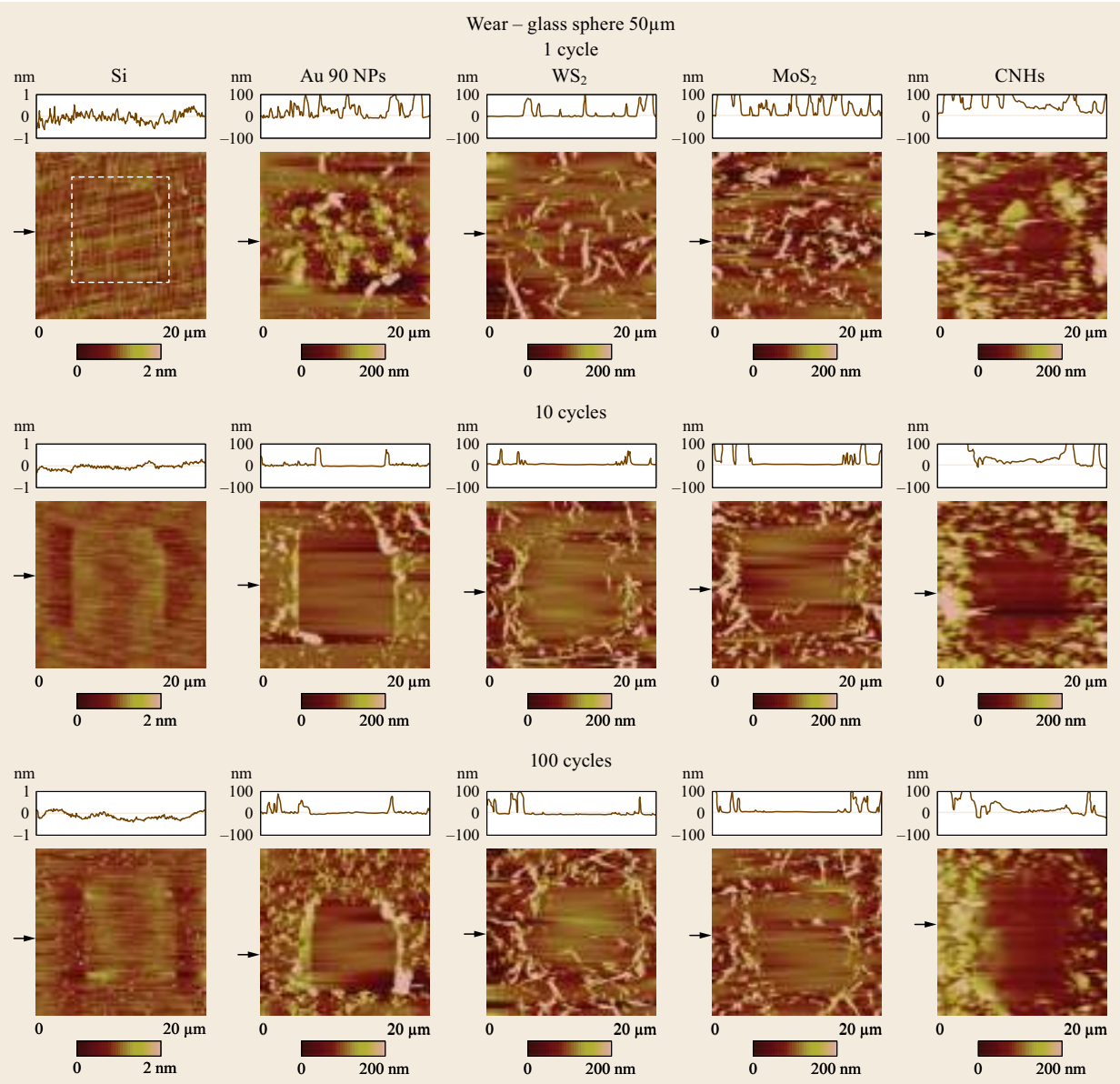


Fig. 38.38 Topography maps and 2-D profiles, at sections shown by the arrows, after sliding at one, 10 and 100 cycles with a normal load of 20 µN on Si and Si coated with Au 90 NPs, CNHs, and MoS_2 and WS_2 nanotubes (after [38.224])

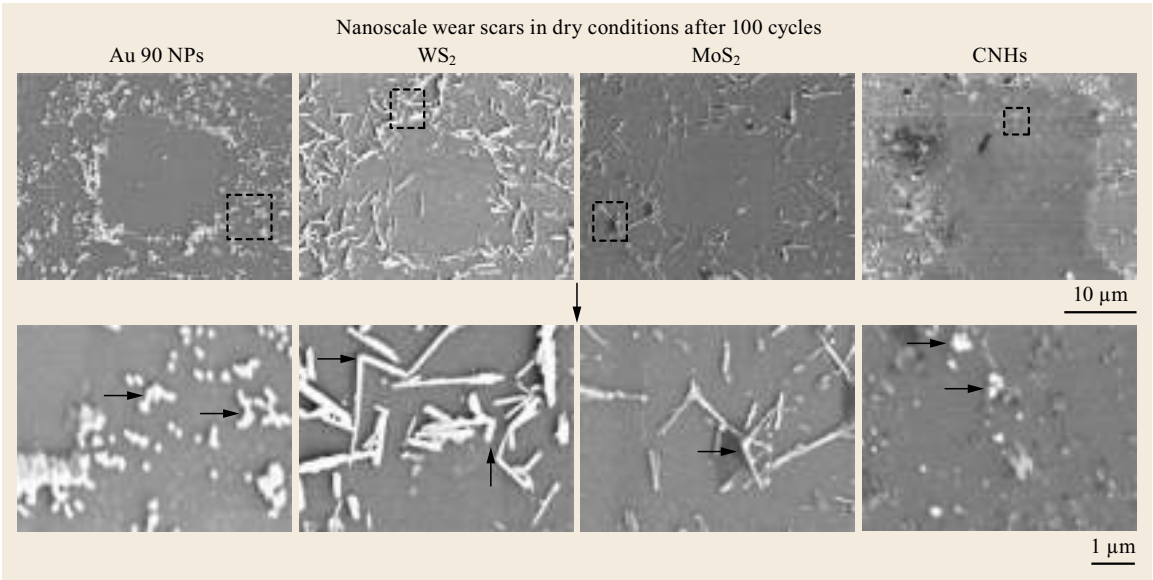


Fig. 38.39 SEM micrographs of nanoscale wear scars for Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs in dry conditions after 100 cycles at a normal load of 20 μN (*first row*). The nano-objects within the *squares* are magnified and the *vertical arrows* point to those micrographs in the *second row*. The magnified micrographs show agglomerated Au 90 NPs and CNHs, and broken MoS_2 and WS_2 MWNTs as indicated by *arrows* within the micrographs (after [38.224])

ond row. Agglomerated Au 90 NPs, CNHs, and broken MoS_2 and WS_2 MWNTs are pointed out by arrows within each micrograph. Agglomeration and breaking of the nano-objects can occur during the wear process due to the high loads being applied. In sample preparation of MoS_2 and WS_2 MWNTs, broken nanotubes also can result from sonication as they collide with one another. Since there is no evidence for crushed nanotubes from the SEM micrographs, it is likely that exfoliation of MoS_2 and WS_2 MWNTs would be limited. Protection of the substrate from wear due to tribofilm formation is not expected to be significant.

After one cycle, the damage of the silicon surface is expected to be less than that of an initially uncoated substrate since the Au 90 NPs, CNHs and MoS_2 and WS_2 MWNTs remain in the wear area. In addition to the reduced contact area provided by the nano-objects it is believed that they roll and slide between the glass sphere and the substrate, which promotes facile shearing of the two surfaces, as shown in Fig. 38.37a–c. Further reductions in wear are expected for the CNHs due to roughness, which results in reduced real area of contact. It is expected that the greatest wear occurs on the bare silicon substrate and the presence of nano-objects provides protection of the substrate with the least wear occurring with the CNHs after 100 cycles.

Macroscale Friction and Wear

In macroscale friction and wear tests, a ball on flat tribometer was used. A sapphire ball of 1.5 mm radius was slid over nano-object-coated and uncoated silicon substrates under a normal load of 200 mN for 500 cycles with a stroke length of 10 mm. Figure 38.40 shows optical micrographs of wear scars in dry and liquid environments with and without nano-objects. The widths of the wear scars shown are smaller for sliding in liquid environments compared to sliding in dry environments, as the amount of wear is less. The presence of nano-objects in dry and liquid environments reduces the amount of wear compared to cases without nano-objects.

In liquid environments, as the viscosity of the fluid increases, the width of the wear scar decreases, with the greatest wear protection occurring in glycerol. This is due to the ability of the viscous fluid to support the normal load and not be squeezed out of the contact area.

Optical and SEM micrographs of the wear scars for all cases are shown in Fig. 38.41. Nano-objects can be observed within the wear scars for all cases, as seen in Fig. 38.41a. The presence of the nano-objects after the wear test further confirms that reduction in wear and size of the wear scars are due to addition of the nano-objects. Figure 38.41b shows magnified SEM micrographs of wear scars for Au 90 NPs, CNHs, and

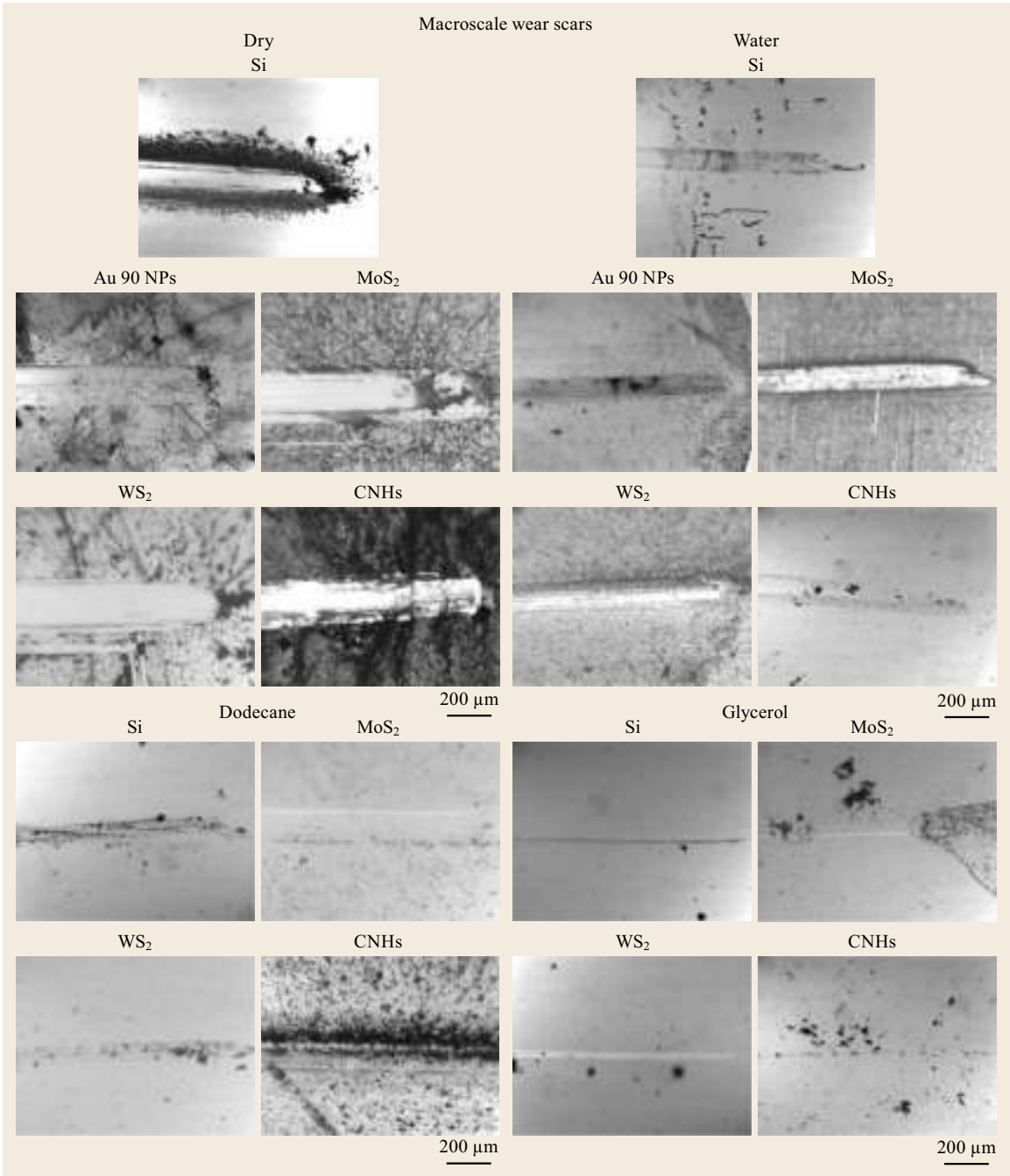


Fig. 38.40 Optical micrographs of macroscale wear scars taken after 500 cycles at a normal load of 200 mN, with 1.5 mm radius sapphire ball sliding in dry, water, dodecane and glycerol conditions with and without Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs (after [38.224]). The lowest wear occurs in glycerol with CNHs

MoS_2 and WS_2 MWNTs in dry environments. Agglomerations of nanoparticles within the squares can be observed in the case of the Au 90 NP and CNH wear scar.

Agglomeration occurs during the wear process as nanoparticles are pressed together and are attracted to each other due to van der Waals and at-

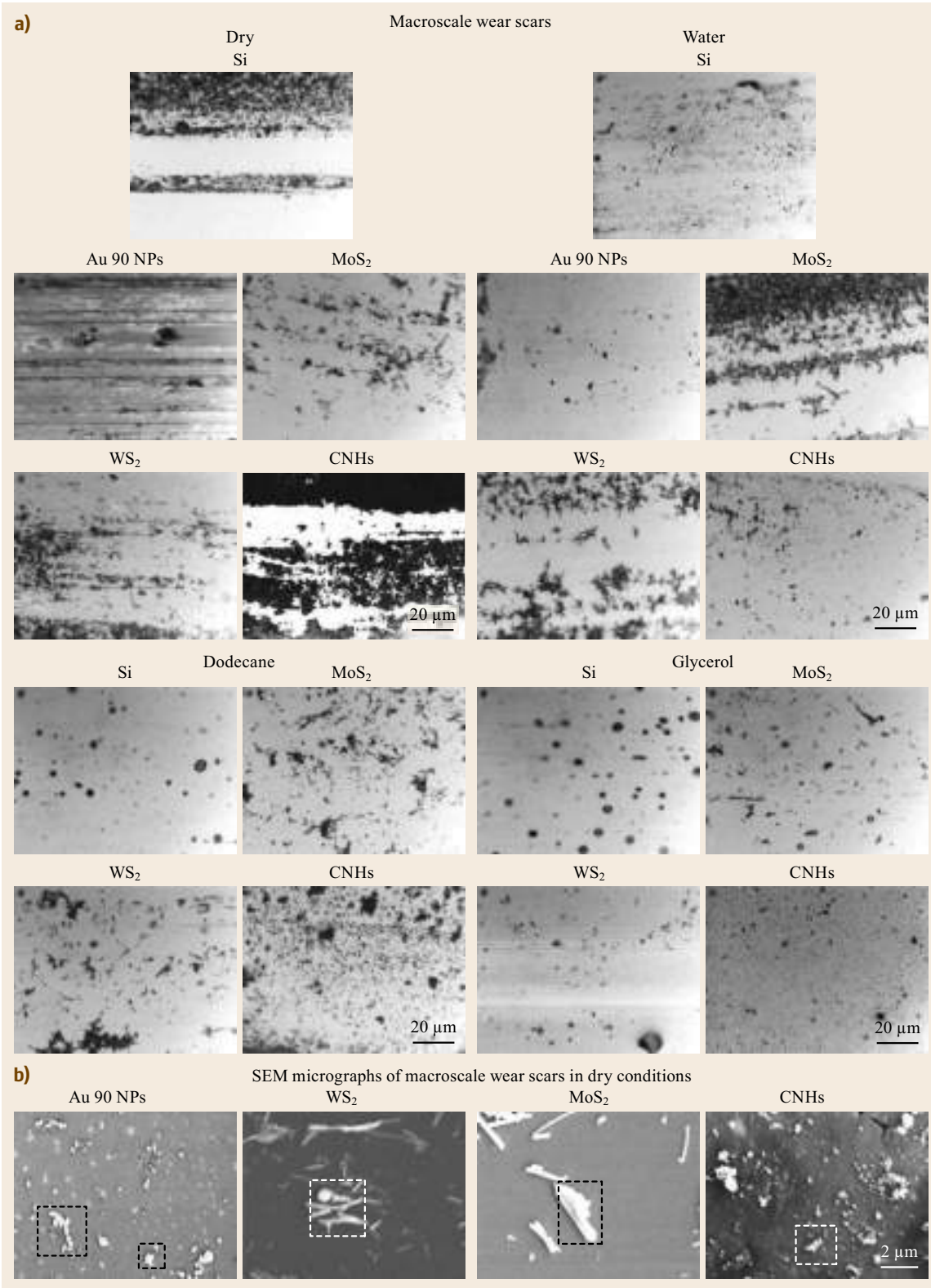

a)
Macroscale wear scars
Dry
Si
Water
Si
Au 90 NPs
MoS2
Au 90 NPs
MoS2
WS2
CNHs
WS2
CNHs
20 µm
20 µm
Si
Dodecane
MoS2
Si
Glycerol
MoS2
WS2
CNHs
WS2
CNHs
20 µm
20 µm
b)
SEM micrographs of macroscale wear scars in dry conditions
Au 90 NPs
WS2
MoS2
CNHs
2 µm

Fig. 38.41 **(a)** Optical micrographs inside the wear scar on the macroscale taken after 500 cycles at a normal load of 200 mN, in dry, water, dodecane, and glycerol conditions with and without Au 90 NPs, CNHs, and MoS_2 and WS_2 MWNTs, and **(b)** SEM micrographs of the wear scar for agglomerated Au 90 NPs and CNHs, along with crushed or broken MoS_2 and WS_2 MWNTs in dry environments shown within the *squares* (after [38.224]) ◄

tractive meniscus forces. For both MoS_2 and WS_2 wear scars, the SEM micrographs show crushed and broken nanotubes within the squares compared to those in Fig. 38.34a,b, which depicts undeformed nanotubes. It is believed that in the process of being crushed, the outer layers of the nanotubes break apart leading to exfoliation of the nanotubes and tribofilm formation, as discussed by *Maharaj* et al. [38.224].

In liquid environments, the results are similar to the nanoscale, where the coefficients of friction are lower than in dry environments in general, as reported by *Maharaj* et al. [38.224]. This is due to the presence of a surface of low shear strength for sliding. For CNHs submerged in water, the coefficient of friction is the lowest of all four nano-objects due to the reduced contact area provided by the roughness of the nanohorns. For MoS_2 and WS_2 MWNTs after 200 cycles, the coefficient of friction increases and becomes greater than that of the other nano-objects, and reaches that of bare silicon. The increase in friction is due to the oxidation of the MoS_2 and WS_2 MWNTs, as previously discussed. For sliding in dodecane and glycerol, the coefficient of friction is lower sliding on a nano-object-coated surface compared to bare silicon. The coefficient of friction is lower for the CNHs compared to the MWNTs, similar to the submerged-in-water case. After approximately 400 cycles for sliding on nanoparticles and nanotubes in dodecane, the coefficient of friction begins to increase as wear of the substrate takes place. In the case of glycerol there is also an increase in coefficient of friction for the nanotubes after 400 cycles. However, the coefficient remains steady for the CNHs. This is due to the combined effect of the reduced real area of contact for CNHs with the ability of glycerol to better support the load due to its high viscosity, which reduces the wear. The lowest coefficient of friction and least wear takes place in glycerol with CNHs.

38.5.3 Summary

Au NPs, MoS_2 and WS_2 MWNTs, and CNHs were investigated to understand friction and wear mechanisms of nano-objects and their effect on friction and wear in dry and liquid environments in water, dodecane, and glycerol. Studies were conducted in both single and multiple nano-object contacts with the aid of an AFM. In macroscale studies, a ball on flat tribometer was used.

For single nano-object contact, the lowest friction forces occur with the CNHs compared to Au NPs and MoS_2 and WS_2 MWNTs due to the reduced contact area provided by the roughness of the nanohorns and the low meniscus force contribution of the dry environments. Lower friction forces occur in submerged-in-liquid environments compared to dry environments due to elimination of meniscus force and sliding on a low shear-strength surface among the various liquids. The highest friction forces occur in glycerol due to high viscosity.

For multiple nano-object contact, sliding over nano-objects decreased the coefficients of friction compared to sliding on the bare silicon substrate due to reduced contact area and rolling and sliding among the various nano-objects. The lowest coefficient of friction occurs with the CNHs compared to the other nano-objects due to the mechanisms mentioned for single nano-object contact in dry environments. In nanoscale wear experiments in dry environments, the addition of nano-objects reduces the real area of contact and reduces the wear of the substrate through rolling and sliding of the nano-objects.

In macroscale studies, sliding over the various nano-objects reduced friction and wear due to the mechanisms mentioned in multiple nano-object contact on the nanoscale. Additionally, in dry environments, because of exfoliation and tribofilm formation, sliding on the MWNTs decreased the coefficient of friction and reduced wear. The lowest coefficient of friction and least wear in all cases occur with CNHs due to the reduced contact area. The coefficients of friction were also lower in submerged-in-liquid environments compared to dry conditions due to mechanisms mentioned on the nanoscale. The more viscous liquids results in better wear protection and lower coefficient of friction due to their ability to support the load and not be easily squeezed out of the contact area.

38.6 Nanotribological Studies of Biological Molecules on Polystyrene and Silicon Surfaces and Coated Polymer Surfaces

Protein layers are routinely deployed on bioMEMS/NEMS as a functional layer allowing for specific molecular recognition, binding properties, or to facilitate biocompatibility. Examples of these include sensors, immunoisolation capsules, and drug-delivery systems [38.1, 60, 235]. In addition, coating material surfaces with a protein layer facilitates biocompatibility along with implant functionality [38.149, 236]. Both the implanted biomaterial and the layer of protein will experience a range of physical and chemical assaults over the life of the device [38.237]. These include degradation by serum or tissue proteases, biochemical dissociation from surfaces, chemical modification, and abrasion by tissue elements [38.238]. In addition, an implanted biomaterial will experience micromotion over time, leading to wear and degradation of the protein layer [38.75, 93]. These can significantly affect and degrade the performance of the biomaterial, eventually leading to failure [38.237]. Optimizing the adhesion, friction, and wear on the adsorbed protein layer therefore is crucial to prevent device failure.

Protein adsorption, conformation, and adhesion also are of interest in cellular adhesion and growth. *Schricker* et al. [38.239, 240] and *Bhushan* and *Schricker* [38.241] studied protein adhesion on block copolymer-based biomaterials used as a scaffold for dental bone generation.

Bioadhesion refers to the adherence of biological material to a natural or synthetic material. The factors that control bioadhesion or cell adhesion and wear include surface morphology, chemical interactions, physiological factors, physical-mechanical interactions, and fluid environment [38.242, 243]. Bioadhesion for a given biological material to be bonded to an engineering surface and its wear properties need to be optimized.

In some applications, surfaces with low bioadhesion are required. Fluorinated self-assembled layers have been used to reduce bioadhesion of polymer surfaces.

38.6.1 Nanoscale Adhesion, Friction, and Wear of Protein Layers on Polystyrene Surfaces

Bhushan and *Utter* [38.242] studied nanoscale adhesion, friction, and wear behavior of four common proteins – bovine serum albumin (BSA), collagen, fibronectin, and streptavidin (STA) – adsorbed to a spin-coated polystyrene surface using atomic force microscopy. Bovine serum albumin is a homologue of human serum albumin and its primary physiological functions include helping to maintain blood oncotic pressure, the binding and transport of important ligands, and several antioxidant functions [38.244]. Type I collagen is the most abundant protein in the human body and is the primary load-bearing component of connective tissue [38.245]. Fibronectin is a major extracellular component required for clot formation and wound healing found throughout the body [38.246]. Finally, STA is used extensively in the biotechnology field for protein purification, biosensors, and the immobilization of biomolecules due to its interaction with biotin [38.247]. Figure 38.42 shows protein structures. Bovine serum albumin, fibronectin and streptavidin are all globular proteins [38.242]. Collagen is the only protein used in this experiment with a linear structure. Table 38.9 compares and contrasts the masses and isoelectric points of the four proteins used [38.242]. The isoelectric point is

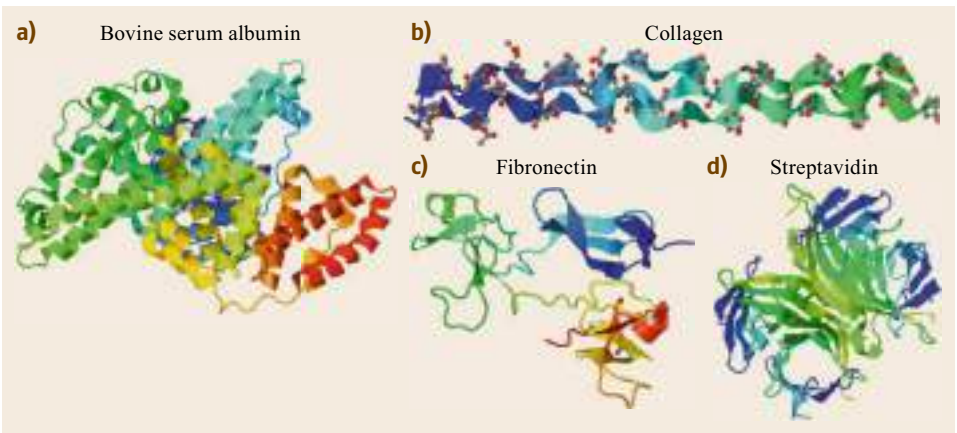


Fig. 38.42a–d Images showing the physical structure of the bovine serum albumin **(a)**, collagen **(b)**, fibronectin **(c)**, and streptavidin **(d)**. Proteins are not shown to scale relative to one another. All figures based on the protein data bank's files (after [38.242])

Table 38.9 Summary of the different protein masses and their isoelectric point [38.242])

Protein	Molecular weight (kDa)	Isoelectric point
Bovine serum albumin	≈ 67[a]	4.7[e]
Collagen	≈ 300[b]	4.7[f]
Fibronectin	≈ 500[c]	5[g]
Streptavidin	≈ 60[d]	5.5[d]

[a] [38.248]; [b] [38.245]; [c] [38.246]; [d] [38.247]; [e] [38.249]; [f] [38.250]; [g] [38.251]

the pH at which the protein will be electrically neutral in solution. Changing the pH of the solution above or below the isoelectric point will change the net charge on each protein. All of the proteins used in this experiment have an acidic isoelectric point giving them a net negative charge in DI water. STA and BSA have masses of 60 and 67 kDa respectively, with fibronectin and collagen being an order of magnitude more massive.

Polystyrene was selected as the substrate surface due to its extensive use in the biotechnology field. 50 µl of protein solution were added to the surface, and the samples were allowed to sit overnight. Before beginning any experiment, the samples were rinsed with DI water to remove the electrolytes and any unbound protein. Phosphate buffered saline (PBS) (pH 7.4) and the two remaining pH levels were created by combining monosodium phosphate (Aldrich) and disodium phosphate (Aldrich) in DI water to create the proper phosphate buffered solution.

Nanoscale adhesion, friction, and wear measurements were made using an AFM in DI and PBS (pH 5.0, 7.4 and 9.0). Adhesion and friction were investigated for all four proteins by varying the concentration of the protein solution from 1 to 100 µg/ml in DI water and PBS (pH 7.4). Nanoscale wear was also studied at a protein concentration of 100 µg/ml in PBS (pH 7.4). In addition, adhesion and friction were investigated by varying the pH and temperature of the liquid medium for both BSA and STA at 100 µg/ml.

Nanoscale Adhesion Measurements

First, the effect of protein concentration on the adhesive force was investigated for all the proteins. Figure 38.43a shows the adhesive force data between the bare silicon nitride AFM tip and each of the various proteins. The adhesive force increases between the protein film and the bare silicon nitride AFM tip as the concentration of protein is increased [38.242]. This increase in adhesive force is believed in part to be due to multiple interactions occurring between the protein film and the AFM tip [38.93]. This is also consistent with the observation that increasing the concentration of protein solution increases the density of the immobilized protein on the polystyrene surface. This increase may also be due to the more hydrophobic domains of the protein interacting with the polystyrene surface, leaving the hydrophilic domains to be exposed to the AFM tip. Other studies have shown that near-maximum binding will occur at roughly two hours, and that more protein will directly adsorb depending on solution concentration [38.252, 253]. It has also been reported that above a concentration of 10 µg/ml, certain proteins will begin to form higher-order structures not present at lower concentrations [38.93].

Figure 38.44 lists the basic interactions that may occur during protein adsorption and is only meant to show some of the interactions that can occur [38.254]. The

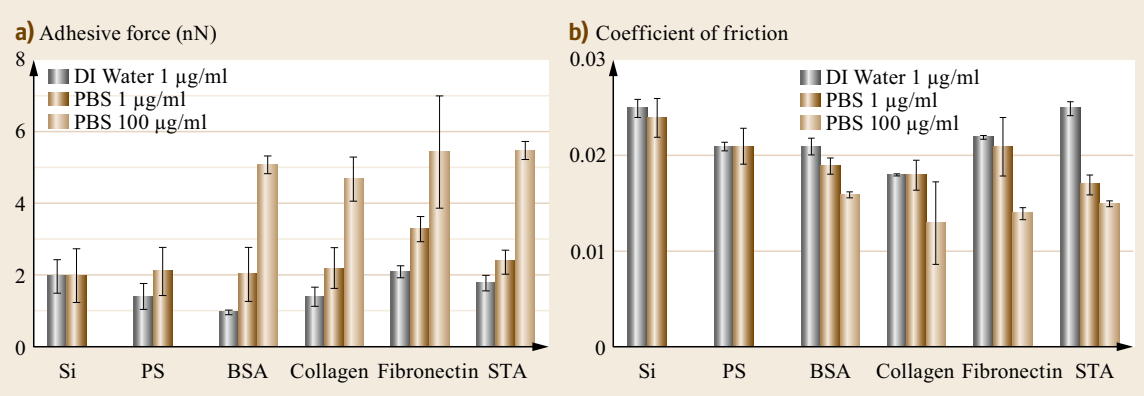


Fig. 38.43a,b Effect of protein concentration on **(a)** the adhesive force and **(b)** the coefficient of friction between the protein film and bare AFM tip. All experiments conducted at ambient temperature 22±1 °C unless otherwise noted. DI water (pH 7.0) and PBS (pH 7.4) were used for all experiments unless otherwise noted. Error bars represent ±1 standard error (after [38.242])

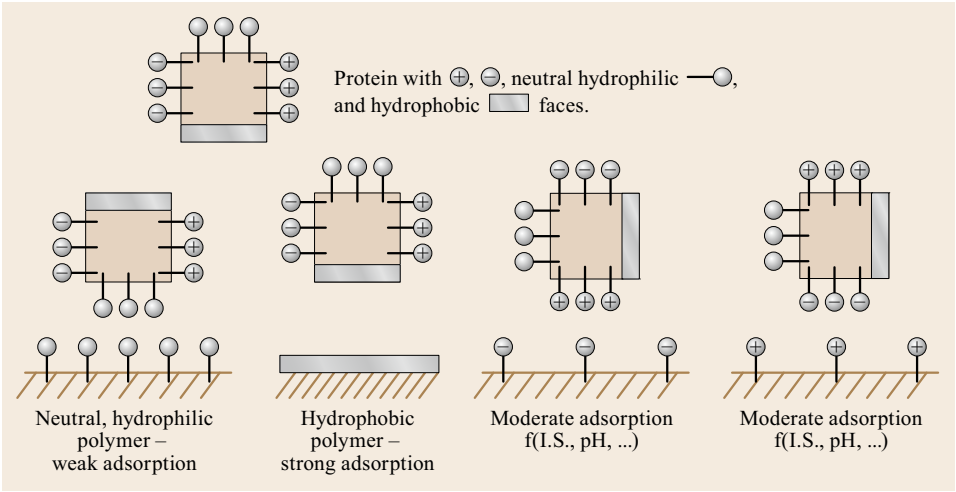


Fig. 38.44 Schematic showing the different protein domains and the potential interactions with the substrate surface (after [38.254])

strongest interaction occurs between the hydrophobic domains of the protein and a hydrophobic surface. Next, moderate adsorption occurs between charged protein domains and charged surfaces. Finally, weak adsorption occurs between hydrophilic protein domains and hydrophilic surfaces. Protein adsorption may involve several of these interactions occurring on the same surface and with the same protein.

Given that polystyrene is a hydrophobic surface, it is to be expected that the hydrophobic domains of the proteins interact with the surface during adsorption. The strength of the hydrophobic interaction is strong enough to induce a conformational change in the protein exposing the inner hydrophobic domains to the surface. This allows for the hydrophilic and charged domains to interact with the solution. It is also well known that by varying the pH of the solution, proteins can be made to denature from their native structure, thus allowing the inner hydrophobic domains to interact with the polystyrene surface. Previous studies have shown that chemically modifying the polystyrene surface has an effect on the amount of adsorbed protein, suggesting that the four proteins used in this study have a higher affinity for hydrophobic surfaces [38.252, 253]. This hydrophobic interaction is believed to be the dominant mechanism for protein adsorption to polystyrene. However, we do not have the data to show the orientation or packing density of the protein films [38.252]. The data from this experiment demonstrates no clear trend to distinguish the proteins from one another. However, what is clear for all of the selected proteins is that an increase in protein concentration leads to an increase in adhesive force between the AFM tip and the protein layer.

Next, the effect of pH on BSA and STA at 100 μg/ml concentration was investigated. Figure 38.45a shows the adhesive force data as a function of pH at 22 °C [38.242]. The general trend is that the adhesive force increases for each protein as the pH of the solution becomes more basic. Both proteins exhibit higher adhesive forces than the polystyrene control surface over the range of pH values. As the protein layer is exposed to a more basic solution, the surface will have a much larger net negative charge. Given that silicon nitride has an isoelectric point of pH 6–7, this may lead to a stronger interaction between the surface and the silicon nitride AFM tip [38.255]. For lower pH values (< pH 7.4), STA displays higher adhesive forces than BSA but this trend does not hold for the most basic solution.

The effect of temperature over the range of pH values also was studied for BSA and STA at 100 μg/ml concentration. Figure 38.45b shows the adhesive force data as a function of pH at 37 °C [38.242]. Adhesive forces for both proteins are much lower than at 22 °C and are not much different than the polystyrene control surface. This decrease in adhesive force at 37 °C may be due in part to the fact that more protein–protein and protein–surface interactions are occurring that result in a lower adhesive force with the AFM tip. The decrease in adhesive force could also be due to a temperature-related conformation change in the protein, which then interacts more weakly with the AFM tip. However, a slight increase occurs for all samples as the solution becomes more basic. This trend is consistent with the data that was obtained at 22 °C. As before, no clear trend is present to distinguish BSA from STA. This may be in part due to their similar protein structure, isoelectric point and mass.

Nanoscale Friction Measurements

The coefficient of friction was measured for all proteins as a function of protein concentration. Figure 38.43b

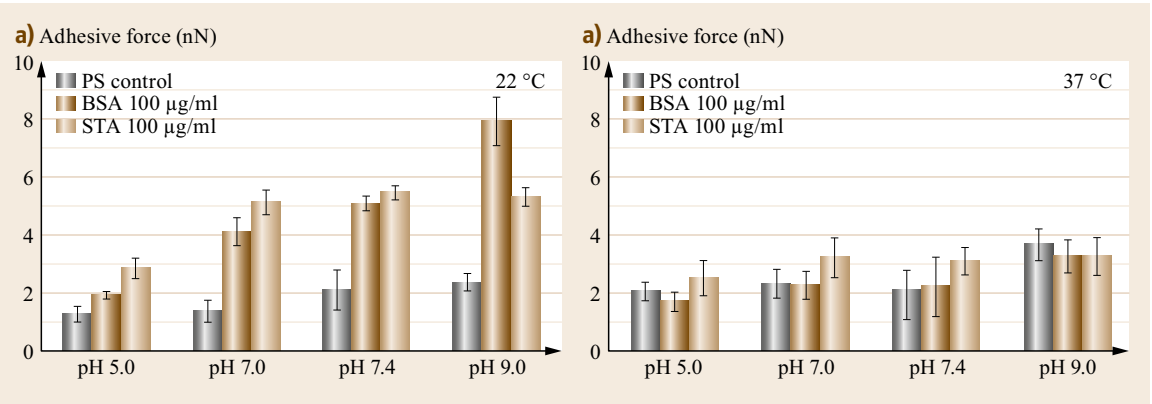


Fig. 38.45a,b Effect of pH on adhesive forces **(a)** at 22 °C and **(b)** 37 °C. Error bars represent ±1 standard error (after [38.242])

shows the coefficient of friction between the bare silicon nitride AFM tip and the various proteins [38.242]. The friction measurements that were taken show a decreasing coefficient of friction as the protein concentration is increased from 1 to 100 µg/ml. This decrease is believed to be due to the proteins acting as a lubricant film between the polystyrene surface and the silicon nitride AFM tip. At a higher protein concentration, a continuous lubrication film is formed due to a more uniform protein layer being formed on the substrate surface [38.93]. Like the adhesion data, there is no clear trend to distinguish the proteins from one another. However, the general trend is that increasing the concentration of these proteins leads to a decrease in the coefficient of friction for each protein tested.

Next, the coefficient of friction was measured for BSA and STA at 100 µg/ml concentration as a function of pH at 22 °C. Figure 38.46a shows the coefficient of friction between the bare silicon nitride AFM tip and two different protein-coated surfaces [38.242]. Both proteins exhibit lower friction than the polystyrene control, but the data appears to show that the coefficient of friction is fairly constant for all of the samples as the pH of the solution is altered. Also, no clear trend is present to distinguish BSA from STA when the pH varied, but the data from this experiment is in good agreement with what was found earlier by varying the protein concentration.

The effect of temperature also was investigated by heating the sample and solution to 37 °C. Figure 38.46b shows the coefficient of friction between the bare silicon nitride AFM tip and two different protein-coated surfaces [38.242]. At an elevated temperature the coefficient of friction increases slightly or stays the same relative to the 22 °C data. Again, the coefficient of friction is lower for the protein-coated samples relative to the polystyrene controls, but the coefficient of friction does not significantly change from one pH level to the next.

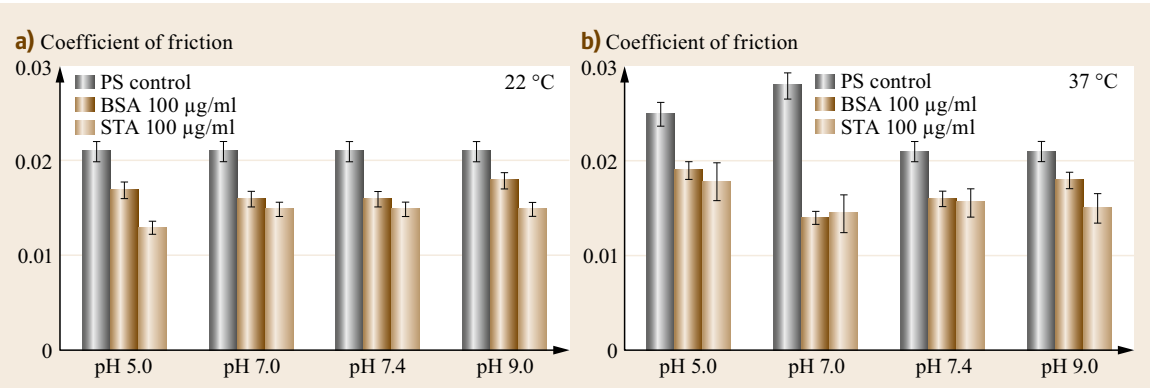


Fig. 38.46a,b Effect of pH on the coefficient of friction **(a)** at 22 °C and **(b)** 37 °C. Error bars represent ±1 standard error (after [38.242])

From these experiments, the coefficient of friction appears to be affected significantly by the concentration of the adsorbed protein film and not by the pH or temperature of the surrounding medium. As reported by *Bhushan* et al. [38.93] and *Fan* et al. [38.256], and confirmed with these experiments, protein concentration appears to be the primary factor in the formation of a continuous lubrication film. Solution pH and temperature appear to have secondary effects on the friction force experienced by the AFM tip.

Nanoscale Wear Tests

Figure 38.47 shows the wear images for each protein and a bare polystyrene control surface [38.242]. Each figure shows the height image and cross-sectional profile of a wear mark after conducting the wear scans. For these tests, the protein concentration was kept constant at 100 μg/ml with the normal load increasing from 3–9 nN. These tests show that the depth of wear increases with an increasing normal load for all of the samples imaged.

Several key differences become apparent on comparing the wear marks between the different protein-coated surfaces. The surface coated with BSA shows the most significant wear relative to all of the other samples. This may be in part due to the smaller mass of BSA and lower isoelectric point. At a pH of 7.4, BSA will have a net negative charge, which will interact more strongly with the AFM tip. When the AFM tip comes into contact with the protein film, the net negative charge and small mass allows for more wear to occur relative to the other samples. Streptavidin has approximately the same mass as BSA, but it does not exhibit the degree of wear that BSA undergoes. Therefore, protein mass appears to have a secondary effect on wear over the range of normal forces tested. As expected, the wear loads are of the same order of magnitude as the adhesive forces found in this experiment.

A trend becomes apparent by comparing the wear of the three globular proteins. As the protein isoelectric point becomes more basic and closer to the pH of saline (pH 7.4), wear marks become smaller. The most significant wear occurs with BSA, which has the lowest isoelectric point (pH 4.7). The surface coated with fibronectin (isoelectric pH 5) displays less wear than the BSA sample, but greater wear than the streptavidin (isoelectric pH 5.5) sample. The difference in wear marks for the globular proteins in part may be due to charge–charge repulsion that occurs when the pH of the solution is different than the isoelectric point of the protein.

However, the most interesting surface of all appears to be the one coated with collagen. This surface appears not to undergo any significant wear as compared to the other surfaces on the length scale in which we are interested. Collagen is the only protein tested that is linear in structure compared to the other proteins, and it has the ability to strongly self-associate [38.257]. Collagen also contains more hydrophobic amino acid groups than other proteins, which results in a stronger interaction with the polystyrene surface [38.258]. As the protein solution is allowed to sit overnight, slow rearrangements take place in the adsorbed collagen, creating a denser mesh with an increased resistance to wear [38.253]. Given that collagen is the primary load-bearing protein in the body and can be found in tissues in which significant mechanical wear can occur, it seems logical that it could be the most wear-resistant protein [38.245]. However, as shown by *Pamuła* et al. [38.253], and confirmed in the study being presented as well, the collagen surface will still undergo slight degradation at higher loads. Although fibronectin has the ability to self-associate, it does not appear to form the wear-resistant mesh that collagen does due to its globular structure.

Figure 38.48 shows a schematic of the hypothetical wear mechanism for the globular proteins at a higher protein concentration [38.242]. Due to the motion of the AFM tip during scanning, the protein molecules displace laterally and begin to interact with the remaining protein molecules. Protein aggregation leads to a pileup formation that keeps the molecules attached to the surface. The increased normal force causes more protein molecules to displace, leading to an increased degradation of the surface.

Summary

Bhushan and *Utter* [38.242] studied the nanoscale adhesion, friction and wear properties of four proteins adsorbed to a spin-coated polystyrene surface using an AFM in DI water and PBS by varying the concentration of the protein solution, pH, and temperature of the liquid medium. They found that under mild operating conditions, all of the proteins had similar adhesion and friction behavior, making a direct comparison between these proteins difficult. They showed that for all four proteins studied, the adhesive force between each protein and the AFM tip increased with increasing protein concentrations. It is believed that this increase in adhesive force in part is due to the formation of multiple interaction sites between the protein film and the AFM tip. Further, as the solution was altered from acidic to basic, adhesive forces also increased for BSA and STA, possibly in part due to the larger net charge of the protein interacting with the AFM tip. Adhesive forces then decreased when the temperature was increased to 37 °C. There was no clear trend present distinguishing BSA from STA, possibly because they have similar structure, isoelectric points, and mass.

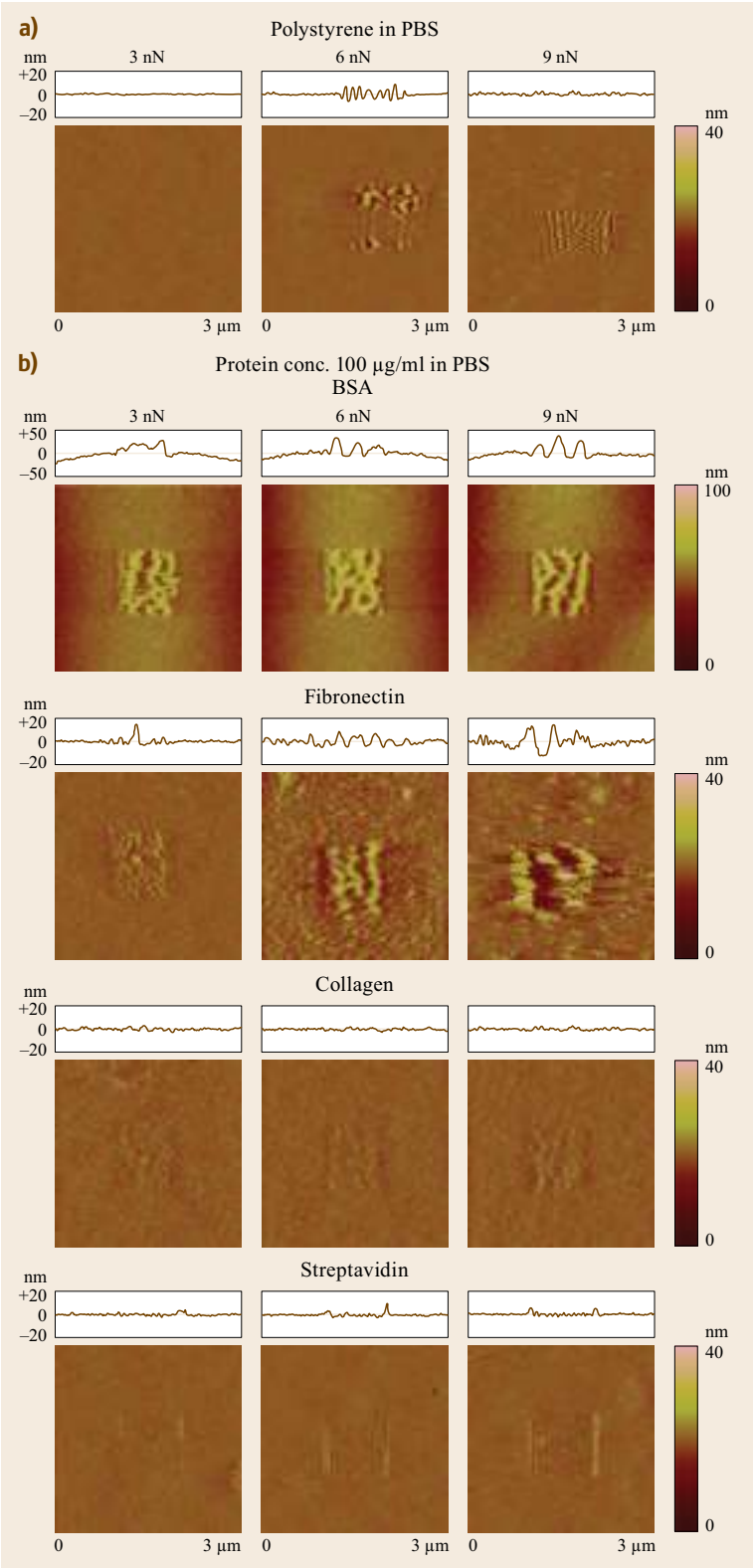


Fig. 38.47a,b AFM height images and corresponding cross-sectional profile of a wear mark in contact mode of **(a)** bare polystyrene and **(b)** four proteins at a concentration of 100 µg/ml in PBS. The normal loads used were 3, 6, and 9 nN. Reprinted from [38.242] with permission from Elsevier

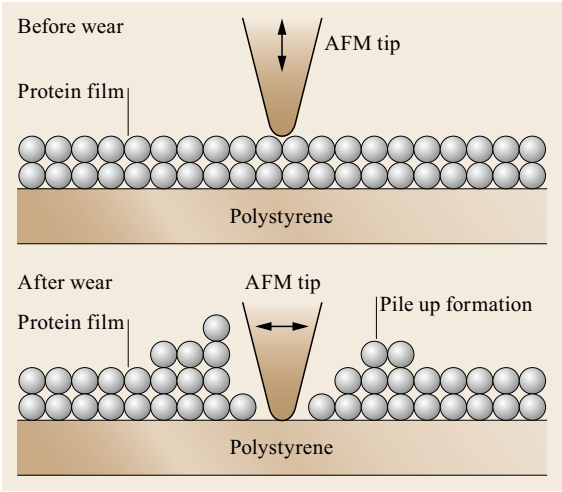


Fig. 38.48 Schematic of hypothetical wear mechanism of the adsorbed protein film at higher concentration (after [38.242])

It was also shown that for all proteins studied in PBS (pH 7.4), the coefficient of friction decreased as the protein concentration increased. It is believed that this is because the protein layer acts as a continuous lubrication film at higher protein concentrations. As compared to the polystyrene control surface, the coefficient of friction also decreased for BSA and STA when each pH level was tested at 22 °C. For each protein studied, the coefficient of friction remained relatively constant across the range of pH values. As compared to the polystyrene control surface, increasing the temperature to 37 °C kept the coefficient of friction relatively constant across the range of pH levels for BSA and STA. It appears that the existence of a lubrication film is the most important factor affecting the coefficient of friction. Solution pH and temperature appear only to have secondary effects on the coefficient of friction. There was no clear trend distinguishing BSA from STA when the pH and temperature of the solution are varied.

Wear is the distinguishing factor for the different proteins. Normal load had an effect on wear. As the normal force was increased, there was increased degradation for all of the samples imaged. For all of the proteins tested, the depth of wear marks increased with increased normal load. The globular proteins degraded in the order of their isoelectric points: bovine serum albumin > fibronectin > streptavidin > collagen. It is possible that collagen is more wear resistant because of its linear structure and ability to self-associate. For the normal forces tested, protein mass appears only to have a secondary effect on wear; the primary wear mechanism results from protein structure and isoelectric point.

38.6.2 Adhesion, Friction, and Wear of Biomolecules on Si-Based Surfaces

Proteins on silicon-based surfaces are of extreme importance in various applications, including silicon microimplants, various bioMEMS such as biosensors, and therapeutics. Silicon is a commonly used substrate in microimplants, but it can have undesired interactions with the human immune system. Therefore, to mimic a biological surface, protein coatings are used on silicon-based surfaces as passivation layers, so that these implants are compatible in the body and avoid rejection. Whether this surface treatment is applied to a large implant or a bioMEMS, the function of the protein passivation is obtained from the nanoscale three-dimensional (3-D) structural conformation of the protein. Proteins are also used in bioMEMS because of their function specificity. For biosensor applications, the extensive array of protein activities provides a rich supply of operations that may be performed at the nanoscale. Many antibodies (proteins) have an affinity to specific protein antigens. For example, pathogens (disease-causing agents, e.g., virus or bacteria) trigger the production of antigens that can be detected when bound to a specific antibody on the biosensor. The specific binding behavior of proteins that has been applied to laboratory assays may also be redesigned for in vivo use as sensing elements of a bioMEMS. The epitope-specific binding properties of proteins to various antigens are useful in therapeutics. Adhesion between the protein and substrate affects the reliability of an application. Among other things, morphology of the substrate affects the adhesion. Furthermore, for in vivo environments, the proteins on the biosensor surface should exhibit high wear resistance during direct contact with the tissue and circulatory blood flow without washing off.

Bhushan et al. [38.75] studied the step-by-step morphological changes and the adhesion of a model protein – streptavidin (STA) – on silicon-based surfaces (see also [38.148, 150, 259]). Figure 38.49a presents a flow chart showing the sequential modification of a silicon surface. In addition to physical adsorption, they also used nanopatterning and chemical linker methods to improve adhesion. A nanopatterned surface contains large edge surface area, leading to high surface energy, which results in high adhesion. In the chemical linker method, sulfo-*N*-hydroxysuccinimido-biotin (sulfo-NHS-biotin) was used as a cross-linker because the bonds between the STA and the biotin molecule are one of the strongest noncovalent bonds known (Fig. 38.49b). Sulfo-NHS-biotin was con-

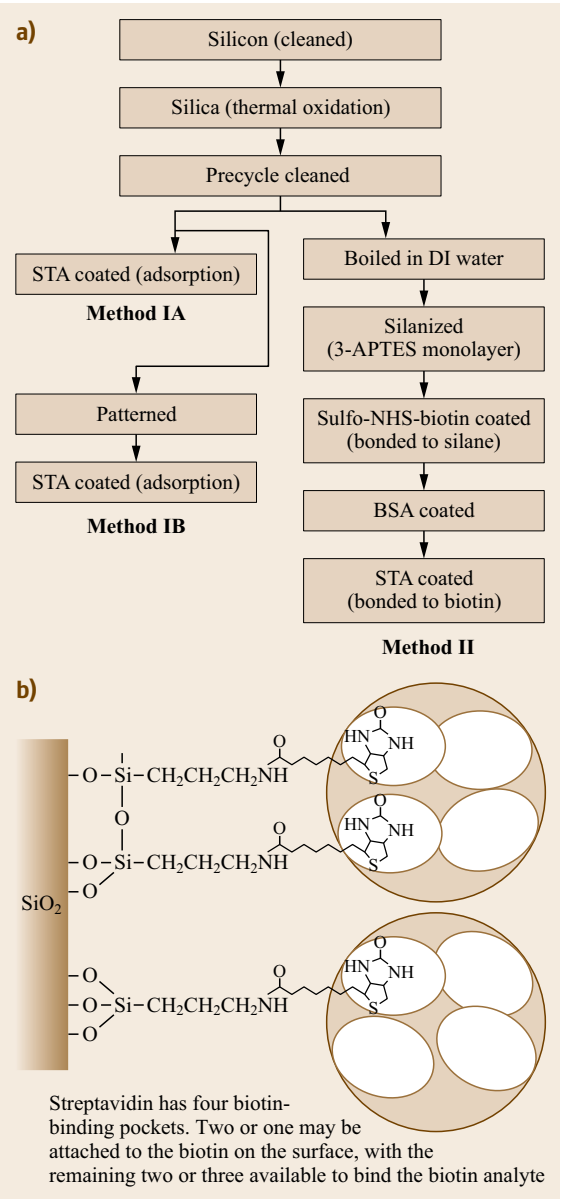


Fig. 38.49 **(a)** Flow chart showing the samples used and their preparation technique, and **(b)** a chemical structure showing streptavidin protein binding to the silica substrate by the chemical linker method

nected to the silica surface through a silane linker, 3-aminopropyltriethoxysilane (3-APTES). In order to make a bond between the silane linker and the silica surface, the silica surface was hydroxylated. Bovine serum albumin (BSA) was used before STA to block nonspecific binding sites of the STA protein with the silica surface. Figure 38.50 shows the step-by-step mor-

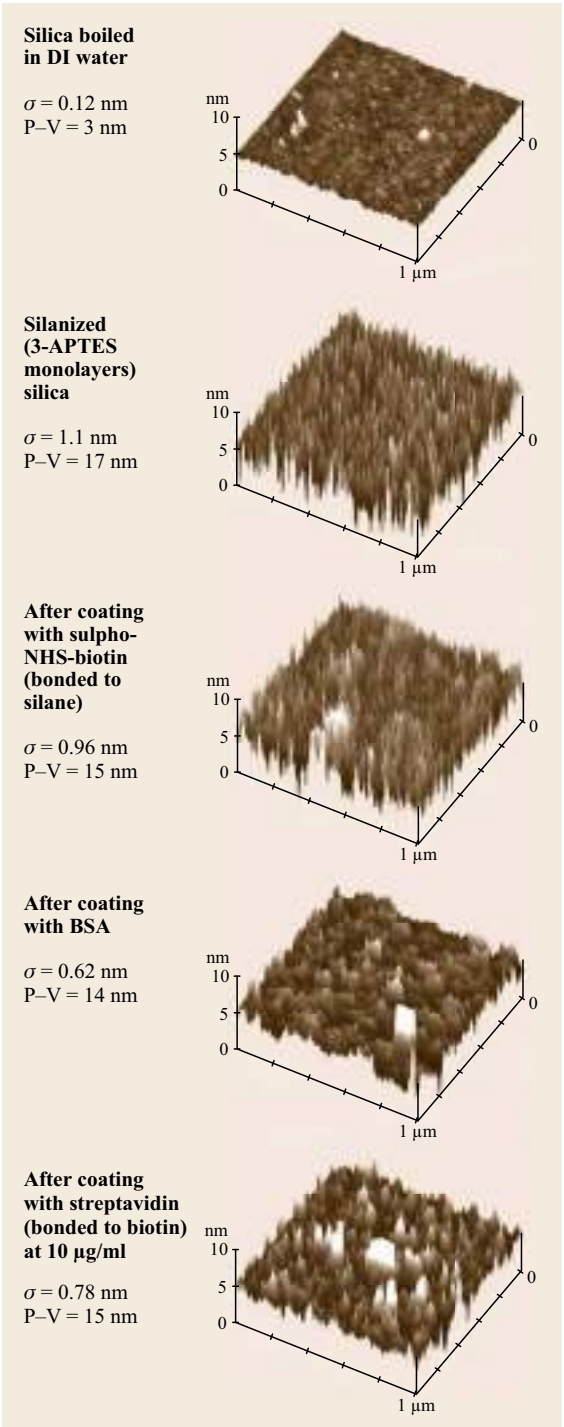


Fig. 38.50 Morphological changes in silica surface during functionalization of silica surface by chemical linker imaged in PBS. Streptavidin is covalently bonded at 10 µg/ml concentration (after [38.75])

phological changes in the silica surface during the deposition process using the chemical linker method. There is an increase in roughness (σ and peak-to-valley (P–V) distance) of the silica surface boiled in deionized (DI) water compared to the bare silica surface. After the silanization process, there are many free silane links on the surface, which cause higher roughness. Once biotin is coated on the silanized surface, the surface becomes smoother. Finally, after the deposition of STA, the surface shows large and small clumps. Presumably, the large clumps represent BSA, and the smaller ones represent STA. To measure adhesion between STA and the corresponding substrates, a STA-coated tip (or functionalized tip) was used and all measurements were made in PBS, a medium commonly used in protein analysis to simulate body fluid. Figure 38.51 shows the adhesion values of various surfaces. The adhesion value between biotin and STA was higher than that for other samples, which is expected. Edges of patterned silica also exhibited high adhesion values. It appears that both nanopatterned surfaces and chemical linker method increase adhesion with STA.

Bhushan et al. [38.93] studied friction and wear of STA deposited by physical adsorption and the chemical linker method (see also [38.150]). Figure 38.52 shows the coefficient of friction between the Si_3N_4 tip and various samples. The coefficient of friction is less for STA-coated silica samples being adsorbed, as compared to an uncoated sample. The STA coating acts as a lubricant film. The coefficient of friction is found to be dependent upon the concentration of STA, and it decreases with an increase in the concentration. *Bhushan* et al. [38.75] have reported that the density and distribution of the biomolecules vary with the concentration. At a higher concentration of the solution, the coating layer is more uniform, and covers more of the silica substrate surface than at a lower concentration. This means that the surface forms a continuous lubricant film at a higher concentration.

In the case of samples prepared by the chemical linker method, the coefficient of friction increases with an increase in the biomolecular chain length due to increased compliance. When normal load is applied to the surface, the surface gets compressed, resulting in a larger contact area between the AFM tip and the biomolecules. Further, the size of STA is much larger than that of APTES and biotin, resulting in a surface tightly packed with biomolecules that results in very little lateral deflection of the sulfo-NHS-biotin linker. Due to the high contact area and low lateral deflection, the friction force increases for the same applied normal load as compared to samples prepared by the physical adsorption method. These tests reveal that the surfaces coated with biomolecules reduce friction, but if the biomolecular coating of the surface is too thick, or the surface has some cushioning effect, as seen in the chemical linker method, the coefficient of friction increases.

Figure 38.53 shows the surface height maps and phase images of wear marks on STA deposited by

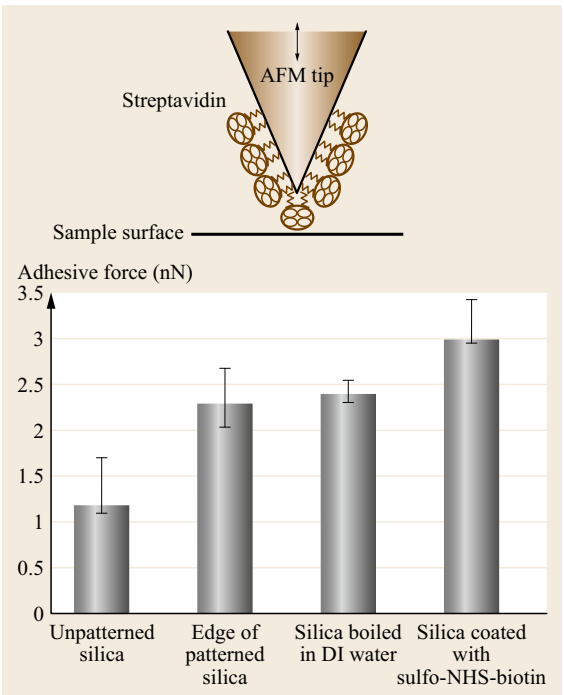


Fig. 38.51 Adhesion measurements of silica, patterned silicon, silica boiled in DI water, and sulfo-NHS-biotin using functionalized (with streptavidin) tips obtained from force–distance curves, captured in PBS

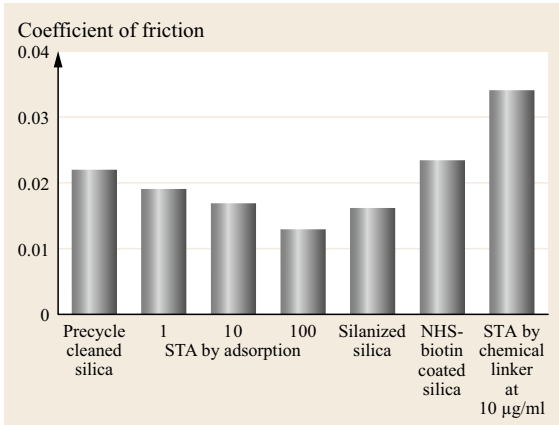


Fig. 38.52 Coefficient of friction for various surfaces with and without biomolecules (after [38.93])

physical adsorption taken after wear tests at three normal loads. The wear depth increases with the increasing normal load. An increase in normal load causes partial damage to the folding structure of the streptavidin molecules. It is unlikely that the chemical (covalent) bonds within the streptavidin molecule are broken, in-

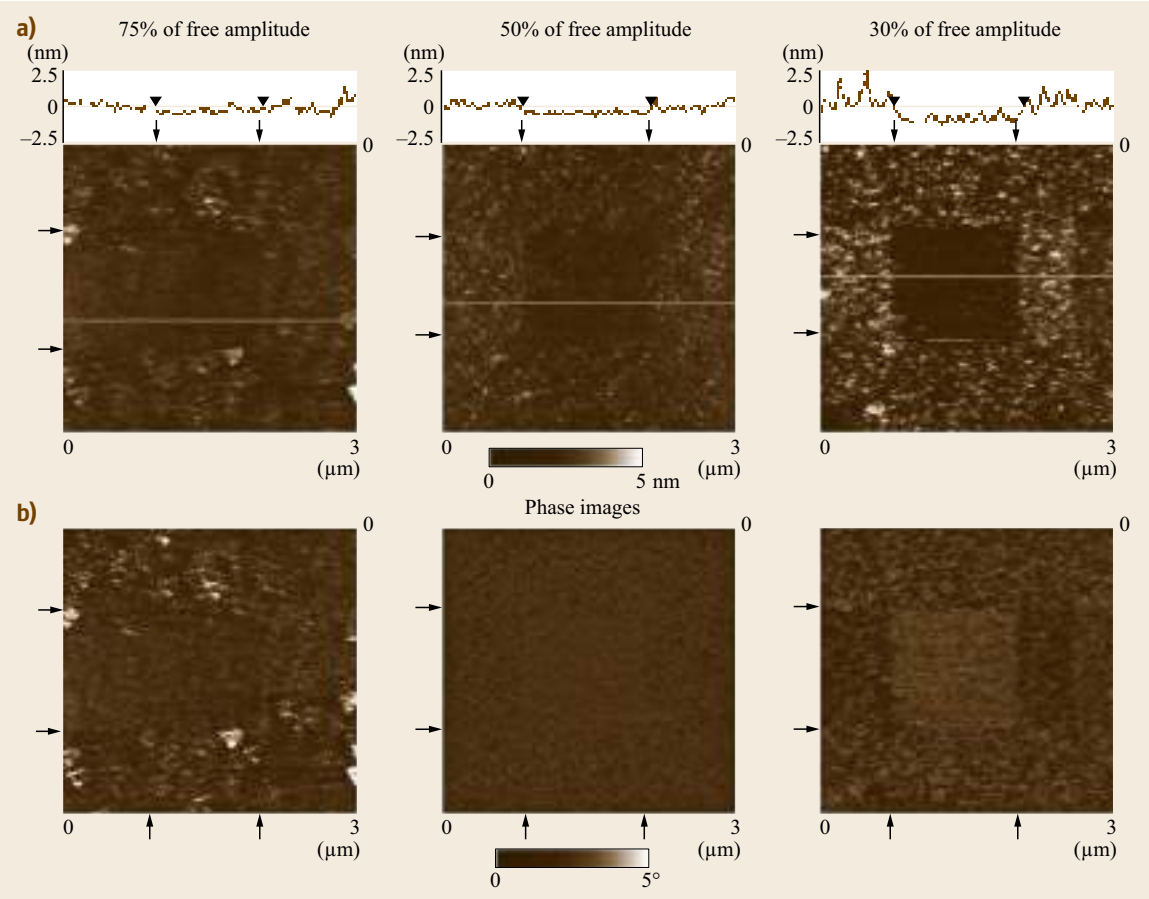


Fig. 38.53a,b Surface height maps and cross-sectional profiles **(a)** and phase images **(b)** of wear marks on precycle cleaned silica coated with streptavidin by physical adsorption after wear tests at three normal loads (increasing from *left* to *right*). The 75, 50, and 30% of free amplitudes correspond to equivalent normal loads of 3, 6, and 8 nN respectively (after [38.93])

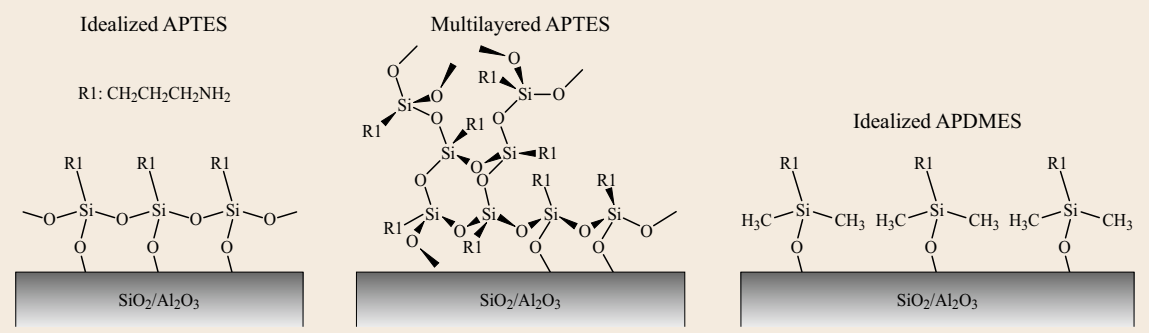


Fig. 38.54 Schematics of idealized and multilayered APTES and idealized APDMES silane polymer films on silicon/SiO_2 and aluminum/Al_2O_3 substrates. Idealized film corresponds to a self-assembled monolayer (after [38.150])

stead, the folding structure is damaged, leading to the wear mark. When the load is high, about 30% of free amplitude (≈ 8 nN), the molecules may be removed by the AFM tip due to the indentation effect. Because of this, there is a significant increase in the wear depth from 50% of free amplitude (≈ 6 nN) to 30% of free amplitude (≈ 8 nN). The data show that biomolecules will be damaged during sliding.

In summary, for the samples prepared using nanopatterning and chemical linker methods, adhesion is higher as compared to the direct adsorption method. The coefficient of friction is smaller for STA-coated silica prepared using the direct adsorption method as compared to an uncoated silica sample. The coefficient of friction decreases with increasing concentration of STA in the solution because the protein acts as a lubricating film. Friction increases for the STA sample prepared using the chemical linker method due to low lateral deformation and the cushioning effect. Wear of STA increases with the increasing load.

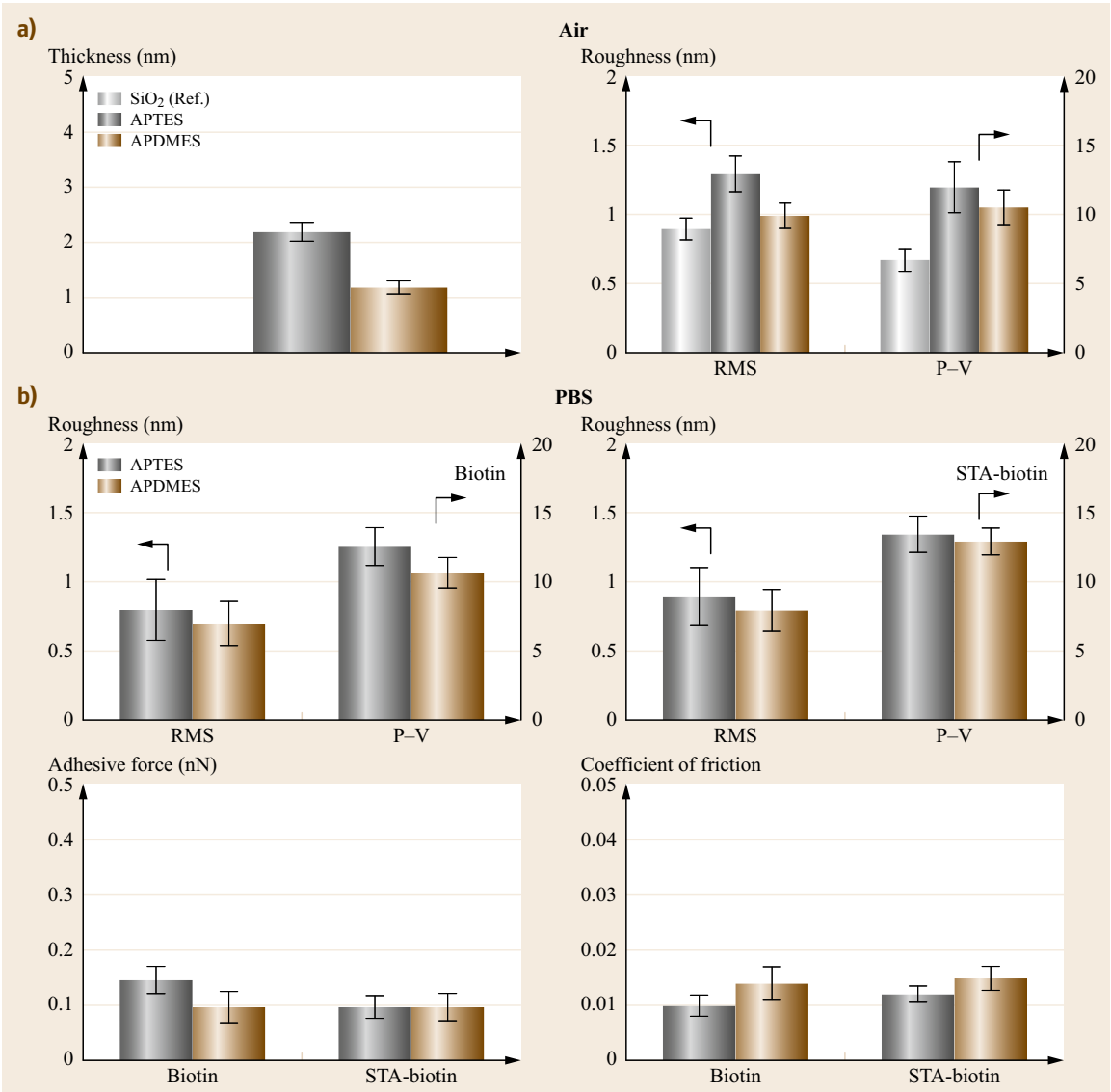


Fig. 38.55 **(a)** Summary of film thickness and surface roughness (RMS and P–V distance) for APTES and APDMES in air, and **(b)** summary of the surface roughness (RMS and P–V distance), adhesive force, and coefficient of biotin and STA-biotin on APTES and APDMES films, all in PBS buffer solution. *Error bars* represent $\pm 1\sigma$ (after [38.150])

APDMES to Improve Adhesion, Friction and Wear of Biomolecular Films

APTES films used in the study reported above were not very smooth, and P–V distances were greater than the summed bond strengths of APTES. Therefore, the biotin/STA deposited on APTES films were not very robust. APTES films are commonly described as SAMs, though this is often inaccurate [38.260–262]. APTES monomers can form infinite, cross-linked siloxane polymer lattices because APTES is a multi-(tri)valent silane. Further, APTES can form a multilayered structure due to intermolecular polymerization with significant cross-linking between monomers. It should consist of sparse cross-links between the polymer and the substrate, Fig. 38.54 [38.150]. The propensity to form multilayers and its low mechanical strength make APTES an undesirable interface material. On the other hand, 3-aminopropyldimethylethoxysilane (APDMES) cannot polymerize into extensive networks because it is monovalent, forming only siloxane dimers or linkages

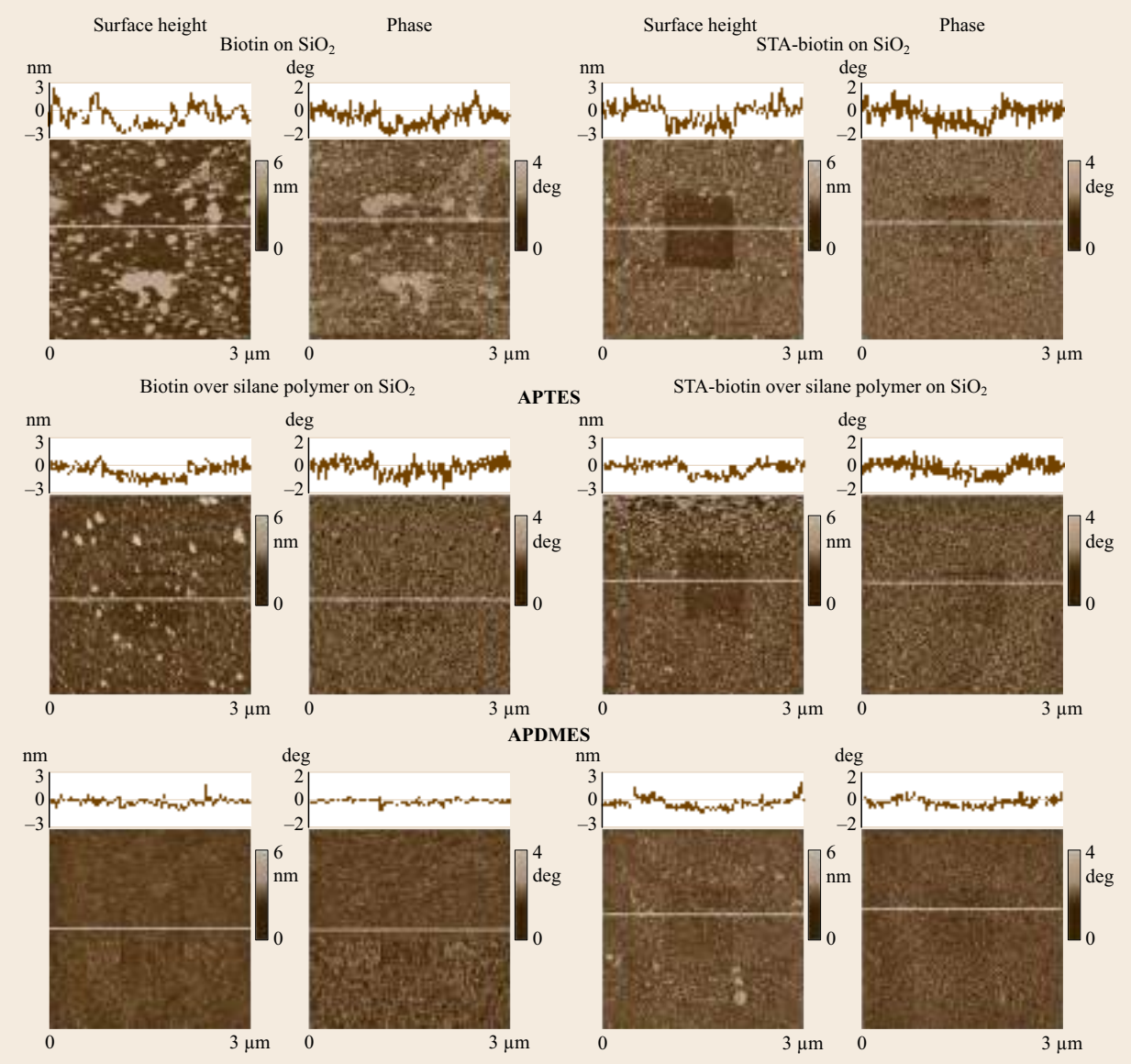


Fig. 38.56 AFM surface height and phase angle images and cross-sectional profiles obtained after wear testing in PBS in tapping mode at 50% of free amplitude (≈ 2 nN) on biotin and STA-biotin on SiO_2 and APTES and APDMES films on SiO_2. The *white lines* indicate the locations of the cross-sections.

to substrate oxides (Fig. 38.54). Therefore, APDMES should produce thinner films of greater mechanical robustness than APTES and provide the robustness to the biomolecular layers deposited on its top. Thinner interfaces theoretically would also increase sensor sensitivity to analytes.

Bhushan et al. [38.150] examined the thickness and durability of APDMES deposited on SiO_2 and Al_2O_3 substrates, with biotin and biotin/STA bound to them. Figure 38.55a shows the thickness and surface roughness (RMS and P–V) of APTES, APDMES, and SiO_2 substrate for reference. Figure 38.55b shows the surface roughness, adhesive force, and coefficient of friction for biotin and STA-biotin deposited on APTES and APDMES films. The data show that the APTES film was not very smooth. It was shown in Fig. 38.50 that biotin and STA deposited on APTES were also not very smooth, with high P–V values. The thickness of the APTES layer was larger than the expected monolayer thickness. The unexpected thickness of the film suggests multilayering by APTES. APDMES films produced the thinner film, with a thickness comparable to the summed bond lengths of the APDMES polymer (Fig. 38.55a). *Bhushan* et al. [38.150] reported that the APDMES film was more uniform, smoother, nearly continuous, and exhibited a higher contact angle and lower adhesive force as compared to the APTES film. The surface roughness and adhesive force of biotin and STA-biotin on APDMES was also slightly lower than that on APTES film. The coefficient of friction on APDMES was slightly higher than that on APTES (Fig. 38.55b).

Bhushan et al. [38.150] also studied wear properties of various films. Contact mode always immediately stripped the surface at low loads; consequently, wear experiments were performed in tapping mode at various loads. AFM surface height and phase angle images and cross-sectional profiles obtained after wear tests for biotin and STA-biotin on APTES and APDMES in PBS are shown in Fig. 38.56. As controls, the wear experiments of biotin and STA-biotin on SiO_2 without SAM were also carried out. The biomolecular films on APDMES were more robust than that on APTES. Given that each molecule of APDMES must be bonded to a surface oxide group and that intrapolymer cross-links are not possible in APDMES, it has the highest density of siloxane linkages to the substrate oxide, making it the more robust. *Bhushan* et al. [38.150] also studied the effect of load on the wear of APDMES and biotin and STA-biotin deposited on APDMES; Fig. 38.57. As expected, wear increased with an increase in the load. The relationship between the average wear depth (and the coefficient of friction) and the average normal load is generally linear. The slope of the wear depth (and coefficient of friction) to load is steepest for the interface to which STA-biotin was bound because of the cushioning effect of the thick film, as suggested earlier.

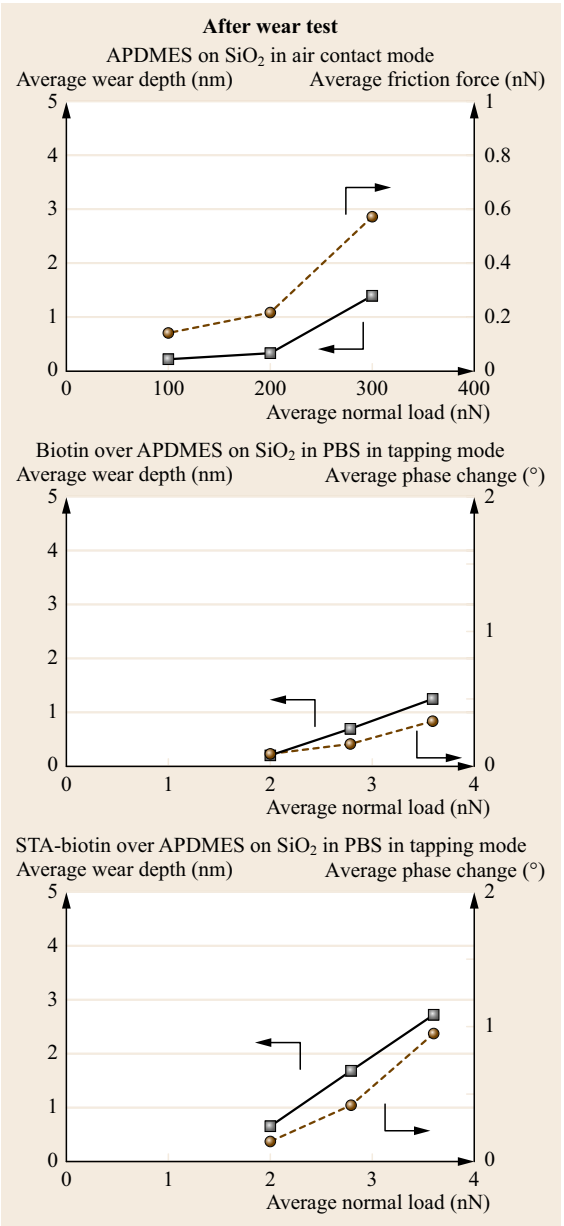


Fig. 38.57 Plot of average wear depth and average friction force/phase angle as a function of average normal load for APDMES on SiO_2 in air in contact mode, biotin over APDMES on SiO_2 in PBS in tapping mode, and on STA-biotin over APDMES on SiO_2 in PBS in tapping mode (after [38.150])

In summary, an APDMES film is more uniform, smoother, and nearly continuous as compared to APTES film. These properties of APDMES provide a good interfacial material for biomolecular films providing a smooth and robust structure.

38.6.3 Adhesion of Coated Polymer Surfaces

As mentioned in Appendix 38.A, poly(methyl methacrylate) (PMMA), PDMS, and other polymers are used in the construction of micro/nanofluidic-based biodevices. Adhesion between the moving parts needs to be minimized. Furthermore, if the adhesion between the microchannel surface and the biofluid is high, the biomolecules will stick to the microchannel surface and restrict flow. In order to facilitate flow, a surface with low bioadhesion is required.

Tambe and *Bhushan* [38.263, 264] and *Bhushan* and *Burton* [38.265] have reported adhesive force data for PMMA and PDMS against an AFM Si_3N_4 tip and a silicon ball. *Tokachichu* and *Bhushan* [38.266] measured contact angle and adhesion of bare PMMA and PDMS and those coated with a perfluoro SAM of perfluorodecyltriethoxysilane (PFDTES). Oxygen plasma treatment was used for hydroxylation of the surface to enhance chemical bonding of the SAM to the polymer surface. They made measurements in ambient conditions, PBS, and fetal bovine serum (FBS); the latter is a blood component. Figures 38.58 and 38.59 show the contact angle and adhesion data. SAM-coated surfaces have high contact angles (Fig. 38.58), as expected. The adhesion value of PDMS in ambient conditions is high because of the electrostatic charge present on the surface. The adhesion values of PDMS are higher than PMMA because PDMS is softer than PMMA (elastic modulus = 5 GPa and hardness = 410 MPa [38.151]), resulting in a higher contact area between the PDMS surface and the AFM tip. Additionally, PMMA does not develop electrostatic charge. When SAM is coated on PMMA and PDMS surfaces the adhesion values are similar, which shows that the electrostatic charge on virgin PDMS plays no role when the surface is coated. In the PBS solution, there is a decrease in adhesion values

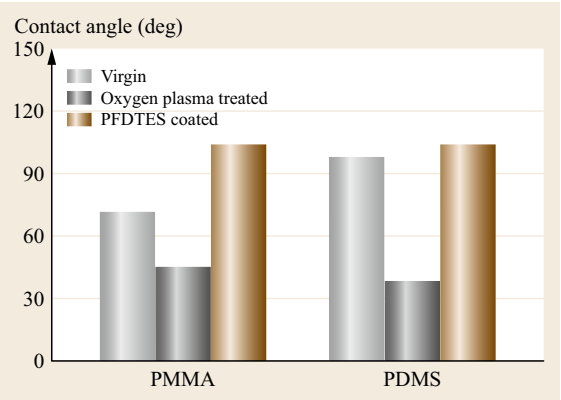


Fig. 38.58 Sessile drop contact angle measurements of virgin, oxygen plasma-treated and PFDTES-coated PMMA and PDMS surfaces. The maximum error in the data is ±2° (after [38.266])

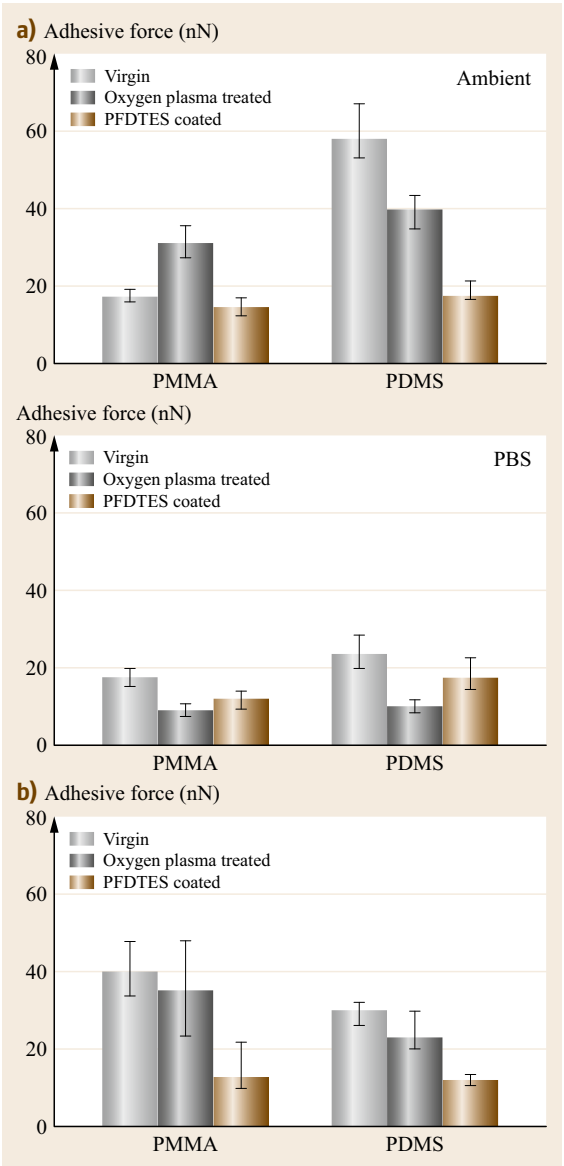


Fig. 38.59a,b Adhesion measurement of virgin, oxygen plasma-treated and PFDTES-coated PMMA and PDMS surfaces **(a)** with bare silicon nitride AFM tip in ambient, and in PBS environment, and **(b)** dip-coated tip with FBS in PBS environment (after [38.266])

because there is no meniscus contribution. The adhesion values for an FBS-coated tip in PBS are generally lower than for an uncoated tip in PBS.

In summary, the adhesion values of SAM-coated surfaces are lower than bare surfaces in various environments.

38.7 Trajectory of Submicron Particles for Therapeutics and Diagnostics

For drug delivery, submicron particles are injected into the bloodstream in human capillaries (as small as 4–5 μm) and employed to deliver drugs to diseased cells, to locate diseased cells or tumoral masses and estimate the state of disease, and to carry diagnostic agents (fluorescent molecules) to diseased cells to enhance imaging [38.163]. Particles exhibiting one of these characteristics are considered smart systems, which can function as purely therapeutic agents or purely diagnostic agents or a combination of both. Small delivery particles include nanocrystals, synthetic vesicles, liposomes, and fabricated silicon.

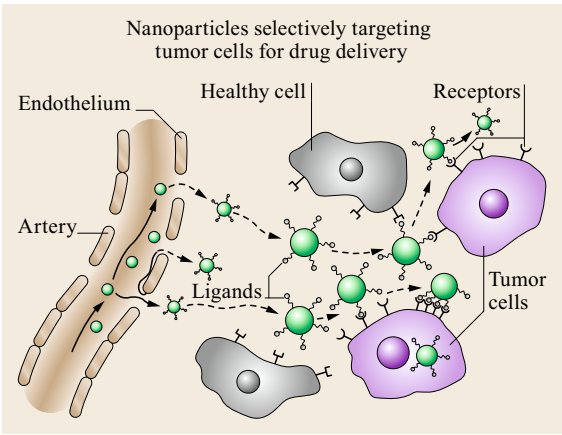


Fig. 38.60 Trajectory of submicron particles injected intravenously for selectively targeting tumor cells for drug delivery (after [38.162]).

A schematic of the trajectory of a submicron particle injected intravenously for selectively targeting tumor cells for drug delivery is shown in Fig. 38.60. Particles are reservoirs containing drug or diagnostic agents. These are covered with a layer of adhesive molecules (ligands) for attaching to selected target sites. When a particle is sufficiently close to select cell surfaces (a few nm), its ligands can interact with the cellular counterpart (receptors), leading to firm attachment. Once the particle is attached to its target, a drug or diagnostic agent is delivered. The speed and the time needed for a particle circulating in the bloodstream to reach the endothelium lining of the human capillary is dependent upon the distance and interactive forces. The particle can interact with the endothelium lining with buoyancy, van der Waals forces, and electrostatic and steric forces. These interactions are a function of the material properties of the particle and any coating on it, in particular, relative density (particle density relative to blood), the electrostatic potential, the dielectric constant, and the particle radius. These properties can be optimized.

The interacting forces are weak, so the particle trajectory should be close to the endothelium lining. Otherwise, the particle may not get attracted to the lining and perform its intended function. *Decuzzi* et al. [38.163] developed an analytical model to predict the trajectory of a particle freely circulating in

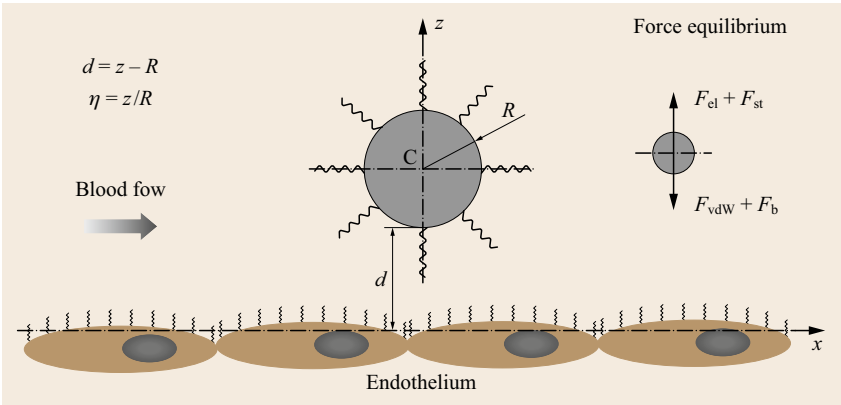


Fig. 38.61 A spherical particle moving close to the endothelium wall in a laminar flow. In the inset, the balance of the forces acting over the particle are sketched. F_{el}, F_{st}, F_{vdW}, and F_b correspond to electrostatic force, steric force, van der Waals force, and buoyancy force respectively (after [38.163])

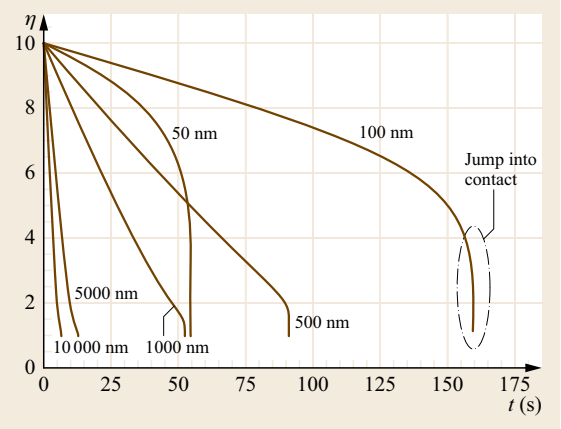


Fig. 38.62 The dimensionless particle position $\eta(= z/R)$ as a function of time t for different values of particle radius ($R = 10\,000$, 5000, 1000, 500, 100, and 50 nm). There exists a critical radius R_c at which the travel time is maximum (after [38.163])

the bloodstream and associated interaction forces. The model can be used to optimize the particle radii and material properties. Figure 38.61 shows a spherical particle freely circulating in the bloodstream with its center at a distance z from the endothelium wall. The particle has a radius R and its trajectory is governed by the forces exerted by the bloodstream, gravitation, and electromagnetic interactions. We assume that the particle is sufficiently far from the endothelium wall that specific interactions (such as ligand–receptor interactions) can be ignored. At short range (below 1−3 nm), solvation and other steric forces dominate. However, the most important long-range forces (5−10 nm) between particles and wall surfaces in the presence of a liquid are buoyancy and van der Waals, electrostatic and steric forces. Buoyancy is related to the radius of the particle R and the relative density of the particle relative to blood. The van der Waals interaction (generally attractive) is related to R, its relative position with respect to the endothelium wall z, and the Hamaker constant A, which depends upon the dielectric constants of the media involved. The electrostatic double layer (EDL) interaction (repulsive and attractive) is related to R, z, the ionic concentration, and the characteristic Debye length of the solution. Finally, steric repulsive interaction is related to the unperturbed radius of gyration of polymer chains grafted on the particle surface, R and z. The value of various forces as a function of the particle radius R for a fixed distance from the wall z is presented in Table 38.10. The dominating force is buoyancy when the particle radius is sufficiently large, and van der Waals when the particle is sufficiently small.

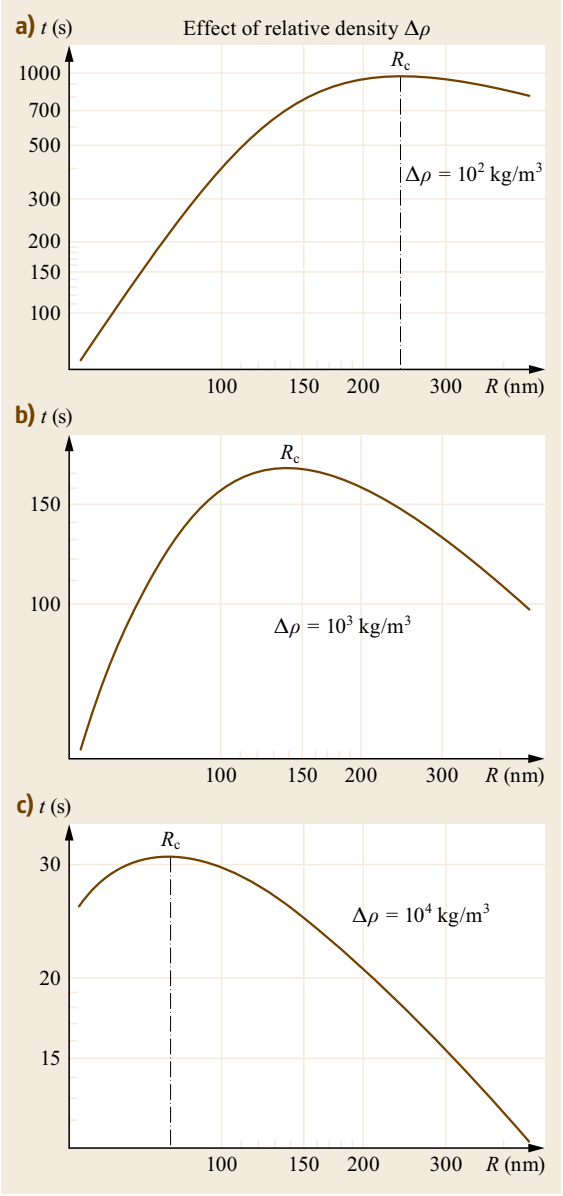


Fig. 38.63a–c The travel time as a function of particle radius R for different values of the relative density **(a)** $\Delta\rho = 10^2$, **(b)** $\Delta\rho = 10^3$, and **(c)** $\Delta\rho = 10^4$ kg/mm^3. The critical radius R_c depends upon $\Delta\rho$ (after [38.163])

Electrostatic and steric forces are negligible as long as the distance is larger than the 50 nm considered in this example.

The travel time needed to reach the wall depends upon the particle size, as shown in Fig. 38.62. In this figure, the dimensionless position of the particle center with respect to the endothelium wall η $(= z/R)$ is plot-

Table 38.10 The value of the buoyancy force F_b, van der Waals force F_{vdW}, electrostatic force F_{el} and steric force F_{sr} as a function of the particle radius for a fixed distance from the wall ($\eta = z/R = 2$ or $d = R$) (after [38.163])

R (nm)	F_b (pN)	F_{vdW} (pN)	F_{el} (pN)	F_{sr} (pN)
10 000	41.092	0.0005	≈ 0	≈ 0
1000	41.092×10^{-3}	0.005	≈ 0	10^{-33}
100	41.092×10^{-6}	0.05	10^{-52}	1.2×10^{-3}
50	51.3×10^{-7}	0.1	10^{-25}	3×10^{-2}

ted as a function of time t for different particle radii, ranging from $R = 10\,\mu m$ to 50 nm. It was assumed that particles were initially at a distance d_0 ($z_0 - R$) equal to $9R$ from the endothelium wall (i. e., the center of the particle is at $z_0 = 10R$, $\eta_0 = z_0/R = 10$). Thus, the distance traveled scales with the size of the particle. The data show that as the radius of the particle decreases, the time needed to reach the wall increases up to a maximum beyond which it decreases as the radius is further reduced. Larger particles are initially far from the endothelium wall; a 10 μm particle is initially 100 μm away from the endothelium ($z = 10R$) where electrostatic and van der Waals interactions are negligible and the particle motion is governed by buoyancy and hemodynamic resistance. As the particle approaches the endothelium wall, van der Waals attraction dominates, and a rapid increase in speed is observed with a *jump into contact* kind of behavior. Such a behavior is more clearly shown by particles with submicron radius closer to the endothelium wall, so the van der Waals force dominates from the beginning. We note that there is a critical radius at which the time taken for the particle to travel to the wall is maximum. The selected radius should be smaller or larger than the critical radius (possibly smaller so that they can circulate freely even in smaller capillaries).

The effect of the relative density of the particle $\Delta\rho$ on the critical radius R_c at which travel time is maximum is plotted in Fig. 38.63. We note that as $\Delta\rho$ reduces, the R_c increases. As $\Delta\rho$ decreases, the effect of buoyancy becomes less important, and van der Waals attractive forces exert a greater influence as particle radii decrease.

In summary, the interacting forces are weak, so the particle trajectory should be close to the endothelium lining. The trajectory and interaction forces depend upon the particle radii and material properties. These results suggest that particles for therapeutics and diagnostics should have a radius larger or smaller than the critical value (possibly smaller so that they can circulate freely even in smaller capillaries). The material properties of the particle and the polymer chains grafted on it, such as the relative density of the particle, can be tuned specifically to the type of malignant tissue and the state of disease, improving the particle affinity with the diseased cell.

38.8 Component-Level Studies

38.8.1 Surface Roughness Studies of Micromotor Components

Most of the friction forces resisting motion in a micromotor are concentrated near the rotor–hub interface where continuous physical contact occurs. Surface roughness usually has a strong influence on the friction characteristics on the micro/nanoscale. A catalog of roughness measurements on various components of a MEMS device does not exist in the literature. Using an AFM, measurements on various component surfaces were made for the first time by *Sundararajan* and *Bhushan* [38.267]. Table 38.11 shows various surface roughness parameters obtained from $5 \times 5\,\mu m$ scans of the various component surfaces of several unlubricated micromotors using the AFM in tapping mode. A surface with a Gaussian height distribution should have a skewness of zero and kurtosis of three. Although the rotor and stator top surfaces exhibit comparable roughness parameters, the underside of the rotors exhibits lower RMS roughness and P–V distance values. More importantly, the rotor underside shows negative skewness and lower kurtosis than the topsides, both of which are conducive to high real area of contact and hence high friction [38.91, 92]. The rotor underside also exhibits a higher coefficient of microscale friction than the rotor topside and stator, as shown in Table 38.11. Figure 38.64 shows representative surface height maps of the various surfaces of a micromotor measured using the AFM in tapping mode. The rotor underside exhibits varying topography from the outer edge to the middle and inner edge. At the outer edges, the topography shows smaller circular asperities, similar to the topside. The middle and inner regions show deep pits with

Table 38.11 Surface roughness parameters and microscale coefficient of friction for various micromotor component surfaces measured using an AFM. Mean and $\pm 1\sigma$ values are given

	RMS roughness[a] (nm)	Peak-to-valley distance[a] (nm)	Skewness[a], *Sk*	Kurtosis[a], *K*	Coefficient of microscale friction[b], μ
Rotor topside	21±0.6	225±23	1.4±0.30	6.1±1.7	0.07±0.02
Rotor underside	14±2.4	80±11	−1.0±0.22	3.5±0.50	0.11±0.03
Stator topside	19±1	246±21	1.4±0.50	6.6±1.5	0.08±0.01

[a] Measured from a tapping mode AFM scan of size 5 μm × 5 μm using a standard Si tip scanning at 5 μm/s in a direction orthogonal to the long axis of the cantilever.
[b] Measured using an AFM in contact mode at 5 μm × 5 μm scan size using a standard Si_3N_4 tip scanning at 10 μm/s in a direction parallel to the long axis of the cantilever

fine edges that may have been created by the etchants used for etching of the sacrificial layer. It is known that etching can affect the roughness of surfaces in surface micromachining. The residence time of the etchant near the inner region is high, responsible for larger pits. Figure 38.65 shows roughness of the surface directly beneath the rotors (the base polysilicon layer). There appears to be a difference in the roughness between the portion of this surface that was initially underneath the rotor (region B) during fabrication and the portion that was away from the rotor and hence always exposed (region A). The former region shows lower roughness than the latter region. This suggests that the surfaces at the rotor–hub interface that come into contact at the end of the fabrication process exhibit large real areas of contact that result in high friction.

38.8.2 Adhesion Measurements of Microstructures

Surface force apparatuses (SFA) and AFMs are used to measure adhesion on micro- to nanoscales between two surfaces. In the SFA, adhesion of liquid films sandwiched between two curved and smooth surfaces is measured. In an AFM, as discussed earlier, adhesion between a sharp tip and the surface of interest is measured. The propensity of adhesion between two surfaces can be evaluated by studying the tendency of microstructures with well-defined contact areas covering a wide spectrum of suspension compliances to stick to the underlying substrate. The test structures that have been used include the cantilever beam array (CBA) technique with different lengths [38.268–271] and stand-off multiple dimples mounted on microstructures with a range of compliances, standing above a substrate [38.272]. The more commonly used CBA technique uses an array of micromachined polysilicon beams (for Si MEMS applications) in the mesoscopic length scale that are anchored to the substrate at one

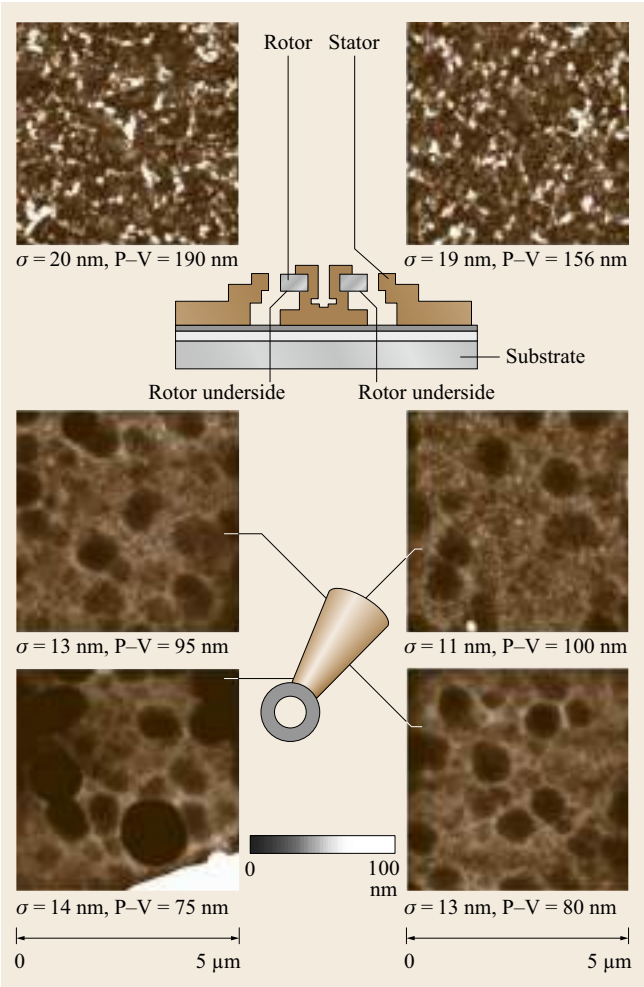


Fig. 38.64 Representative AFM surface height images obtained in tapping mode (5 μm × 5 μm scan size) of various component surfaces of a micromotor. RMS roughness and P–V values of the surfaces are given. The underside of the rotor exhibits drastically different topography from the topside (after [38.267])

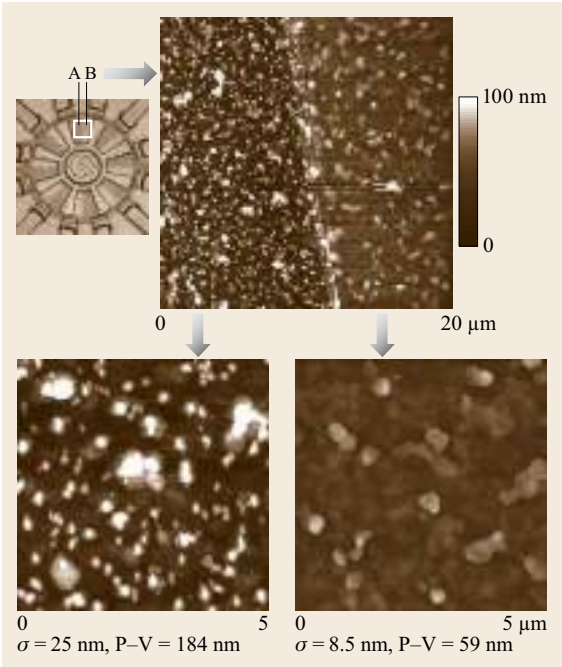


Fig. 38.65 Surface height images of polysilicon regions directly below the rotor. Region A is away from the rotor while region B was initially covered by the rotor prior to the release etch of the rotor. During this step, slight movement of the rotor caused region B to be exposed (after [38.267])

end and with different lengths parallel to the surface. CBA relies on peeling and detachment of cantilever beams. Change in free energy or reversible work done to separate unit areas of two surfaces from contact is called work of adhesion. To measure the work of adhesion, electrostatic actuation is used to bring all beams into contact with the substrate; Fig. 38.66 [38.268, 270]. Once the actuation force is removed, the beams begin to peel themselves off the substrate, which can be observed with an optical interference microscope (e.g., Wyko surface profiler). For beams shorter than a characteristic length, the so-called detachment length, their stiffness is sufficient to free them completely from the substrate underneath. Beams larger than the detachment length remain adhered. The beams at the transition region start to detach and remain attached to the substrate just at the tips. For this case, by equating the elastic energy stored within the beam and the beam–substrate interfacial energy, the work of adhesion, W_{ad}, can be calculated by the following equation [38.268]

$$W_{ad} = \frac{3Ed^2t^3}{8\ell_d^4}\,, \tag{38.1}$$

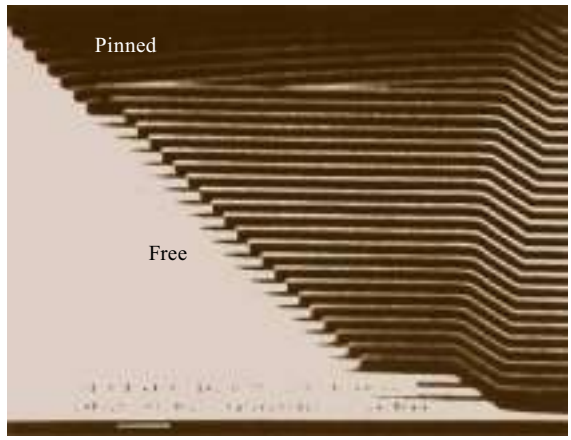


Fig. 38.66 SEM micrograph of micromachined array of polysilicon cantilever beams of increasing length. The micrograph shows the onset of pinning for beams longer than 34 μm (after [38.268])

where E is the Young's modulus of the beam, d is the spacing between the undeflected beam and the substrate, t is the beam thickness, and ℓ_d is the detachment length. The technique has been used to screen methods for adhesion reduction in polysilicon microstructures.

38.8.3 Microtriboapparatus for Adhesion, Friction, and Wear of Microcomponents

To measure adhesion, friction, and wear between two microcomponents, a microtriboapparatus has been used. Figure 38.67 shows a schematic of a microtriboapparatus capable of using MEMS components for tests [38.273]. In this apparatus, an upper specimen, mounted on a soft cantilever beam, comes in contact with a lower specimen mounted on a lower specimen holder. The apparatus consists of two piezoelectric bodies (x- and z-piezoelectric bodies) and four fiber-optic sensors (x- and z-displacement sensors, and x- and z-force sensors). For adhesion and friction studies, the z-piezoelectric body is used to bring the upper specimen and lower specimen into contact, and the x-piezoelectric body is used to apply a relative motion in the lateral direction. The x- and z-displacement sensors are used to measure the lateral position of the lower specimen and vertical position of the upper specimen respectively. By monitoring the deflection of the cantilever, the x- and z-force sensors are used to measure friction force and normal load/adhesive force between the two specimens respectively.

As most MEMS/NEMS devices are fabricated from silicon, study of silicon-on-silicon contacts is important. This contact was simulated by a flat single-crystal

Si(100) wafer (phosphorus-doped) specimen sliding against a single-crystal Si(100) ball (1 mm in diameter, 5×10^{17} atoms/cm^3 boron doped) mounted on a stainless steel cantilever [38.273, 274]. Both of them have a native oxide layer on their surfaces. With an objective to reduce adhesion and friction and humidity effects, the other materials studied were 10 nm-thick DLC deposited by filtered cathodic arc deposition on Si(100), 2.3 nm-thick chemically bonded PFPE (Z-DOL, BW) on Si(100), and hexadecane thiol (HDT) monolayer on evaporated Au(111) film to investigate their anti-adhesion performance. For completeness, other approaches to reduce liquid-mediated adhesion include the formation of dimples on the contact surfaces to reduce contact area [38.13, 87, 88, 91, 92, 173, 270] and bioinspired superhydrophobic surfaces [38.275].

It is well known that in computer rigid disk drives, the adhesive force increases rapidly with an increase in rest time between a magnetic head and a magnetic disk [38.32]. Considering that adhesion and friction are the major issues that lead to the failure of MEMS/NEMS devices, the rest time effect on the microscale for Si(100), DLC, PFPE, and HDT was studied, and the results are summarized in Fig. 38.68a. It was found that the adhesive force of Si(100) logarithmically increases with the rest time to a certain equilibrium time ($t = 1000$ s), after which it remains constant. Figure 38.68a also shows that the adhesive force of DLC, PFPE, and HDT does not change with rest time. Single-asperity contact modeling of the dependence of meniscus force on the rest time has been carried out by *Chilamakuri* and *Bhushan* [38.276], and the modeling results (Fig. 38.68b) verify experimental observations. Due to the presence of a thin film of water adsorbed on Si(100), a meniscus forms around the contacting asperities and grows with time until equilibrium occurs, which causes the rest time effect on its adhesive force. The adhesive forces of DLC, PFPE, and HDT do not change with rest time, which suggests that the water meniscus is not present on their surfaces.

The measured adhesive forces of Si(100), DLC, PFPE, and HDT at rest time of 1 s are summarized in Fig. 38.69. It shows that the presence of solid films of DLC, PFPE, and HDT greatly reduced the adhesive force of Si(100), whereas HDT film had the lowest adhesive force. It is well known that the native oxide layer (SiO_2) on the top of a Si(100) wafer exhibits hydrophilic properties, and water molecules, produced by capillary condensation of water vapor from the environment, can easily be adsorbed on this surface. The condensed water will form a meniscus as the upper specimen approaches the lower specimen surface. The meniscus force is a major contributor to the adhesive force. In the case of DLC, PFPE, and HDT, the films

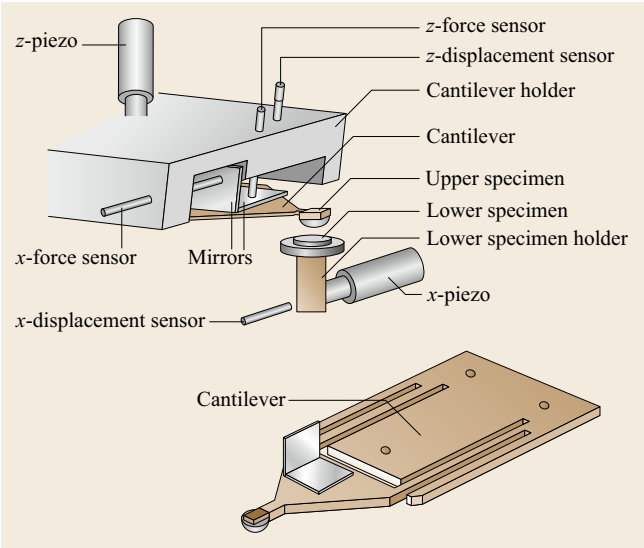


Fig. 38.67 Schematic of the microtriboapparatus including specially designed cantilever (with two perpendicular mirrors attached on the end), lower specimen holder, two piezoelectric bodies (x- and z-piezoelectric bodies), and four fiber-optic sensors (x- and z-displacement sensors and x- and z-force sensors) (after [38.273])

were found to be hydrophobic based on contact angle measurements, and the amount of condensed water vapor was low as compared to that on Si(100). It should be noted that the measured adhesive force is generally higher than that measured in AFM, because the larger radius of a Si(100) ball as compared to that of an AFM tip induces larger meniscus and van der Waals forces.

To investigate the velocity effect on friction, the friction force as a function of velocity was measured and is summarized in Fig. 38.70a. For Si(100), the friction force initially decreases with increasing velocity until equilibrium occurred. Figure 38.70a also shows that the velocity had almost no effect on the friction properties of DLC, PFPE, and HDT. This implies that the friction mechanisms of DLC, PFPE, and HDT do not change with the variation of velocity. For Si(100), at high velocity, the meniscus is broken and does not have enough time to rebuild. In addition, it is also believed that tribochemical reaction plays an important role. The high velocity leads to tribochemical reactions of Si(100) (which has native oxide SiO_2) with water molecules to form a $Si(OH)_4$ film. This film is removed and continuously replenished during sliding. The $Si(OH)_4$ layer at the sliding surface is known to be of low shear strength. The breaking of the water meniscus and the formation of a $Si(OH)_4$ layer results in a decrease in friction force of Si(100). For DLC, PFPE, and HDT, their surfaces exhibit hydrophobic properties, and only can adsorb a few water molecules

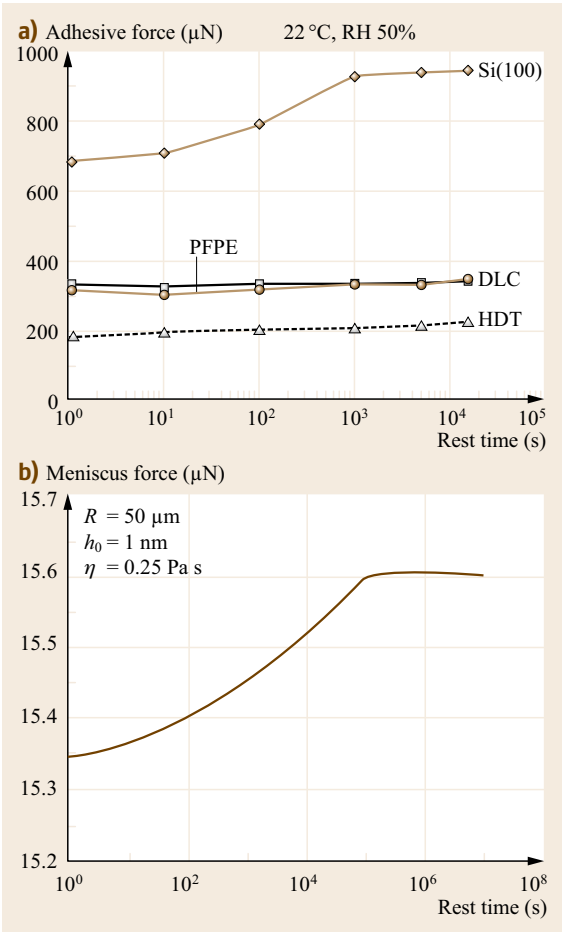


Fig. 38.68 **(a)** The influence of rest time on the adhesive force of Si(100), DLC, chemically bonded PFPE, and HDT, and **(b)** single-asperity contact modeling results of the rest time effect on the meniscus force for an asperity of radius R in contact with a flat surface with a water film of thickness of h_0 and absolute viscosity of η_0 (after [38.276])

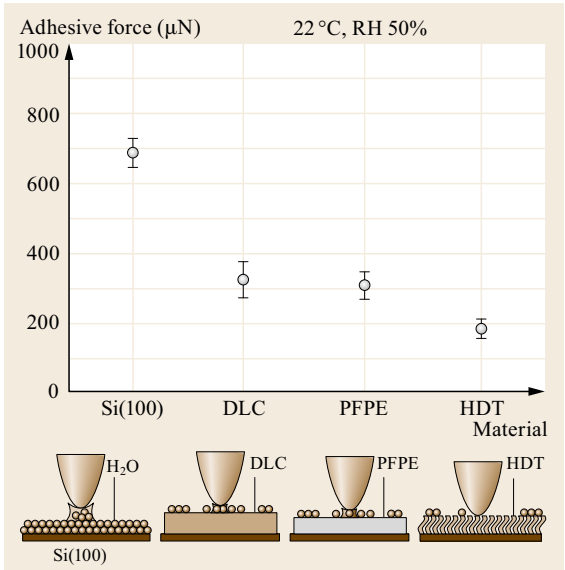


Fig. 38.69 Adhesive forces of Si(100), DLC, chemically bonded PFPE, and HDT at ambient conditions, and a schematic showing the relative size of water meniscus on different specimens

in ambient conditions. The above-mentioned meniscus breaking and tribochemical reaction mechanisms do not exist for these films. Therefore, their friction force does not change with velocity.

The influence of relative humidity was studied in an environmentally controlled chamber. The adhesive force and friction force were measured by making measurements at increasing relative humidity, and the results are summarized in Fig. 38.70b. It shows that for Si(100), the adhesive force increased with relative humidity, but the adhesive force of DLC and PFPE only showed a slight increase when humidity was higher than 45%, while the adhesive force of HDT did not change with humidity. Figure 38.70b also shows that for Si(100), the friction force increased with an increase in relative humidity up to 45%, and then it shows a slight decrease with a further increase in the relative humidity. For PFPE, there was an increase in the friction force when humidity was higher than 45%. In the whole testing range, relative humidity did not have any apparent influence on the friction properties of DLC and HDT. In the case of Si(100), the initial increase of relative humidity up to 45% caused more adsorbed water molecules, and formed a bigger water meniscus, which led to an increase of friction force. But at very high humidity of 65%, large quantities of adsorbed water can form a continuous water layer that separates the tip and sample surfaces, and acts as a kind of lubricant, which causes a decrease in the friction force. For PFPE, dewetting of the lubricant film at humidities higher than 45% resulted in an increase in adhesive and friction forces. For DLC and HDT, their surfaces show hydrophobic properties, and increasing relative humidity did not play much of a role in their friction force.

The influence of temperature was studied using a heated stage. The adhesive force and friction force were measured by making measurements at increasing temperatures from 22 to 125 °C. The results are presented in Fig. 38.70c. It shows that once the temperature was higher than 50 °C, increasing temperature caused a significant decrease of adhesive and friction forces of Si(100) and a slight decrease in the case of DLC and PFPE. But the adhesion and friction forces of HDT did not show any apparent change with test

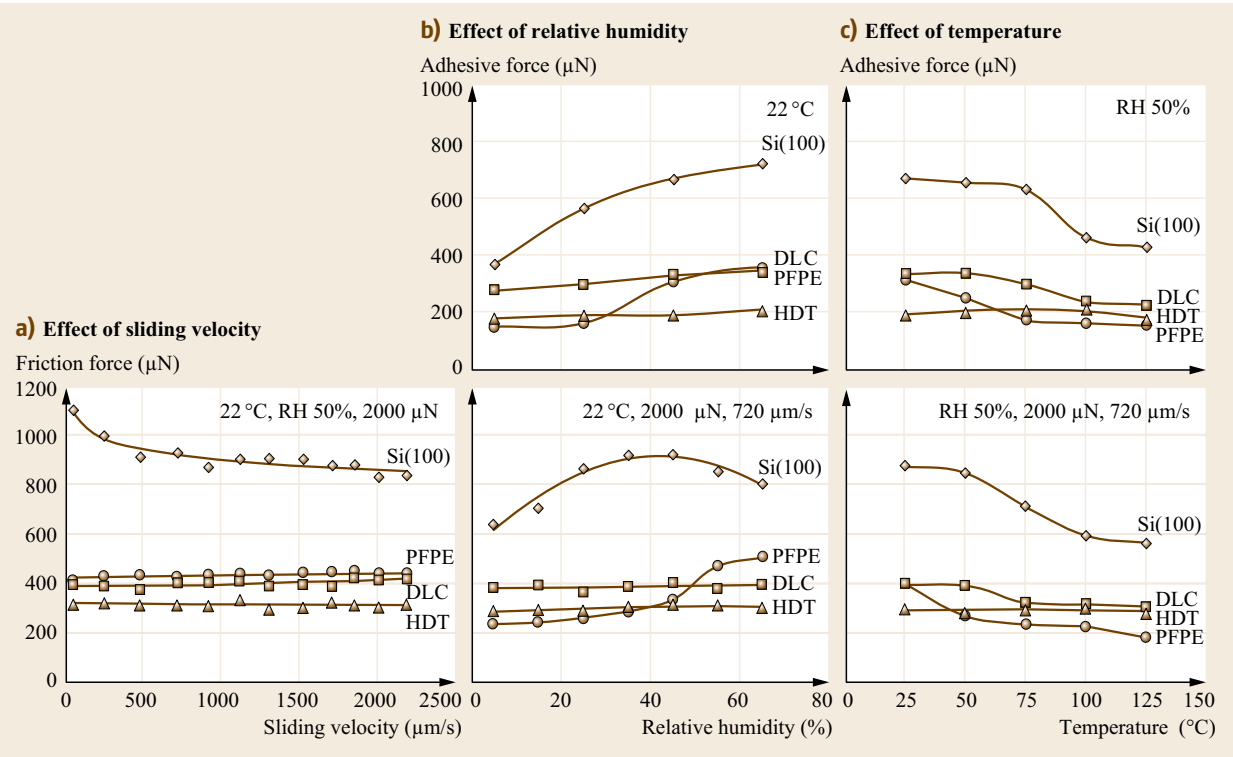


Fig. 38.70a–c The influence of (a) sliding velocity on the friction forces, (b) relative humidity on the adhesive and friction forces, and (c) temperature on the adhesive and friction forces of Si(100), DLC, chemically bonded PFPE, and HDT

temperature. At high temperature, desorption of water and reduction of the surface tension of water led to the decrease of adhesive and friction forces of Si(100), DLC, and PFPE. However, in the case of HDT film, as only a few water molecules were adsorbed on the surface, the above-mentioned mechanisms did not play a big role. Therefore, the adhesive and friction forces of HDT did not show any apparent change with temperature. Figure 38.70 shows that in the whole velocity, relative humidity, and temperature test range, the adhesive force and friction force of DLC, PFPE, and HDT were always smaller than that of Si(100), whereas HDT had the smallest value.

To summarize, several methods can be used to reduce adhesion in microstructures. MEMS/NEMS surfaces can be coated with hydrophobic coatings such as PFPEs, SAMs, and passivated DLC coatings.

38.8.4 Static Friction Force (Stiction) Measurements in MEMS

In MEMS devices involving parts in relative motion to each other, such as micromotors, large friction forces become the limiting factor to the successful operation and reliability of the device. It is generally known that most micromotors cannot be rotated as manufactured and require some form of lubrication. Therefore, it is critical to determine the friction forces present in such MEMS devices. To measure in situ the static friction of a rotor-bearing interface in a micromotor, *Tai* and *Muller* [38.277] measured the starting torque (voltage) and pausing position for different starting positions under a constant-bias voltage. A friction-torque model was used to obtain the coefficient of static friction. To measure the in-situ kinetic friction of the turbine and gear structures, *Gabriel* et al. [38.278] used a laser-based measurement system to monitor the steady-state spins and decelerations. *Lim* et al. [38.279] designed and fabricated a polysilicon microstructure to measure in situ the static friction of various films. The microstructure consisted of a shuttle suspended above the underlying electrode by a folded beam suspension. A known normal force was applied, and lateral force was measured to obtain the coefficient of static friction. *Beerschwinger* et al. [38.280] developed a cantilever-deflection rig to measure friction of LIGA-processed micromotors (see also [38.281]) These techniques employ indirect methods to determine the friction forces or involve fabrication of complex structures.

A novel technique to measure the static friction force (stiction) encountered in surface micromachined polysilicon micromotors using an AFM was developed by *Sundararajan* and *Bhushan* [38.267]. Continuous physical contact occurs during rotor movement (rotation) in the micromotors between the rotor and lower hub flange. In addition, contact occurs at other locations between the rotor and the hub surfaces and between the rotor and the stator. Friction forces will be present at these contact regions during motor operation. Although the actual distribution of these forces is not known, they can be expected to be concentrated near the hub where there is continuous contact. If we therefore represent the static friction force of the micromotor as a single force F_s acting at point P_1 (as shown in Fig. 38.71a), then the magnitude of the frictional torque about the center of the motor (O) that must be overcome before rotor movement can be initiated is

$$T_s = F_s \ell_1 \,, \tag{38.2}$$

where ℓ_1 is the distance OP_1, which is assumed to be the average distance from the center at which the friction force F_s occurs. Now, consider an AFM tip moving against a rotor arm in a direction perpendicular to the long axis of the cantilever beam (the rotor arm edge closest to the tip is parallel to the long axis of the cantilever beam), as shown in Fig. 38.71a. When the tip encounters the rotor at point P_2, the tip will twist, generating a lateral force between the tip and the rotor (event A in Fig. 38.71b). This reaction force will generate a torque about the center of the motor. Since the tip is trying to move further in the direction shown, the tip will continue to twist to a maximum value at which the lateral force between the tip and the rotor becomes high enough such that the resultant torque T_f about the center of the motor equals the static friction torque T_s. At this point, the rotor will begin to rotate, and the twist of the cantilever decreases sharply (event B in Fig. 38.71b). The twist of the cantilever is measured in the AFM as a change in the lateral deflection signal (in volts), which is the underlying concept of friction force microscopy (FFM). The change in the lateral deflection signal corresponding to the above-mentioned events as the tip approaches the rotor is shown schematically in Fig. 38.71c. The value of the peak V_f is a measure of the force exerted on the rotor by the tip just before the static friction torque is matched and the rotor begins to rotate.

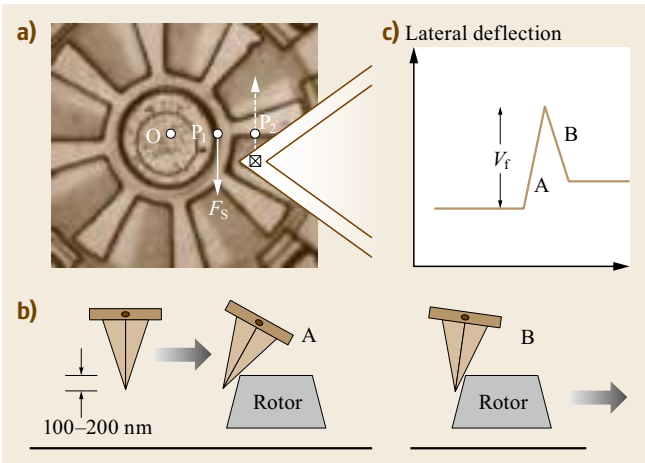


Fig. 38.71 (a) Schematic of the technique used to measure the force, F_s, required to initiate rotor movement using an AFM/FFM. **(b)** As the tip is pushed against the rotor, the lateral deflection experienced by the rotor due to the twisting of the tip prior to rotor movement is a measure of static friction force, F_s, of the rotors. **(c)** Schematic of lateral deflection expected from the above-mentioned experiment. The peak V_f is related to the state of the rotor (after [38.267])

Using this technique, the viability of PFPE lubricants for micromotors has been investigated, and the effect of humidity on the friction forces of unlubricated and lubricated devices was studied as well. Figure 38.72 shows static friction forces, normalized over the weight of the rotor, of unlubricated and lubricated micromotors as a function of rest time and relative humidity. Rest time here is defined as the time elapsed between the first experiment conducted on a given motor (solid symbol at time zero) and subsequent experiments (open symbols). Each open symbol data point is an average of six measurements. It can be seen that for the unlubricated motor and the motor lubricated with a bonded layer of Z-DOL(BW), the static friction force was highest for the first experiment, and then dropped to an almost constant level. In the case of the motor with an as-is mobile layer of Z-DOL, the values remained very high up to 10 days after lubrication. In all cases, there was negligible difference in the static friction force at 0 and 45% RH. At 70% RH, the unlubricated motor exhibited a substantial increase in the static friction force, while the motor with bonded Z-DOL showed no increase due to the hydrophobicity of the lubricant layer. The motor with an as-is mobile layer of the lubricant showed consistently high values of static friction force that varied little with humidity.

Figure 38.73 summarizes static friction force data for two motors, M1 and M2, along with schematics of the meniscus effects for the unlubricated and lubricated surfaces. Capillary condensation of water vapor from the environment resulted in the formation of meniscus bridges between the contacting and near-contacting asperities of two surfaces in close proximity to each other, as shown in Fig. 38.73. For unlubricated surfaces, more menisci were formed at higher humidity resulting in

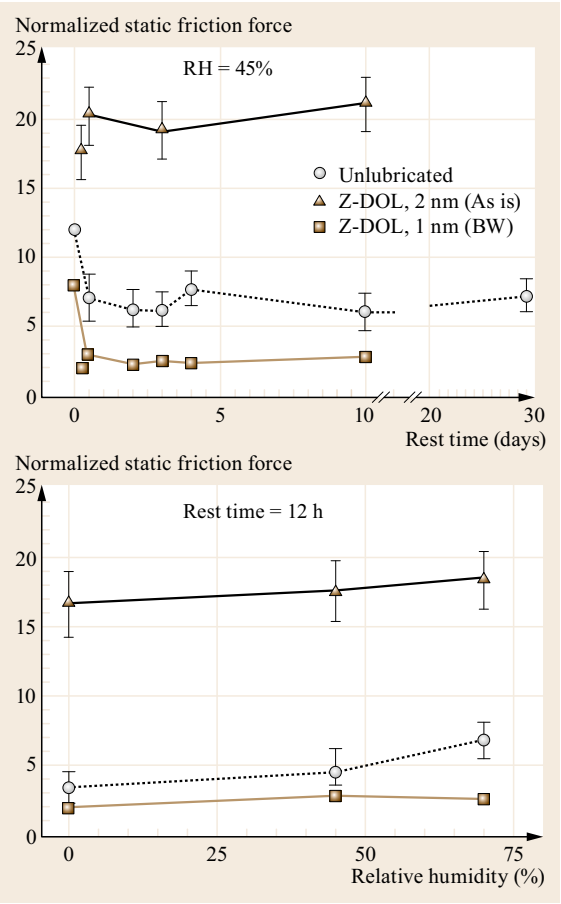


Fig. 38.72 Static friction force values of unlubricated motors and motors lubricated using PFPE lubricants, normalized over the rotor weight, as a function of rest time and relative humidity. Rest time is defined as the time elapsed between a given experiment and the first experiment in which motor movement was recorded (time 0). The motors were allowed to sit at a particular humidity for 12 hours prior to measurements (after [38.267])

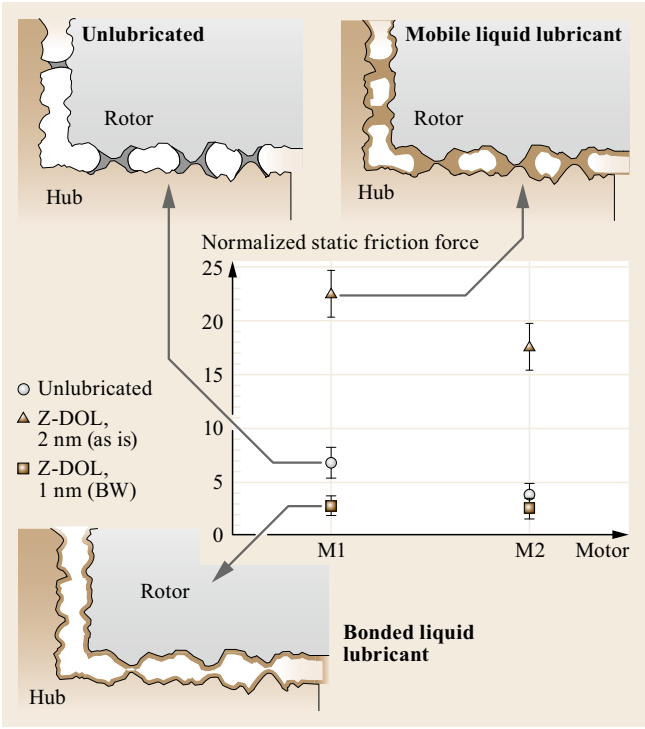


Fig. 38.73 Summary of effect of liquid and solid lubricants on static friction force of micromotors. Despite the hydrophobicity of the lubricant used (Z-DOL), a mobile liquid lubricant (Z-DOL as-is) leads to very high static friction force due to increased meniscus forces whereas a solid-like lubricant (bonded Z-DOL, BW) appears to provide some amount of reduction in static friction force

higher friction force between the surfaces. The formation of meniscus bridges is supported by the fact that the static friction force for unlubricated motors increases at high humidity (Fig. 38.73). Solid bridging may occur near the rotor–hub interface due to silica residues after the first etching process. In addition, the drying process after the final etch can result in liquid bridging formed by the drying liquid due to meniscus force at these areas [38.91, 92, 268, 269]. Therefore, the initial static friction force will be quite high, as evidenced by the solid data points in Fig. 38.73. Once the first movement of the rotor permanently breaks these solid and liquid bridges, the static friction force of the motors will drop (as seen in Fig. 38.73) to a value dictated predominantly by the adhesive energies of rotor and hub surfaces, the real area of contact between these surfaces and meniscus forces due to water vapor in the air. At that point, the effect of lubricant films can be observed. Lubrication with a mobile layer, even a thin one, results in very high static friction forces due to meniscus effects of the lubricant liquid itself at and near the contact regions. It should be noted that a motor submerged in a liquid lubricant would result in a fully flooded lubrication regime. In this case, there is no meniscus contribution, and only the viscous contribution to the friction forces would be relevant. However, submerging the device in a lubricant may not be a practical method. A solid-like hydrophobic lubricant layer (such as bonded Z-DOL) results in favorable friction characteristics for the motor. The hydrophobic nature of the lubricant inhibits meniscus formation between the contact surfaces and maintains low friction even at high humidity (Fig. 38.73). This suggests that solid-like hydrophobic lubricants are ideal for lubrication of MEMS, while mobile lubricants result in increased values of static friction force.

38.8.5 Mechanisms Associated with Observed Stiction Phenomena in Digital Micromirror Devices (DMDs) and Nanomechanical Characterization

DMDs are used in digital projection displays, as described earlier. The DMD has a layered structure, consisting of an aluminum alloy micromirror layer, yoke and hinge layer, and metal layer on a CMOS memory array [38.5, 26, 27]. A blown-up view of the DMD and the corresponding AFM surface height images are presented in Fig. 38.74 [38.117]. Single-layered aluminum alloy films are used for the construction of micromirrors. Sometimes these also are used for the construction of hinges, spring tips, and landing sites. The aluminum alloy films are overwhelmingly comprised of aluminum; trace elements (including Ti and Si) are present to suppress contact spiking and electromigration, which may occur if current densities become high during electrostatic operation. Multilayered sputtered SiO_2/TiN/Al alloy films are now generally used for the landing-site structure to minimize refraction throughout the visible region of the electromagnetic spectrum in order to increase the contrast ratio in projection display systems [38.282, 283]. These multilayered films generally also are used for hinges and spring tips. A low surface energy SAM is maintained on the surfaces of the DMD, which is packaged in a hermetic environment to minimize stiction during contact between the spring tip and the landing site. A SAM of perfluorinated n-alkanoic acid ($C_nF_{2n-1}O_2H$) (e.g., perfluorodecanoic acid or PFDA, $CF_3(CF_2)_8COOH$) applied by the vapor phase deposition process is used. A getter strip of PFDA is included inside the hermetically sealed enclosure containing the chip, which acts as a reservoir in order to maintain a PFDA vapor within the package.

In order to identify a stuck mirror and characterize its nanotribological properties, the chip was scanned using an AFM [38.117]. It was found that it is hard to tilt the stuck micromirror back to its normal position by adding a normal load at the rotatable corner of the micromirror; thus it is called a *hard* stuck micromirror. An example of a stuck micromirror is shown in Fig. 38.75a. Once the stuck micromirror was found, the region was repeatedly scanned at a large normal load, up to 300 nN. After several scans, the stuck micromirror was removed. Once the stuck micromirror was removed, the surrounding micromirrors also could be removed by continuous scanning under a large normal load (Fig. 38.75a bottom row). The adhesive force of the landing site underneath the stuck micromirror and the normal micromirror are presented in Fig. 38.75b. It clearly indicates that the landing site underneath the stuck micromirror had much larger adhesion. 1 μm × 1 μm view of landing sites under stuck and normal micromirrors are also shown in Fig. 38.75b. The landing site under the stuck micromirror had an apparent U-shaped wear mark, which was surrounded by a smeared area.

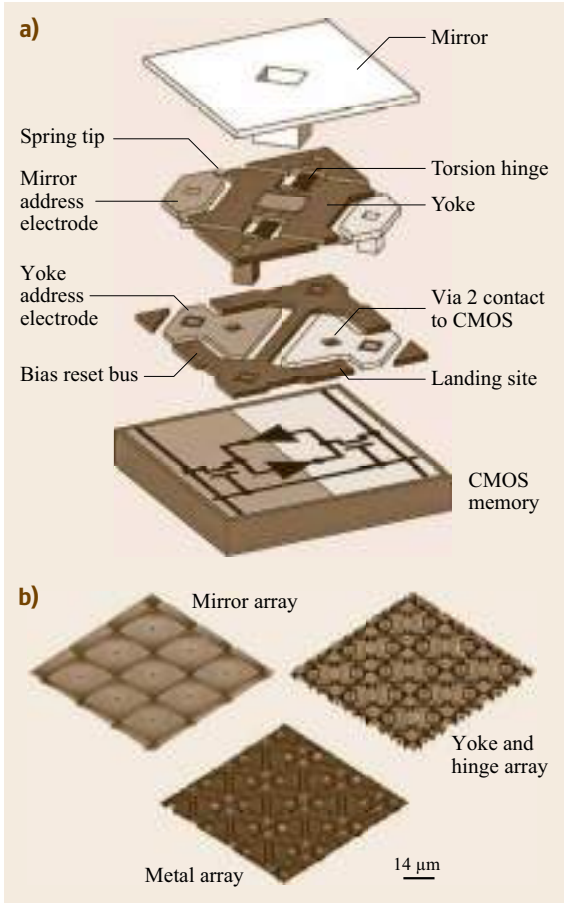


Fig. 38.74 (a) Exploded view of a DMD pixel and **(b)** AFM surface height images of various arrays. The DMD layers were removed by an ultrasonic method (after [38.117])

Liu and *Bhushan* [38.117] calculated contact stresses to examine whether the stresses were high enough to cause wear at the spring tip landing site interface. The calculated contact stress value was about 33 MPa, which is substantially lower than the hardness. Therefore, not much plastic deformation and consequent wear was expected. A wear mark was only found on a very few landing sites on the DMD, which means that the SAM coating generally can endure such high contact stresses. Based on data reported in the literature, coverage for vapor-deposited SAMs is expected to be about 97%. The bond strength of the molecules

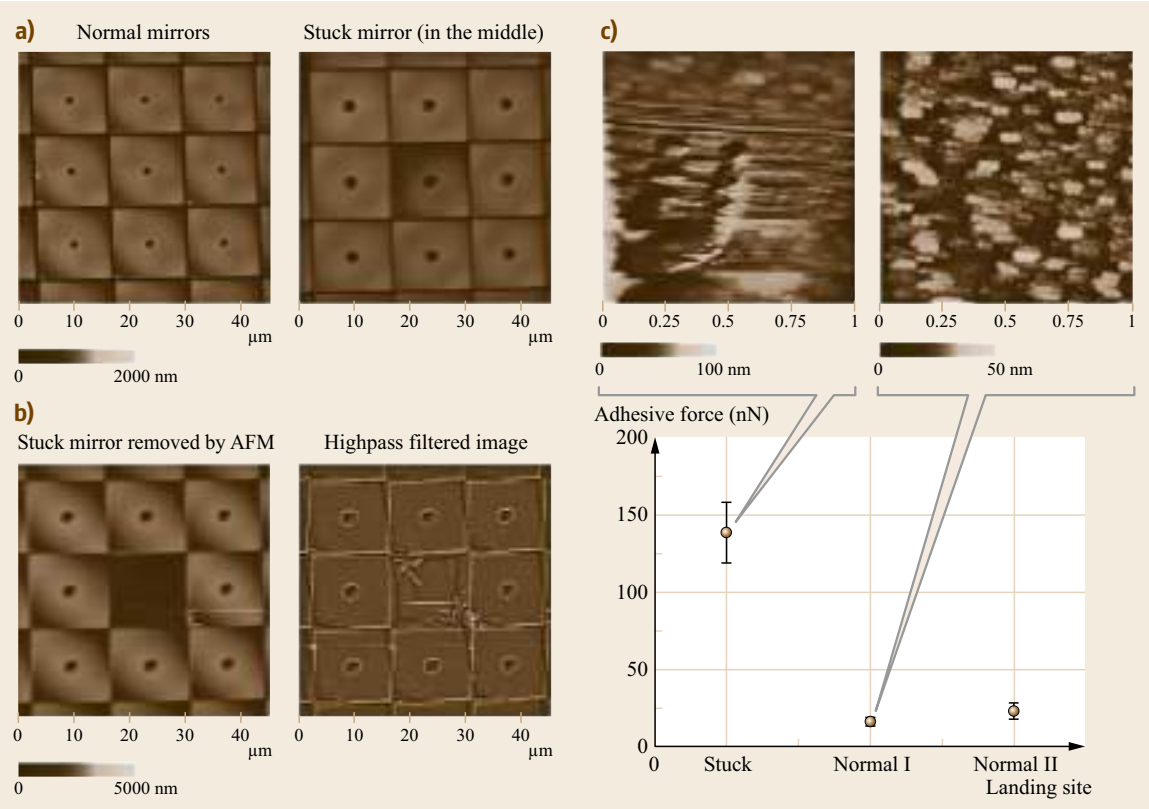


Fig. 38.75 **(a)** shows AFM surface height images of a stuck micromirror surrounded by eight normal micromirrors. **(b)** *Left image* shows the stuck micromirror that was removed by an AFM tip after repeated scanning at high normal load. The *right image* row presents the high-pass filtered image showing that the residual hinge that sits underneath the removed micromirror is clearly observed. **(c)** AFM surface height images and adhesive forces of the landing sites underneath the two normal micromirrors and the stuck micromirror (after [38.117])

close to the boundary of the uncovered sites is expected to be weak. Thus, the uncovered sites and the adjacent molecules are referred to as defects in the SAM coating. Occasionally, if contact occurs at the defect sites, the large cyclic stress may be close to the critical load, and lead to the initial delamination of the SAM coating at the interface. Continuous contact leads to the formation of a high surface-energy surface by exposure of the fresh substrate and formation of SAM fragments. This eventually leads to an increase in stiction by the formation of large menisci. Once this happens, the stress at the contact area is increased, which would accelerate the wear. Based on this hypothesis, suggested mechanisms for the wear and stiction of the landing site are summarized in Fig. 38.76. Wear initiates at the defect sites and consequently high stiction can result in high wear. Improving the coverage and wear resistance of SAM coatings could enhance the yield of DMD.

In some cases, the micromirrors are not fully stuck, and can be moved by applying a load at the rotatable corner of the micromirror with a discontinuous motion, thus called *soft* stiction. Soft stuck micromirrors studied by *Liu* and *Bhushan* [38.118] were identified in quality inspection. These micromirrors encountered slow transition from one end to the other end $(+1/-1)$. Figure 38.77 shows the AFM surface height images of a location showing a stuck mirror (S) and surrounding normal micromirrors N_i ($i = 1, 2$, and 3). Surprisingly, the images of the stuck and normal micromirror array are almost the same. On the micromirrors of interest, a tilting test was performed at the corner of the micromirrors; the rotatable direction of the microarray is indicated by an arrow bar in Fig. 38.77. The load displacement curve for the stuck micromirror is presented in Fig. 38.78. The curve is not smooth and appears serrated. It clearly indicated that, although the S micromirror can be rotated, it rotates with hesitation. In regimes 1 and 2, as marked in Fig. 38.78, the slopes are much higher. In order to understand the occurrence mechanisms of stiction, stiction of the landing sites

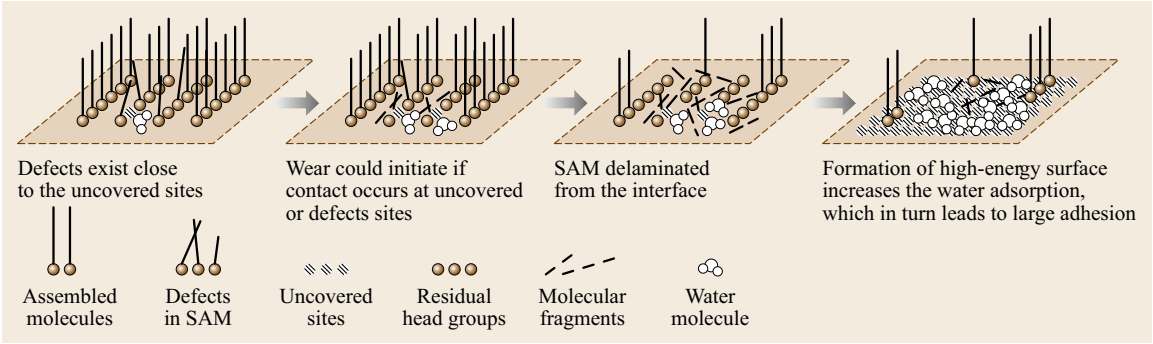


Fig. 38.76 Suggested mechanisms for wear and stiction (after [38.117])

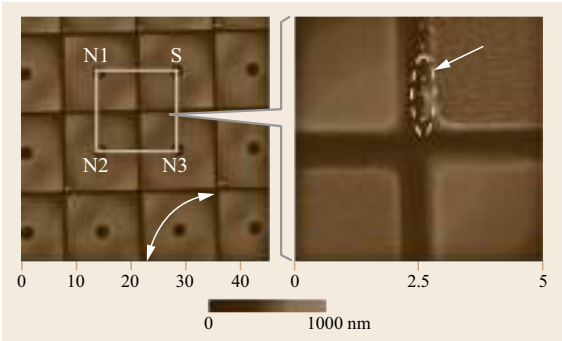


Fig. 38.77 AFM surface height images of normal micromirrors and a soft-stuck micromirror. The soft-stuck micromirror was labeled as S, and the normal micromirrors studied are labeled as N1, N2, and N3 (after [38.118])

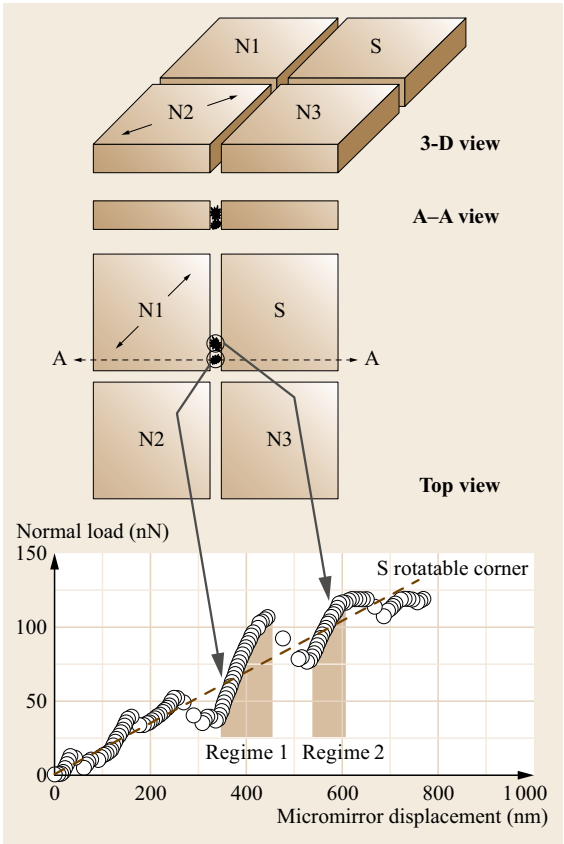


Fig. 38.78 Load–displacement curve obtained on the rotatable corner of S micromirror and schematic to illustrate the suggested mechanism for soft stiction occurrence

of normal and stuck mirrors were measured. Unlike a hard-stuck mirror, adhesive forces of soft stuck and normal mirrors were comparable, which suggests that the SAM coating is intact with the soft-stuck mirror. It was found that a high normal load (about 900 nN) and on the order of couple of hundred scans were required to remove the soft-stuck micromirrors by an AFM. For comparison, only about 300 nN and about ten scans were required to remove the hard-stuck mirror. After careful examination of the AFM images of the micromirror sidewalls in Fig. 38.77 (bottom left), it is noted that there are contaminant particles attached to the sidewalls of the S mirror. Therefore, it is believed that during the tilting test for the S micromirror (see schematic in Fig. 38.78), a sharper slope regime will occur in the displacement curve. Extra force is required to overcome the resistance that is induced by the sidewall contamination particles. This is believed to be the reason for the slow transition of the micromirror during quality inspection.

Finally, nanomechanical characterization of various layers used in the construction of landing sites, hinge, and micromirror materials have been measured by *Wei* et al. [38.282, 283]. Bending and fatigue studies of the hinge have been carried out by *Liu* and *Bhushan* [38.284] and *Bhushan* and *Liu* [38.285] to measure stiffness and fatigue properties. For these studies, the micromirror was removed. During removal,

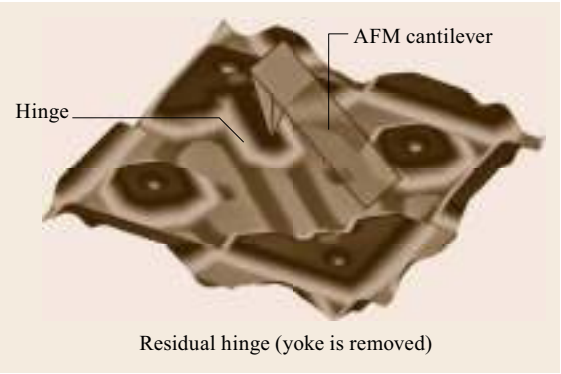


Fig. 38.79 AFM surface height image of the residual hinge and schematic diagram of relative position of the hinge and AFM tip during the nanoscale bending and fatigue tests. The tip is located at the free end of the hinge (after [38.284]) ◄

the micromirror/yoke structure was removed simultaneously, leaving the hinge mounted on one end of the array; see Fig. 38.79. The stiffness of the Al hinge was reported to be comparable to the stiffness of bulk Al. The Al hinge exhibited higher modulus than the SiO_2 hinge. The fatigue properties depended upon the preparation of the hinge for testing.

38.9 Conclusion

The field of MEMS/NEMS and bioMEMS/bioNEMS has expanded considerably since its inception in the late 1980s. Large surface-to-volume ratios of the devices results in very high surface forces, such as adhesion and friction/stiction, which seriously undermine the performance and reliability of the devices. There is a need for a fundamental understanding of adhesion, friction, stiction, wear, and lubrication, and the role of contamination and environment, all on the nanoscale. Most mechanical properties are known to be scale dependent. Therefore, the properties of nanostructures need to be measured. Using AFM-based techniques, researchers have conducted nanotribological and nanomechanical studies of materials and devices. In addition, component-level testing has also been carried out to aid in better understanding of the observed nanotribological phenomena in MEMS/NEMS.

Macroscale and microscale tribological studies of silicon and polysilicon films have been performed. The effect of doping and oxide films and environment on the tribological properties of these popular MEMS/NEMS materials have also been studied. SiC film is found to be a good candidate material for use in high-temperature MEMS/NEMS devices. Perfluoroalkyl SAMs and bonded perfluoropolyether lubricants are well suited for lubrication of micro/nanodevices under a range of environmental conditions. DLC coatings can also be used for low friction and wear.

Nano-objects exhibit enhanced mechanical, electrical, magnetic, and chemical properties. Many nano-objects also exhibit low friction and wear properties due to their size, shape, and unique properties. Nano-objects of interest include Au nanoparticles, MoS_2 and WS_2 MWNTs, and CNHs. Nanoscale friction and wear properties have been investigated in single nano-object contact and multiple nano-object contact modes using an AFM. All nano-objects exhibit low friction and wear. Lower friction forces occur in submerged-in-liquid environments, as compared to dry environments.

For bioMEMS/bioNEMS, nanoscale adhesion, friction, and wear of a number of protein layers on polystyrene surfaces have been investigated. Adhesion between biological molecular layers and the substrate, and friction and wear of biological layers can be important. Adhesion of biomolecules on Si substrate surfaces for various bioMEMS applications can be improved by nanopatterning and the chemical linker method. Friction and wear mechanisms of protein layers have been studied. Trajectory of submicron particles in the bloodstream used for therapeutic and diagnostics purposes needs to be optimized in order for them to bond to the target sites on endothelium walls. An analytical model has been developed for this purpose.

Surface roughness measurements of micromachined polysilicon surfaces have been made using an AFM. The roughness distribution on surfaces is strongly dependent upon the fabrication process. Adhesion and friction of microstructures can be measured using a novel microtriboapparatus. Adhesion and friction measurements on silicon-on-silicon confirm AFM measurements that show hexadecane thiol and bonded perfluoropolyether films exhibit superior adhesion and friction properties. Static friction force measurements of micromotors have been performed using an AFM. The forces are found to vary considerably with humidity. A bonded layer of perfluoropolyether lubricant is found to satisfactorily reduce the friction forces in the micromotor. Tribological failure modes of digital

micromirror devices are either *hard* stiction or *soft* stiction. In hard stiction, the tip on the yoke remains stuck with the landing site underneath. The mechanism responsible for the hard stiction is the localized damage to the SAM on the landing site. In soft stiction, the mirror-yoke assembly rotates with hesitation. The mechanism responsible for the soft stiction is contaminant particles present at the mirror sidewalls.

AFM/FFM-based techniques show the capability to study and evaluate nanotribology and nanomechanics related to MEMS/NEMS and bioMEMS/bioNEMS devices.

38.A Appendix: Micro/Nanofabrication Techniques

Micro/nanofabrication techniques include top-down methods, in which one builds down from the large to the small, and bottom-up methods, in which one builds up from the small to the large.

38.A.1 Top-Down

The top-down fabrication methods used in the construction of MEMS/NEMS include lithographic and nonlithographic techniques to produce micro- and nanostructures. The lithographic techniques fall into three basic categories: bulk micromachining, surface micromachining, and LIGA (a German acronym for Lithographie Galvanoformung Abformung), which translates to lithography, electroplating, and molding. The first two approaches, bulk and surface micromachining, mostly use planar photolithographic fabrication processes developed for semiconductor devices to produce two-dimensional (2-D) structures [38.7, 13, 286–289]. The various steps involved in these two fabrication processes are shown schematically in Fig. 38.80. Bulk micromachining employs anisotropic etching to remove sections through the thickness of a single-crystal silicon wafer, typically 250–500 μm thick. Bulk micromachining is a proven high-volume production process and is routinely used to fabricate microstructures such as accelerometers, pressure sensors, and flow sensors. In surface micromachining, structural and sacrificial films are alternatively deposited, patterned, and etched to produce a freestanding structure. These films are typically made of low-pressure chemical vapor deposition (LPCVD) polysilicon film with 2–20 μm thickness. Surface micromachining is used to produce sensors, actuators, micromirror arrays, motors, gears, and grippers. The resolution in photolithography is dependent upon the wavelength of light. A commonly used light source is an argon fluoride excimer laser with 193 nm wavelength (ultraviolet or UV) used in patterning 90 nm lines and spaces. Deep-UV wavelengths, x-ray lithography, electron beam (e-beam) lithography, focused ion beam lithography, maskless lithography, liquid-immersion lithography, and STM writing by removing material atom by atom are some of the recent developments for sub-100 nm patterning.

The fabrication of nanostructures such as nanochannels with sub-10 nm resolution can be accomplished through several routes: e-beam lithography and sacrificial layer lithography (SLL). The process for e-beam

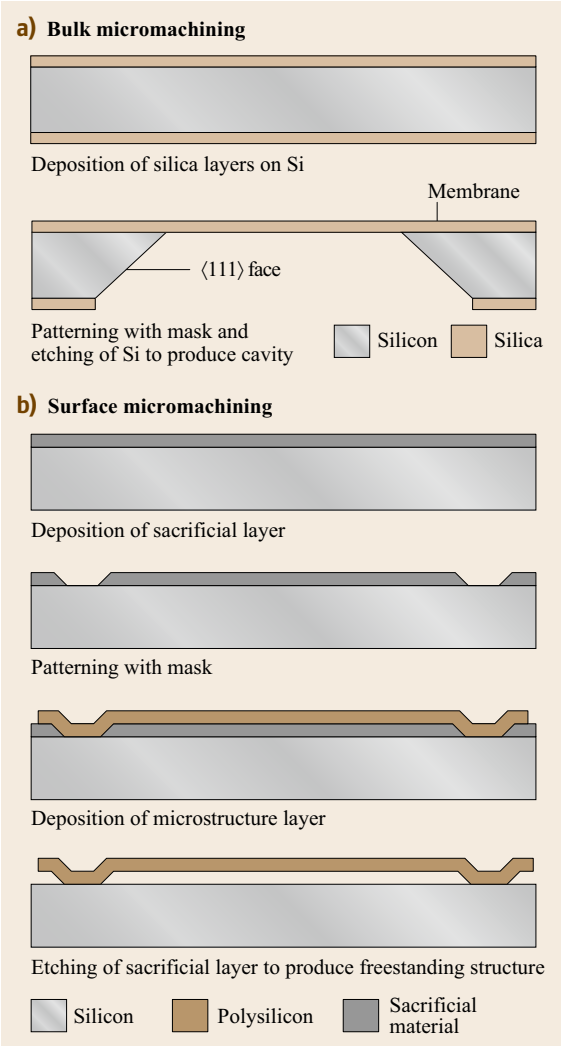


Fig. 38.80a,b Schematic of process steps involved in (a) bulk micromachining and (b) surface micromachining fabrication of MEMS

lithography is a finely focused electron beam that is exposed over a resist surface, and the exposure duration and location is controlled with the use of a computer [38.290, 291]. When the resist is exposed to the electron beam, the electrons either break or join the molecules in the resist, so the local characteristics are changed in such a way that further processes can either remove the exposed part (positive resist) or remove the unexposed part (negative resist). The resist material determines if the molecules will either break or join together; thus, determining if a positive or negative image is produced. E-beam lithography can be used either to create photolithographic masks for replication or to create the devices directly. The masks that are created can be used for either optical or x-ray lithography. One limitation of e-beam lithography is that throughput is drastically reduced since a single electron beam is used to create the entire exposure pattern on the resist. While this technique is slower than conventional lithographic techniques, it is ideal for prototype fabrication because no masks are required.

In SLL process, the use of a sacrificial layer allows the direct control of nanochannel dimensions so long as there exists a method for removing the sacrificial layer with absolute selectivity to the structural layers. A materials system with such selectivity is the silicon/silicon oxide system used widely in the microfabrication of MEMS devices. The use of sidewall deposition of the sacrificial layer and subsequent etching allows for the fabrication of high-density nanochannels for biomedical applications. It is based on surface micromachining [38.83]. Figure 38.81 shows a schematic of the process steps in SLL based on *Hansford* et al.'s [38.83] work on fabrication of polysilicon membranes with nanochannels. As with all the membrane protocols, the first step in the fabrication is the etching of the support ridge structure into the bulk silicon substrate. A low-stress silicon nitride (LSN or simply nitride), which functions as an etch stop layer, is then deposited using LPCVD. The base structural polysilicon layer (base layer) is deposited on top of the etch stop layer. The plasma etching of holes in the base layer is what defines the shape of the pores. The buried nitride etch stop acts as an etch stop for the plasma etching of a polysilicon base layer. After the pore holes are etched through the base layer, the pore sacrificial thermal oxide layer is grown on the base layer. The basic requirement of the sacrificial layer is the ability to control the thickness with high precision across the entire wafer. Anchor points are defined in the sacrificial oxide layer to mechanically connect the base layer with the plug layer (necessary to maintain the pore spacing between layers). This is accomplished by using the same mask shifted from the pore holes. This produces anchors in one or two corners of each pore hole, which provides the desired connection between the structural layers while opening as much pore area as possible. After the anchor points are etched through the sacrificial oxide, the plug polysilicon layer is deposited (using LPCVD) to fill in the holes. To open the pores at the surface, the plug layer is planarized using chemical mechanical polishing (CMP) down to the base layer, leaving the final structure with the plug layer only in the pore hole openings. As the silicon wafer is ready for release, a protective nitride layer is deposited on the wafer (completely covering both sides of the wafer). The backside etch windows are etched in the protective layer, exposing the silicon wafer in the desired areas,

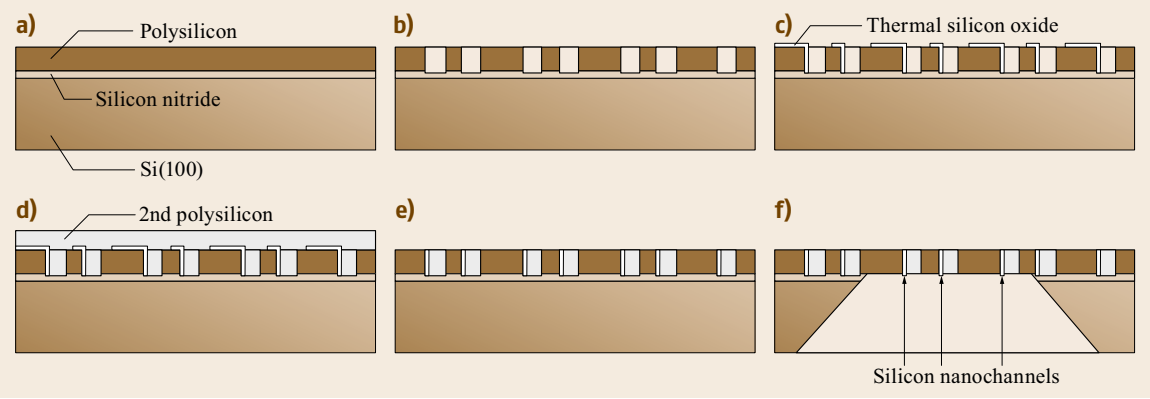


Fig. 38.81a–f Schematic of process steps involved in sacrificial layer lithography: **(a)** growth of silicon nitride layer (etch stop) and base polysilicon deposition, **(b)** hole definition in base, **(c)** growth of thin sacrificial oxide and patterning of anchor points, **(d)** deposition of plug polysilicon, **(e)** planarization of plug layer, and **(f)** deposition and patterning of protective nitride layer through etch, followed by etching of protective, sacrificial and etch layers and final release of the structure in HF (after [38.83])

and the wafer is placed in a KOH bath to etch. After the silicon wafer is completely removed up to the membrane (as evidenced by the smooth buried etch stop layer), the protective, sacrificial, and etch stop layers are removed by etching in concentrated HF. Etching of the sacrificial layer in polysilicon film defines nanochannels.

The LIGA process is based on the combined use of x-ray lithography, electroplating, and molding processes. X-rays produced by synchrotron radiation are used to prepare the mold. The steps involved in the LIGA process are shown schematically in Fig. 38.82. LIGA is used to produce high-aspect-ratio MEMS (HARMEMS) devices that are up to 1 mm in height and only a few microns in width or length [38.292]. The LIGA process yields very sturdy 3-D structures due to their increased thickness. One of the limitations of silicon microfabrication processes originally used for fabrication of MEMS devices is the lack of suitable materials that can be processed. With LIGA, a variety of nonsilicon materials such as metals, ceramics and polymers can be processed.

Nonlithographic micromachining processes, primarily in Europe and Japan, are also being used for fabrication of millimeter-scale devices using direct material microcutting or micromechanical machining (such as microturning, micromilling, and microdrilling) or removal by energy beams (such as microspark erosion, focused ion beam, laser ablation, and laser polymerization) [38.7, 293]. Hybrid technologies including LIGA and high-precision micromachining techniques have been used to produce miniaturized motors, gears, actuators, and connectors [38.98, 100, 101, 294]. These millimeter-scale devices may find more immediate applications.

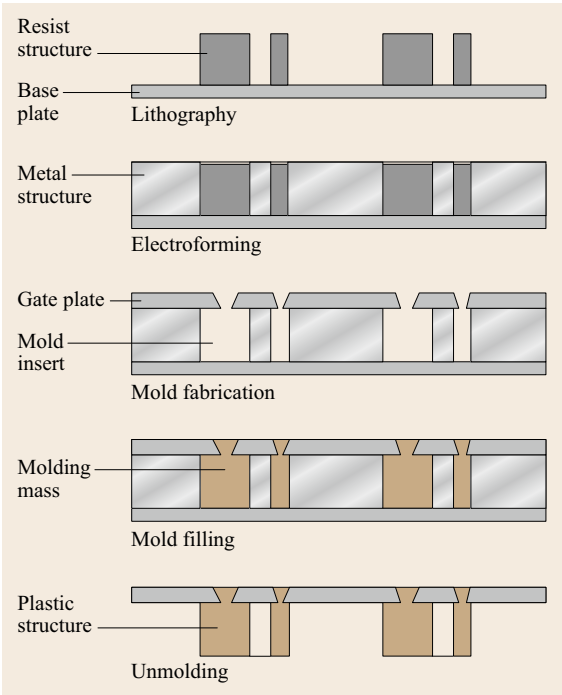


Fig. 38.82 Schematic of process steps involved in LIGA fabrication of MEMS

A micro/nanofabrication technique, so-called *soft lithography*, is a nonlithographic technique [38.63, 69, 295, 296] in which a master or mold is used to generate patterns, defined by the relief on its surface, on polymers by replica molding [38.297], hot embossing (nanoimprint lithography) [38.298], or by contact printing (known as microcontact printing or μCP) [38.299]. Soft lithography is faster, less expensive, and more suitable for most biological applications than glass or silicon micromachining. Polymers have established an important role in bioMEMS/bioNEMS because of their reduced cost. The use of polymers also offers a wide range of material properties to allow tailoring of biological interactions for improved biocompatibility. Polymer fabrication is believed to be about an order of magnitude cheaper than silicon fabrication.

Replica molding is the transfer of a topographic pattern by curing or solidifying a liquid precursor against the original patterned mold. The mold or stamp is generally made of a two-part polymer (elastomer and curing agent), such as poly(dimethylsiloxane) (PDMS) from a photolithographically generated photoresist master. Solvent-based embossing or imprinting uses a solvent to restructure a polymer film. Hot embossing, also called nanoimprint lithography, usually refers to the transfer of pattern from a micromachined quartz or metal master to a pliable plastic sheet. Heat and high pressure allow the plastic sheet to become imprinted. These sheets can then be bonded to various plastics such as polymethyl methacrylate (PMMA). Nanoimprint lithography can produce patterns on a surface having 10 nm resolution. Contact printing uses a patterned stamp to transfer ink (mostly self-assembled monolayer) onto a surface in a pattern defined by the raised regions of a stamp. These techniques can be used to pattern line widths as small as 60 nm.

Replica molding is commonly used for mass-produced, disposable plastic micro/nanocomponents, for example micro/nanofluidic chips, generally made of PDMS and PMMA [38.296, 300] and more flexible in choice of materials for construction than conventional photolithography.

To assemble microsystems, microrobots are used. Microrobotics include building blocks, such as steering links, microgrippers, conveyor systems, and locomotive robots [38.17].

38.A.2 Bottom-Up Fabrication (Nanochemistry)

Bottom-up approaches (from small to large) largely rely on nanochemistry [38.37, 38, 40–45]. The bottom-up approach includes chemical synthesis, the spontaneous *self-assembly* of molecular clusters (molecular self-assembly) from simple reagents in solution or biological molecules as building blocks to produce three-dimensional nanostructures as done by nature, quantum dots (nanocrystals) of arbitrary diameter (about $10–10^5$ atoms), molecular beam epitaxy (MBE) and organometallic vapor phase epitaxy (OMVPE) to create specialized crystals one atomic or molecular layer at a time, and manipulation of individual atoms by a scanning tunneling microscope or an AFM or atom optics. The self-assembly must be encoded, that is, one must be able to precisely assemble one object next to another to form a designed pattern. A variety of nonequilibrium plasma chemistry techniques are also used to produce layered nanocomposites, nanotubes, and nanoparticles. Nanostructures can also be fabricated using mechanosynthesis with proximal probes.

References

38.1 B. Bhushan: *Encyclopedia of Nanotechnology*, 2nd edn. (Springer International, Cham 2016)

38.2 R.J. Chen, H.C. Choi, S. Bangsaruntip, E. Yenilmez, X. Tang, Q. Wang, Y.L. Chang, H. Dai: An investigation of the mechanisms of electrode sensing of protein adsorption on carbon nanotube devices, J. Am. Chem. Soc. **126**, 1563–1568 (2004)

38.3 D. Srivastava: Computational nanotechnology of carbon nanotubes. In: *Carbon Nanotubes: Science and Applications*, ed. by M. Meyyappan (CRC, Boca Raton 2004) pp. 25–63

38.4 W.G. van der Wiel, S. De Franceschi, J.M. Elzerman, T. Fujisawa, S. Tarucha, L.P. Kouwenhoven: Electron transport through double quantum dots, Rev. Mod. Phys. **75**, 1–22 (2003)

38.5 L.J. Hornbeck: A Digital Light Processing update – Status and future applications, Proc. Soc. Photo-Opt. Eng. **3634**, 158–170 (1999)

38.6 B. Bhushan: Governance, policy, and legislation of nanotechnology: A perspective, Microsyst. Technol. **21**, 1137–1155 (2015)

38.7 M. Madou: *Fundamentals of Microfabrication and Nanotechnology*, 3rd edn. (CRC, Boca Raton 2011)

38.8 B. Bhushan, D. Luo, S.R. Schricker, W. Sigmund, S. Zauscher: *Handbook of Nanomaterials Properties*, Vol. I & II (Springer, Berlin, Heidelberg 2014)

38.9 https://www.nano.gov

38.10 R.S. Muller, R.T. Howe, S.D. Senturia, R.L. Smith, R.M. White: *Microsensors* (IEEE, New York 1990)

38.11 I. Fujimasa: *Micromachines: A New Era in Mechanical Engineering* (Oxford Univ. Press, Oxford 1996)

38.12 W.S. Trimmer (Ed.): *Micromachines and MEMS, Classic and Seminal Papers to 1990* (IEEE, New York 1997)

38.13 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

38.14 G.T.A. Kovacs: *Micromachined Transducers Sourcebook* (McGraw-Hill, Boston 1998)

38.15 S.D. Senturia: *Microsystem Design* (Kluwer Academic, Boston 2000)

38.16 M. Elwenspoek, R. Wiegerink: *Mechanical Microsensors* (Springer, Berlin, Heidelberg 2001)

38.17 M. Gad-el-Hak: *The MEMS Handbook* (CRC, Boca Raton 2002)

38.18 T.R. Hsu: *MEMS and Microsystems: Design and Manufacture* (McGraw-Hill, Boston 2002)

38.19 A. Hierlemann: *Integrated Chemical Microsensor Systems in CMOS Technology* (Springer, Berlin, Heidelberg 2005)

38.20 T.A. Core, W.K. Tsang, S.J. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. **36**(10), 39–47 (1993)

38.21 J. Bryzek, K. Peterson, W. McCulley: Micromachines on the march, IEEE Spectr. **5**, 20–31 (1994)

38.22 J.S. Aden, J.H. Bohorquez, D.M. Collins, M.D. Crook, A. Garcia, U.E. Hess: The third-generation HP thermal inkjet printhead, HP Journal **45**(1), 41–45 (1994)

38.23 H. Le: Progress and trends in ink-jet printing technology, J. Imaging Sci. Technol. **42**, 49–62 (1998)

38.24 R. Baydo, A. Groscup: Getting to the heart of ink jet: Printheads, Beyond Recharg. **10**(May), 10–12 (2001)

38.25 E.R. Lee: *Microdrop Generation* (CRC, Boca Raton 2003)

38.26 L.J. Hornbeck, W.E. Nelson: Bistable deformable mirror device, OSA Tech. Dig. Ser. **8**, 107–110 (1988)

38.27 L.J. Hornbeck: The DMD projection display chip: A MEMS-based technology, MRS Bulletin **26**, 325–328 (2001)

38.28 K. Suzuki: Micro electro mechanical systems (MEMS) micro-switches for use in DC, RF, and optical applications, Jpn. J. Appl. Phys. **41**, 4335–4339 (2002)

38.29 V.M. Lubecke, J.C. Chiao: MEMS technologies for enabling high frequency communication cicuits. In: *Proc. IEEE 4th Int. Conf. Telecom. Modern Satell. Cable Broadcast. Serv.* (IEEE, New York 1999) pp. 1–8

38.30 C.R. Giles, D. Bishop, V. Aksyuk: MEMS for lightwave networks, MRS Bulletin **4**, 328–329 (2001)

38.31 A. Hierlemann, O. Brand, C. Hagleitner, H. Baltes: Microfabrication techniques for

chemical/biosensors. In: *Proc. IEEE, Chemical and Biological Microsensors*, Vol. 91, ed. by S. Casalnuovo, R.B. Brown (IEEE, New York 2003) pp. 839–863
38.32 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)
38.33 H. Hamilton: Contact recording on perpendicular rigid media, J. Mag. Soc. Jpn. **15**(S2), 483–481 (1991)
38.34 D.A. Horsley, M.B. Cohn, A. Singh, R. Horowitz, A.P. Pisano: Design and fabrication of an angular microactuator for magnetic disk drives, J. Microelectromech. Syst. **7**, 141–148 (1998)
38.35 T. Hirano, L.S. Fan, D. Kercher, S. Pattanaik, T.S. Pan: HDD tracking microactuator and its integration issues. In: *Proc. ASME Int. Mech. Eng. Congr. Exp.*, Vol. 2, ed. by A.P. Lee, J. Simon, F.K. Foster, R.S. Keynton (ASME, New York 2000) pp. 449–452
38.36 T. Fukuda, F. Arai, L. Dong: Assembly of nanodevices with carbon nanotubes through nanorobotic manipulations, Proc. IEEE **91**, 1803–1818 (2003)
38.37 K.E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)
38.38 G. Timp (Ed.): *Nanotechnology* (Springer, New York 1999)
38.39 M.S. Dresselhaus, G. Dresselhaus, P. Avouris: *Carbon Nanotubes – Synthesis, Structure, Properties and Applications* (Springer, Berlin, Heidelberg 2001)
38.40 E.A. Rietman: *Molecular Engineering of Nanosystems* (Springer, New York 2001)
38.41 W.A. Goddard, D.W. Brenner, S.E. Lyshevski, G.J. Iafrate (Eds.): *Handbook of Nanoscience, Engineering, and Technology* (CRC, Boca Raton 2002)
38.42 H.S. Nalwa (Ed.): *Nanostructured Materials and Nanotechnology* (Academic, San Diego 2002)
38.43 C.P. Poole, F.J. Owens: *Introduction to Nanotechnology* (Wiley, Hoboken 2003)
38.44 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)
38.45 B. Bhushan: Nanotribology and nanomechanics, Wear **259**, 1507–1531 (2005)
38.46 J.A. Stroscio, D.M. Eigler: Atomic and molecular manipulation with a scanning tunneling microscope, Science **254**, 1319 (1991)
38.47 P. Vettiger, J. Brugger, M. Despont, U. Drechsler, U. Dürig, W. Häberle, M. Lutwyche, H. Rothuizen, R. Stutz, R. Widmer, G. Binning: Ultrahigh density, high data-rate NEMS based AFM data storage system, Microelectron. Eng. **46**, 11–27 (1999)
38.48 C. Stampfer, A. Jungen, C. Hierold: Fabrication of discrete carbon nanotube based nanoscaled force sensor. In: *Proc. IEEE Sens. 2004, Vienna* (IEEE, New York 2004) pp. 1056–1059
38.49 C. Hierold: *Carbon Nanotube Devices* (Wiley-VCH, Weinheim 2008)
38.50 B. Bhushan: *Mechanics and Reliability of Flexible Magnetic Media*, 2nd edn. (Springer, New York 2000)
38.51 Anonymous: *International Technology Roadmap for Semiconductors* (2004) http://public.itrs.net/
38.52 H. Lodish, A. Berk, P. Matsudaira, C.A. Kaiser, M. Krieger, M.P. Scott, L. Zipursky, J. Darnell: *Molecular Cell Biology*, 5th edn. (W.H. Freeman, New York 2003)
38.53 A. Koch: *Bacterial Growth and Form* (Springer, Dordrecht 2001)
38.54 D.L. Nelson, M.M. Cox: *Lehninger's Principles of Biochemistry*, 4th edn. (W.H. Freeman, New York 2005)
38.55 P. Gravesen, J. Branebjerg, O. Jensen: Microfluidics – A review, J. Micromech. Microeng. **3**, 168–182 (1993)
38.56 A. van der Berg (Ed.): *Lab-on-a-Chip: Chemistry in Miniaturized Synthesis and Analysis Systems* (Elsevier, Amsterdam 2003)
38.57 S.N. Bhatia, C.S. Chen: Tissue engineering at the micro-scale, Biomed. Microdevices **2**, 131–144 (1999)
38.58 R.P. Lanza, R. Langer, J. Vacanti (Eds.): *Principles of Tissue Engineering*, 2nd edn. (Academic, San Diego 2000)
38.59 E. Leclerc, K.S. Furukawa, F. Miyata, T. Sakai, T. Ushida, T. Fujii: Fabrication of microstructures in photosensitive biodegradable polymers for tissue engineering applications, Biomaterials **25**, 4683–4690 (2004)
38.60 K. Park (Ed.): *Controlled Drug Delivery: Challenges and Strategies* (American Chemical Society, Washington 1997)
38.61 R.S. Shawgo, A.C.R. Grayson, Y. Li, M.J. Cima: BioMEMS for drug delivery, Curr. Opin. Solid State Mater. Sci. **6**, 329–334 (2002)
38.62 P.A. Oeberg, T. Togawa, F.A. Spelman: *Sensors in Medicine and Health Care* (Wiley, New York 2004)
38.63 A. Manz, H. Becker (Eds.): *Microsystem Technology in Chemistry and Life Sciences*, Topics in Current Chemistry, Vol. 194 (Springer, Berlin, Heidelberg 1998)
38.64 J. Cheng, L.J. Kricka (Eds.): *Biochip Technology* (Publishers, Philadelphia 2001)
38.65 M.J. Heller, A. Guttman (Eds.): *Integrated Microfabricated Biodevices* (Marcel Dekker, New York 2001)
38.66 C.L.P. San, E.P.H. Yap (Eds.): *Frontiers in Human Genetics* (World Scientific, Singapore 2001)
38.67 C.H. Mastrangelo, H. Becker (Eds.): *Microfluidics and bioMEMS*, Proc. SPIE, Vol. 4560 (SPIE, Bellingham 2001)
38.68 H. Becker, L.E. Locascio: Polymer microfluidic devices, Talanta **56**, 267–287 (2002)
38.69 D.J. Beebe, G.A. Mensing, G.M. Walker: Physcis and applications of microfluidics in biology, Annu. Rev. Biomed. Eng. **4**, 261–286 (2002)
38.70 J.V. Zoval, M.J. Madou: Centrifuge-based fluidic platforms, Proc. IEEE **92**, 140–153 (2000)
38.71 W.C. Tang, A.P. Lee: Defense applications of MEMS, MRS Bulletin **26**, 318–319 (2001), http://www.darpa.mil/mto/mems

38.72 M.R. Taylor, P. Nguyen, J. Ching, K.E. Peterson: Simulation of microfluidic pumping in a genomic DNA blood-processing cassette, J. Micromech. Microeng. **13**, 201–208 (2003)

38.73 S. Prakash, M. Pinti, B. Bhushan: Theory, fabrication and applications of microfluidic and nanofluidic biosensors, Phil. Trans. Roy. Soc. A **370**, 2269–2303 (2012)

38.74 R. Raiteri, M. Grattarola, M. Butt, P. Skladal: Micromechanical cantilever-based biosensor, Sens. Actuators B **79**, 115–126 (2001)

38.75 B. Bhushan, D.R. Tokachichu, M.T. Keener, S.C. Lee: Morphology and adhesion of biomolecules on silicon based surfaces, Acta Biomater. **1**, 327–341 (2005)

38.76 H.P. Lang, M. Hegner, C. Gerber: Cantilever array sensors, Mater. Today **4**, 30–36 (2005)

38.77 F. Patolsky, C. Lieber: Nanowire nanosensors, Mater. Today **4**, 20–28 (2005)

38.78 T.R. Nicholson III, S. Gupta, X. Wen, H.H. Wu, R. Anisha, P. Casal, K.J. Kwak, B. Bhushan, P.R. Berger, W. Lu, L.J. Brillson, S.C. Lee: Rational enhancement of nanobiotechnological device function illustrated by partial optimization of a protein sensing field effect transistor, Proc. Inst. Mech. Eng. Part N: J. Nanoeng. Nanosyst. **223**, 149–161 (2010)

38.79 S. Gupta, H.-H. Wu, K.J. Kwak, P. Casal, T.R. Nicholson III, X. Wen, R. Anisha, B. Bhushan, P.R. Berger, W. Lu, L.J. Brillson, S.C. Lee: Interfacial design and structure of protein/polymer films on oxidated AlGaN surfaces, J. Phys. D **44**, 034010 (2011)

38.80 P. Casal, X. Wen, S. Gupta, T. Nicholson III, Y. Wang, A. Theiss, B. Bhushan, L. Brillson, W. Lu, S.C. Lee: ImmunoFET feasibility in physiological salt environments, Phil. Trans. Roy. Soc. A **370**, 2474–2488 (2012)

38.81 M. Scott: MEMS and MOEMS for national security applications. In: *Reliability, Testing, and Characterization of MEMS/MOEMS II, Proc. SPIE*, Vol. 4980 (SPIE, Bellingham 2003) pp. xxxvii–xliv

38.82 T.A. Desai, D.J. Hansford, L. Kulinsky, A.H. Nashat, G. Rasi, J. Tu, Y. Wang, M. Zhang, M. Ferrari: Nanopore technology for biomedical applications, Biomed. Devices **2**, 11–40 (1999)

38.83 D. Hansford, T. Desai, M. Ferrari: Nano-scale size-based biomolecular separation technology. In: *Biochip Technology*, ed. by J. Cheng, L.J. Kricka (Harwood Academic, New York 2001) pp. 341–361

38.84 F.J. Martin, C. Grove: Microfabricated drug delivery systems: Concepts to improve clinical benefits, Biomed. Microdevices **3**, 97–108 (2001)

38.85 B. Bhushan, A.V. Kulkarni, W. Bonin, J.T. Wyrobek: Nano/picoindentation measurement using a capacitance transducer system in atomic force microscopy, Philos. Mag. **74**, 1117–1128 (1996)

38.86 B. Bhushan: Nanotribology and nanomechanics of MEMS/NEMS and bioMEMS/bioNEMS materials and devices, Microelectron. Eng. **84**, 387–412 (2007)

38.87 B. Bhushan: Nanotribology and nanomechanics in nano/biotechnology, Phil. Trans. R. Soc. A **366**, 1499–1537 (2008)

38.88 B. Bhushan: *Nanotribology and Nanomechanics – An Introduction*, 4th edn. (Springer International, Cham 2017)

38.89 B. Bhushan (Ed.): *Modern Tribology Handbook*, Vol. 1/2 (CRC, Boca Raton 2001)

38.90 B. Bhushan: Adhesion and stiction: Mechanisms, measurement techniques, and methods for reduction, J. Vac. Sci. Technol. B **21**, 2262–2296 (2003)

38.91 B. Bhushan: *Principles and Applications of Tribology*, 2nd edn. (Wiley, New York 2013)

38.92 B. Bhushan: *Introduction to Tribology*, 2nd edn. (Wiley, New York 2013)

38.93 B. Bhushan, D. Tokachichu, M.T. Keener, S.C. Lee: Nanoscale adhesion, friction, and wear studies of biomolecules on silicon based surfaces, Acta Biomater. **2**, 39–49 (2006)

38.94 Y.C. Tai, L.S. Fan, R.S. Muller: IC-processed micro-motors: Design, technology and testing, Proc. IEEE Micro Electro Mech. Syst. (1989) doi:10.1109/memsys.1989.77950

38.95 S.M. Spearing, K.S. Chen: Micro-gas turbine engine materials and structures, Ceramic Eng. Sci. Proc. **18**, 11–18 (2001)

38.96 L.G. Fréchette, S.A. Jacobson, K.S. Breuer, F.F. Ehrich, R. Ghodssi, R. Khanna, C.W. Wong, X. Zhang, M.A. Schmidt, A.H. Epstein: High-speed microfabricated silicon turbomachinery and fluid film bearings, J. MEMS **14**, 141–152 (2005)

38.97 L.X. Liu, Z.S. Spakovszky: Effect of bearing stiffness anisotropy on hydrostatic micro gas journal bearing dynamic behavior. In: *Proc. ASME Turbo Expo 2005, Reno* (2005), Paper No. GT-2005-68199

38.98 H. Lehr, S. Abel, J. Doppler, W. Ehrfeld, B. Hagemann, K.P. Kamper, F. Michel, C. Schulz, C. Thurigen: Microactuators as driving units for microrobotic systems. In: *Proc. SPIE*, Vol. 2906, ed. by A. Sulzmann (SPIE, Bellingham 1996) pp. 202–210

38.99 M. Mehregany, K.J. Gabriel, W.S.N. Trimmer: Integrated fabrication of polysilicon mechanisms, IEEE Trans. Electron Devices **35**, 719–723 (1988)

38.100 H. Lehr, W. Ehrfeld, B. Hagemann, K.P. Kamper, F. Michel, C. Schulz, C. Thurigen: Development of micro-millimotors, Min. Invas. Ther. Allied Technol. **6**, 191–194 (1997)

38.101 F. Michel, W. Ehrfeld: Microfabrication technologies for high performance microactuators. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 53–72

38.102 E.J. Garcia, J.J. Sniegowski: Surface micromachined microengine, Sens. Actuators A **48**, 203–214 (1995)

38.103 D.M. Tanner, N.F. Smith, L.W. Irwin, W.P. Eaton, K.S. Helgesen, J.J. Clement, W.M. Miller, J.A. Walraven, K.A. Peterson, P. Tangyunyong, M.T. Dugger, S.L. Miller: *MEMS Reliability: Infrastructure, Test Structures, Experiments, and Failure Modes, SAND2000-0091* (Sandia National Laboratories,

Albuquerque 2000), http://www.prod.sandia.gov
38.104 S.S. Mani, J.G. Fleming, J.A. Walraven, J.J. Sniegowski, M.P. de Beer, L.W. Irwin, D.M. Tanner, D.A. LaVan, M.T. Dugger, H. Jakubczak, W.M. Miller: Effect of W coating on microengine performance. In: *Proc. 38th Annual Inter. Reliab. Phys. Symp.* (IEEE, New York 2000) pp. 146–151
38.105 M.G. Hankins, P.J. Resnick, P.J. Clews, T.M. Mayer, D.R. Wheeler, D.M. Tanner, R.A. Plass: Vapor deposition of amino-functionalized self-assembled monolayers on MEMS, Proc. SPIE **4980**, 238–247 (2003)
38.106 J.K. Robertson, K.D. Wise: An electrostatically actuated integrated microflow controller, Sens. Actuators A **71**, 98–106 (1998)
38.107 B. Bhushan: Nanotribology and nanomechanics of MEMS devices. In: *Proc. 9th Annu. Workshop MEMS* (IEEE, New York 1996) pp. 91–98
38.108 R.E. Sulouff: MEMS opportunities in accelerometers and gyros and the microtribology problems limiting commercialization. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 109–120
38.109 J. R. Martin, Y. Zhao: Micromachined Device Packaged to Reduce Stiction, US Patent 5694740 (1997)
38.110 M. Parsons: Design and manufacture of automotive pressure sensors, Sensors **18**, 32–46 (2001)
38.111 G. Smith: The application of microtechnology to sensors for the automotive industry, Microelectronics J **28**, 371–379 (1997)
38.112 L.S. Chang, P.L. Gendler, J.H. Jou: Thermal mechanical and chemical effects in the degradation of the plasma-deposited α-SC:H passivation layer in a multlayer thin-film device, J. Mat. Sci. **26**, 1882–1290 (1991)
38.113 V.A. Aksyuk, F. Pardo, D. Carr, D. Greywall, H.B. Chan, M.E. Simon, A. Gasparyan, H. Shea, V. Lifton, C. Bolle, S. Arney, R. Frahm, M. Paczkowski, M. Haueis, R. Ryf, D.T. Neilson, J. Kim, R. Giles, D. Bishop: Beam-steering micromirrors for large optical cross-connects, J. Lightwave Technol. **21**, 634–642 (2003)
38.114 S.A. Henck: Lubrication of digital micromirror devices, Tribol. Lett. **3**, 239–247 (1997)
38.115 M.R. Douglass: Lifetime estimates and unique failure mechanisms of the digital micromirror devices (DMD). In: *Proc. 36th Annual Inter. Reliab. Phys. Symp.* (IEEE, New York 1998) pp. 9–16
38.116 M.R. Douglass: DMD reliability: A MEMS success story. In: *Reliability, Testing, and Characterization of MEMS/MOEMS II, Proc. SPIE*, Vol. 4980 (SPIE, Bellingham 2003) pp. 1–11
38.117 H. Liu, B. Bhushan: Nanotribological characterization of digital micromirror devices using an atomic force microscope, Ultramicroscopy **100**, 391–412 (2004)
38.118 H. Liu, B. Bhushan: Investigation of nanotribological and nanomechanical properties of the digital micromirror device by atomic force microscope, J. Vac. Sci. Technol. A **22**, 1388–1396 (2004)
38.119 L. J. Hornbeck: Low Surface Energy Passivation Layer for Micromechanical Devices, US Patent 5602671 (1997)
38.120 R. A. Robbins, S. J. Jacobs: Lubricant Delivery for Micromechanical Devices, US Patent 6300294 B1 (2001)
38.121 I. DeWolf, W.M. van Spengen: Techniques to study the reliability of metal RF MEMS capacitive switches, Microelectron. Reliab. **42**, 1789–1794 (2002)
38.122 U. Zaghloul, G. Papaioannou, B. Bhushan, F. Coccetti, P. Pons, R. Plana: On the reliability of electrostatic NEMS/MEMS devices: Review of present knowledge on the dielectric charging and stiction failure mechanisms and novel characterization methodologies, Microelectron. Reliab. **51**, 1810–1818 (2011)
38.123 U. Zaghloul, G.J. Papaioanou, B. Bhushan, F. Coccetti, P. Pons, R. Plana: New insight into reliability of electrostatic capacitive RF MEMS switches, Int. Technol. Microw. Wirel. **3**, 571–586 (2011)
38.124 B. Bhushan, K.J. Kwak: Platinum-coated probes sliding at up to 100 mm s^{-1} against coated silicon wafers for AFM probe-based recording technology, Nanotechnology **18**, 345504 (2007)
38.125 B. Bhushan, K.J. Kwak: Noble metal-coated probes sliding at up to 100 mm s^{-1} against PZT films for AFM probe-based ferroelectric recording technology, J. Phys.: Condens. Matter **20**, 225013 (2008)
38.126 K.J. Kwak, B. Bhushan: Platinum-coated probes sliding at up to 100 mm/s against lead zirconate titanate films for atomic force microscopy probe-based ferroelectric recording technology, J. Vac. Sci. Technol. A **26**, 783–793 (2008)
38.127 B. Bhushan, K. Kwak, M. Palacio: Nanotribology and nanomechanics of AFM probe-based data recording technology, J. Phys.: Condens. Matter **20**, 365207 (2008)
38.128 B. Bhushan, M. Palacio, B. Kinzig: AFM-based nanotribological and electrical characterization of ultrathin wear-resistant ionic liquid films, J. Colloid Interf. Sci. **317**, 275–287 (2008)
38.129 M. Palacio, B. Bhushan: Ultrathin wear-resistant ionic liquid films for novel MEMS/NEMS applications, Adv. Mater. **20**, 1194–1198 (2008)
38.130 M. Palacio, B. Bhushan: Molecularly thick dicationic liquid films for nanolubrication, J. Vac. Sci. Technol. A **27**, 986–995 (2009)
38.131 J.M. Berlin, J. Yu, W. Lu, E.E. Walsh, L. Zhang, P. Zhang, W. Chen, A.T. Kan, M.S. Wong, M.B. Tomson, J.M. Tour: Engineered nanoparticles for hydrocarbon detection in oil-field rocks, Energy Environ. Sci. **4**, 505–509 (2011)
38.132 P.R. Johnson, N. Sun, M. Elimelech: Colloid transport in geochemically heterogeneous porous media: Modeling and measurements, Environ. Sci. Techol. **30**, 3284–3293 (1996)
38.133 A. Pavía-Sanders, S. Zhang, J.A. Flores, J.E. Sanders, J.E. Raymond, K.L. Wooley: Robust magnetic/polymer hybrid nanoparticles designed for crude oil entrapment and recovery

in aqueous environments, ACS Nano **7**, 7552–7561 (2013)
38.134 Z. Chen, D. Cummins, B.N. Reinecke, E. Clark, M.K. Sunkara, T.F. Jaramillo: Core–shell MoO_3–MoS_2 nanowires for hydrogen evolution: A functional design for electrocatalytic materials, Nano Lett. **11**, 4168–4175 (2011)
38.135 T. Ohnstein, T. Fukiura, J. Ridley, U. Bonne: Micromachined silicon microvalve. In: *Proc. IEEE-MEMS Workshop* (IEEE, New York 1990) pp. 95–98
38.136 S. Shoji, M. Esashi: Microflow devices and systems, J. Micromech. Microeng. **4**, 157–171 (1994)
38.137 M. Stehr, S. Messner, H. Sandmaier, R. Zenergle: The VAMP – A new device for handing liquids or gases, Sens. Actuators A **57**, 153–157 (1996)
38.138 P. Woias: Micropumps – Summarizing the first two decades. In: *Proc. SPIE – Microfluid. bioMEMS*, Vol. 4560, ed. by C.H. Mastrangelo, H. Becker (SPIE, Bellingham 2001) pp. 39–52
38.139 N.T. Nguyen, X. Huang, T.K. Chuan: MEMS-Micropumps: A review, ASME J. Fluids Eng. **124**, 384–392 (2002)
38.140 B. Bustgens, W. Bacher, W. Menz, W.K. Schomburg: Micropump manufactured by thermoplastic molding. In: *Proc. IEEE-MEMS Workshop* (IEEE, Piscataway 1994) pp. 18–21
38.141 C.H. Ahn, M.G. Allen: Fluid micropumps based on rotary magnetic actuators. In: *MEMS '95: IEEE 8th Int. Workshop MEMS* (IEEE, Piscataway 1995) pp. 408–412
38.142 J. Doepper, M. Clemens, W. Ehrfeld, S. Jung, K.P. Kaemper, H. Lehr: Micro gear pumps for dosing of viscous fluids, J. Micromech. Microeng. **7**, 230–232 (1997)
38.143 J. Henniker: Triboelectricity in polymers, Nature **196**, 474 (1962)
38.144 M. Sakaguchi, H. Kashiwabara: A generation mechanism of triboelectricity due to the reaction of mechaniradicals with mechanoions which are produced by mechanical fracture of solid polymer, Colloid Poly. Sci. **270**, 621–626 (1992)
38.145 G.R. Freeman, N.H. March: Triboelectricity and some associated phenomena, Mat. Sci. Eng. **15**, 1454–1458 (1999)
38.146 S.K. Cho, H. Moon, C.-J. Kim: Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits, J. MEMS **12**, 70–80 (2003)
38.147 A.R. Wheeler, H. Moon, C.A. Bird, R.R.O. Loo, C.-J. Kim, J.A. Loo, R.L. Garrell: Digital microfluidics with in-line sample purification for proteomics analysis with MALDI-MS, Anal. Chem. **77**, 534–540 (2005)
38.148 S.C. Lee, M.T. Keener, D.R. Tokachichu, B. Bhushan, P.D. Barnes, B.R. Cipriany, M. Gao, L.J. Brillson: Protein binding on thermally grown silicon dioxide, J. Vac. Sci. Technol. B **23**, 1856–1865 (2005)
38.149 J. Black: *Biological Performance of Materials: Fundamentals of Biocompatibility* (Marcel Dekker, New York 1999)
38.150 B. Bhushan, K. Kwak, S. Gupta, S.C. Lee: Nanoscale adhesion, friction and wear studies of biomolecules on sam-coated silica and alumina based surfaces, J. Royal Soc. Inter. **6**, 719–733 (2009)
38.151 G. Wei, B. Bhushan, N. Ferrell, D. Hansford: Microfabrication and nanomechanical characterization of polymer microelectromechanical systems for biological applications, J. Vac. Sci. Technol. A **23**, 811–819 (2005)
38.152 B. Bhushan, T. Kasai, G. Kulik, L. Barbieri, P. Hoffmann: AFM study of perfluorosilane and alkylsilane self-assembled monolayers for anti-stiction in MEMS/NEMS, Ultramicroscopy **105**, 176–188 (2005)
38.153 B. Bhushan, D. Hansford, K.K. Lee: Surface modification of silicon and PDMS surfaces with vapor phase deposited ultrathin fluorosilane films for biomedical nanodevices, J. Vac. Sci. Technol. A **24**, 1197–1202 (2006)
38.154 T. Kasai, B. Bhushan, G. Kulik, L. Barbieri, P. Hoffmann: Nanotribological study of perfluorosilane SAMs for anti-stiction and low wear, J. Vac. Sci. Technol. B **23**, 995–1003 (2005)
38.155 K.K. Lee, B. Bhushan, D. Hansford: Nanotribological characterization of perfluoropolymer thin films for bioMEMS applications, J. Vac. Sci. Technol. A **23**, 804–810 (2005)
38.156 R. Duncan: The dawning era of polymer therapeutics, Nat. Rev. Drug Discov. **2**, 347–360 (2003)
38.157 T. Murakami, K. Ajima, J. Miyawaki, M. Yudasaka, S. Iijima, K. Shiba: Drug-loaded carbon nanohorns: Adsorption and release of dexamethasone in vitro, Mol. Pharm. **1**, 399–405 (2004)
38.158 M. Ferrari: Cancer nanotechnology: Opportunities and challenges, Nat. Rev. Cancer **5**, 161–171 (2005)
38.159 P. Decuzzi, R. Pasqualini, W. Arap, M. Ferrari: Intravascular delivery of particulate systems: Does geometry really matter?, Pharmaceu. Res. **26**, 235–243 (2009)
38.160 N.R. Panyala, E.M. Pena-Mendez, J. Havel: Gold and nano-gold in medicine: Overview, toxicology and perspectives, J. Appl. Biomed. **7**, 75–91 (2009)
38.161 M. Ferrari: Frontiers in cancer nanomedicine: directing mass transport through biological barriers, Trends Biotechnol. **28**, 181–188 (2010)
38.162 D.J. Irvine: Drug delivery: One nanoparticle, one kill, Nat. Mater. **10**, 342–343 (2011)
38.163 P. Decuzzi, S. Lee, B. Bhushan, M. Ferrari: A theoretical model for the margination of particles with blood vessels, Ann. Biomed. Eng. **33**, 179–190 (2005)
38.164 B. Bhushan: Nanotribology of carbon nanotubes, J. Phys.: Condens. Matter **20**, 365214 (2008)
38.165 H. Guckel, D.W. Burns: Fabrication of micromechanical devices from polysilicon films with smooth surfaces, Sens. Actuators **20**, 117–122 (1989)
38.166 G.T. Mulhern, D.S. Soane, R.T. Howe: Supercritical carbon dioxide drying of microstructures. In: *Proc. Int. Conf. Solid-State Sens. Actuators* (IEEE, New York 1993) pp. 296–299

38.167 K.F. Man, B.H. Stark, R. Ramesham: *A Resource Handbook for MEMS Reliability, Rev. A* (JPL, Pasadena 1998)

38.168 S. Kayali, R. Lawton, B.H. Stark: MEMS reliability assurance activities at JPL, EEE Links **5**, 10–13 (1999)

38.169 S. Arney: Designing for MEMS Reliability, MRS Bull. **26**, 296–299 (2001)

38.170 K. F. Man: MEMS reliability for space applications by elimination of potential failure modes through testing and analysis, http://www.rel.jpl.nasa.gov/0rg/5053/atop/products/Prod-map.html (NASA 2001)

38.171 S. Sundararajan, B. Bhushan: Development of AFM-based techniques to measure mechanical properties of nanoscale structures, Sens. Actuators A **101**, 338–351 (2002)

38.172 B. Bhushan, J.N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374**, 607–616 (1995)

38.173 B. Bhushan: Nanotribology, nanomechanics and nanomaterials characterization, Phil. Trans. R. Soc. A **366**, 1351–1381 (2008)

38.174 J.S. Shor, D. Goldstein, A.D. Kurtz: Characterization of n-type β-SiC as a piezoresistor, IEEE Trans. Electron Devices **40**, 1093–1099 (1993)

38.175 M. Mehregany, C.A. Zorman, N. Rajan, C.H. Wu: Silicon carbide MEMS for harsh environments, Proc. IEEE **86**, 1594–1610 (1998)

38.176 Y.T. Yang, K.L. Ekinci, X.M.H. Huang, L.M. Schiavone, M.L. Roukes, C.A. Zorman, M. Mehregany: Monocrystalline silicon carbide nanoelectromechanical systems, Appl. Phys. Lett. **78**, 162–164 (2001)

38.177 D. Young, J. Du, C.A. Zorman, W.H. Ko: High-temperature single crystal 3C-SiC capacitive pressure sensor, IEEE Sens. J. **4**, 464–470 (2004)

38.178 C.A. Zorman, A.J. Fleischmann, A.S. Dewa, M. Mehregany, C. Jacob, S. Nishino, P. Pirouz: Epitaxial growth of 3C-SiC films on 4 in. diam Si(100) silicon wafers by atmospheric pressure chemical vapor deposition, J. Appl. Phys. **78**, 5136–5138 (1995)

38.179 C.A. Zorman, S. Roy, C.H. Wu, A.J. Fleischman, M. Mehregany: Characterization of polycrystalline silicon carbide films grown by atmospheric pressure chemical vapor deposition on polycrystalline silicon, J. Mater. Res. **13**, 406–412 (1998)

38.180 A.A. Yasseen, C.H. Wu, C.A. Zorman, M. Mehregany: Fabrication and testing of surface micromachined polycrystalline SiC micromotors, IEEE Elect. Dev. Lett. **21**, 164–166 (2000)

38.181 B. Bhushan, B.K. Gupta: *Handbook of Tribology: Materials, Coatings and Surface Treatments* (McGraw Hill, New York 1991)

38.182 J.F. Shackelford, W. Alexander, J.S. Park (Eds.): *CRC Material Science and Engineering Handbook*, 2nd edn. (CRC, Boca Raton 1994)

38.183 B.K. Gupta, J. Chevallier, B. Bhushan: Tribology of ion bombarded silicon for micromechanical applications, ASME J. Tribol. **115**, 392–399 (1993)

38.184 B.K. Gupta, B. Bhushan, J. Chevallier: Modification of tribological properties of silicon by boron ion implantation, Tribol. Trans. **37**, 601–607 (1994)

38.185 B.K. Gupta, B. Bhushan: Nanoindentation studies of ion implanted silicon, Surf. Coat. Technol. **68/69**, 564–570 (1994)

38.186 B. Bhushan, V.N. Koinkar: Tribological studies of silicon for magnetic recording applications, J. Appl. Phys. **75**, 5741–5746 (1994)

38.187 G.M. Pharr: The anomalous behavior of silicon during nanoindentation. In: *Thin Films: Stresses and Mechanical Properties III*, Vol. 239, ed. by W.D. Nix, J.C. Bravman, E. Arzt, L.B. Freund (Materials Research Soc., Pittsburgh 1991) pp. 301–312

38.188 D.L. Callahan, J.C. Morris: The extent of phase transformation in silicon hardness indentation, J. Mater. Res. **7**, 1612–1617 (1992)

38.189 N.A. Fleck, G.M. Muller, M.F. Ashby, J.W. Hutchinson: Strain gradient plasticity: Theory and experiment, Acta Metall. Mater. **42**, 475–487 (1994)

38.190 B. Bhushan, S. Venkatesan: Friction and wear studies of silicon in sliding contact with thin-film magnetic rigid disks, J. Mater. Res. **8**, 1611–1628 (1993)

38.191 S. Venkatesan, B. Bhushan: The role of environment in the friction and wear of single-crystal silicon in sliding contact with thin-film magnetic rigid disks, Adv. Info Storage Syst. **5**, 241–257 (1993)

38.192 S. Venkatesan, B. Bhushan: The sliding friction and wear behavior of single-crystal, polycrystalline and oxidized silicon, Wear **171**, 25–32 (1994)

38.193 B. Bhushan: Chemical, mechanical and tribological characterization of ultra-thin and hard amorphous carbon coatings as thin as 3.5 nm: Recent developments, Diam. Rel. Mater. **8**, 1985–2015 (1999)

38.194 B. Bhushan, S. Sundararajan, X. Li, C.A. Zorman, M. Mehregany: Micro/nanotribological studies of single-crystal silicon and polysilicon and SiC films for use in MEMS devices. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 407–430

38.195 S. Sundararajan, B. Bhushan: Micro/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear **217**, 251–261 (1998)

38.196 X. Li, B. Bhushan: Micro/nanomechanical characterization of ceramic films for microdevices, Thin Solid Films **340**, 210–217 (1999)

38.197 V.N. Koinkar, B. Bhushan: Micro/nanoscale studies of boundary layers of liquid lubricants for magnetic disks, J. Appl. Phys. **79**, 8071–8075 (1996)

38.198 V.N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14**, 2378–2391 (1996)

38.199 H. Liu, B. Bhushan: Nanotribological characterization of molecularly-thick lubricant films for applications to MEMS/NEMS by AFM, Ultramicroscopy **97**, 321–340 (2003)

38.200 Z. Tao, B. Bhushan: Bonding, degradation, and environmental effects on novel perfluoropolyether lubrications, Wear **259**, 1352–1361 (2005)

38.201 B. Bhushan, M. Cichomski, Z. Tao, N.T. Tran, T. Ethen, C. Merton, R.E. Jewett: Nanotribological characterization and lubricant degradation studies of metal-film magnetic tapes using novel lubricants, ASME J. Tribol. **129**, 621f–6627 (2007)

38.202 M. Palacio, B. Bhushan: Surface potential and resistance measurements for detecting wear of chemically-bonded and unbonded molecularly-thick perfluoropolyether lubricant films using atomic force microscopy, J. Colloid Interf. Sci. **315**, 261–269 (2007)

38.203 M. Palacio, B. Bhushan: Wear detection of candidate MEMS/NEMS lubricant films using atomic force microscopy-based surface potential measurements, Scripta Mater. **57**, 821–824 (2007)

38.204 T. Stifter, O. Marti, B. Bhushan: Theoretical investigation of the distance dependence of capillary and van der Waals forces in scanning force microscopy, Phys. Rev. B **62**, 13667–13673 (2000)

38.205 B. Bhushan, A.V. Kulkarni, V.N. Koinkar, M. Boehm, L. Odoni, C. Martelet, M. Belin: Microtribological characterization of self-assembled and Langmuir–Blodgett monolayers by atomic force and friction force microscopy, Langmuir **11**, 3189–3198 (1995)

38.206 B. Bhushan, H. Liu: Nanotribological properties and mechanisms of alkylthiol and biphenyl thiol self-assembled monolayers studied by AFM, Phys. Rev. B **63**, 245412 (2001)

38.207 H. Liu, B. Bhushan, W. Eck, V. Stadler: Investigation of the adhesion, friction, and wear properties of biphenyl thiol self-assembled monolayers by atomic force microscopy, J. Vac. Sci. Technol. A **19**, 1234–1240 (2001)

38.208 H. Liu, B. Bhushan: Investigation of nanotribological properties of self-assembled monolayers with alkyl and biphenyl spacer chains, Ultramicroscopy **91**, 185–202 (2002)

38.209 B. Bhushan, M. Cichomski, E. Hoque, J.A. DeRose, P. Hoffmann, H.J. Mathieu: Nanotribological characterization of perfluoroalkylphosphonate self-assembled monolayers deposited on aluminum-coated silicon substrates, Microsyst. Technol. **12**, 588–596 (2006)

38.210 N.S. Tambe, B. Bhushan: Nanotribological characterization of self assembled monolayers deposited on silicon and aluminum substrates, Nanotechnology **16**, 1549–1558 (2005)

38.211 Z. Tao, B. Bhushan: Degradation mechanisms and environmental effects on perfluoropolyether, self assembled monolayers, and diamondlike carbon films, Langmuir **21**, 2391–2399 (2005)

38.212 Anonymous: *Properties of Silicon*, EMIS Data Rev., Vol. 4 (INSPEC, London 1998)

38.213 D. Maharaj, B. Bhushan: Friction, wear and mechanical behavior of nano-objects on the nanoscale, Mater. Sci. Eng. R **95**, 1–43 (2015)

38.214 O.J. Cayre, V.N. Paunov: Contact angles of colloid silica and gold particles at air–water and oil–water interfaces determined with the gel trapping technique, Langmuir **20**, 9594–9599 (2004)

38.215 M.C. Fuerstenau, K.N. Han (Eds.): *Principles of Mineral Processing* (Society for Mining, Metallurgy, and Exploration (SME), Littleton 2003)

38.216 S. Zhang, X.T. Zeng, Z.G. Tang, M.J. Tan: Exploring the antisticking properties of solid lubricant thin films in transfer molding, Int. J. Mod. Phys. B **16**, 1080–1085 (2002)

38.217 S. Wang, Y. Zhang, N. Abidi, L. Cabrales: Wettability and surface free energy of graphene films, Langmuir **25**, 11078–11081 (2009)

38.218 D. Maharaj, B. Bhushan: Scale effects of nanomechanical properties and deformation behavior of Au nanoparticle and thin film using depth sensing nanoindentation, Beilstein J. Nanotechnol. **5**, 822–836 (2014)

38.219 G.V. Samsonov (Ed.): *Handbook of the Physiochemical Properties of the Elements* (Plenum, New York 1968)

38.220 D. Maharaj, B. Bhushan: Nanomechanical behavior of MoS_2 and WS_2 multi-walled nanotubes and carbon nanohorns, Sci. Rep. – Nature **5**, 8539 (2015)

38.221 J.L. Feldman: Elastic constants of 2H-MoS_2 and 2H-$NbSe_2$ extracted from measured dispersion curves and linear compressibilities, J. Phys. Chem. Solids **37**, 1141–1144 (1976)

38.222 C. Lee, X. Wei, J.W. Kysar, J. Hone: Measurement of the elastic properties and intrinsic strength of monolayer graphene, Science **32**, 385–388 (2008)

38.223 Y. Zhang, C. Pan: Measurements of mechanical properties and number of layers of graphene from nano-indentation, Diam. Relat. Mater. **24**, 1–5 (2012)

38.224 D. Maharaj, B. Bhushan, S. Iijima: Effect of carbon nanohorns on nanofriction and wear reduction in dry and liquid environments, J. Colloid Interface Sci. **400**, 147–160 (2013)

38.225 K. Mougin, E. Gnecco, A. Rao, M.T. Cuberes, S. Jayaraman, E.W. McFarland, H. Haidara, E. Meyer: Manipulation of gold nanoparticles: Influence of surface chemistry, temperature, and environment (vacuum versus ambient atmosphere), Langmuir **24**, 1577–1581 (2008)

38.226 M. Palacio, B. Bhushan: A nanoscale friction investigation during manipulation of nanoparticles in controlled environments, Nanotechnology **19**, 315710 (2008)

38.227 D. Maharaj, B. Bhushan: Effect of spherical Au nanoparticles on nanofriction and wear reduction in dry and liquid environments, Beilstein J. Nanotechnol. **3**, 759–772 (2012)

38.228 D. Maharaj, B. Bhushan: Nanomanipulation, nanotribology and nanomechanics of Au nanorods in dry and liquid environments using an AFM and depth sensing nanoindenter, Nanoscale **6**, 5838–5852 (2014)

38.229 D. Maharaj, B. Bhushan: Effect of MoS_2 and WS_2 nanotubes on nanofriction and wear reduction in

dry and liquid environments, Tribol. Lett. **49**, 323–339 (2013)
38.230 K.-S. Kim, H.-J. Lee, C. Lee, S.-K. Lee, H. Jang, J.-H. Ahn, J.-H. Kim, H.-J. Lee: Chemical vapor deposition-grown graphene: The thinnest solid lubricant, ACS Nano **5**, 5107–5114 (2011)
38.231 E. Sani, S. Barison, C. Pagura, L. Mercatelli, P. Sansoni, D. Fontani, D. Jafrancesco, F. Francini: Carbon nanohorns-based nanofluids as direct sunlight absorbers, Opt. Express **18**, 5179–5187 (2010)
38.232 A.K. Geim, K.S. Novoselov: The rise of graphene, Nature Mater. **6**, 183–191 (2007)
38.233 S. Iijima, M. Yudasaka, R. Yamada, S. Bandow, K. Suenaga, F. Kokai, K. Takahashi: Nano-aggregates of single-walled graphitic carbon nanohorns, Chem. Phys. Lett. **309**, 165–170 (1999)
38.234 B. Bhushan, S. Sundararajan: Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Act. Mater. **46**, 3793–3804 (1998)
38.235 A.C.R. Grayson, R.S. Shawgo, A.M. Johnson, N.T. Flynn, Y. Li, M.J. Cima, R. Langer: A bioMEMS review: MEMS technology for physiologically integrated devices, Proc. IEEE **92**, 6–21 (2004)
38.236 D.L. Wise: *Biomaterials and Bioengineering Handbook* (CRC, New York 2000)
38.237 B.D. Ratner, A.S. Hoffman, F.J. Schoen, J.E. Lemons: *Biomaterials Science: An Introduction to Materials in Medicine* (Elsevier, San Diego 2004)
38.238 P. Turbill, T. Beugeling, A.A. Poot: Proteins involved in the Vroman effect during exposure of human blood plasma to glass and polyethylene, Biomaterials **17**, 1279–1287 (1996)
38.239 S.R. Schricker, M.L.B. Palacio, B. Bhushan: Antibody sensed protein surface conformation, Mater. Today **14**, 616–621 (2011)
38.240 S. Schricker, M.L.B. Palacio, B. Bhushan: Designing nanostructured block copolymer surfaces to control protein adhesion, Phil. Trans. Roy. Soc. A **370**, 2348–2380 (2012)
38.241 B. Bhushan, S.R. Schricker: A review of block copolymer-based biomaterials that control protein and cell interactions, J. Biomed. Mater. Res. A **102A**, 2467–2480 (2014)
38.242 B. Bhushan, J. Utter: Nanoscale adhesion, friction and wear of proteins on polystyrene, Colloids Surf. B: Biointerfaces **102**, 484–491 (2013)
38.243 M.L.B. Palacio, B. Bhushan: Bioadhesion: A review of concepts and applications, Phil. Trans. Roy. Soc. A **370**, 2321–2347 (2012)
38.244 G.L. Francis: Albumin and mammalian cell culture: Implications for biotechnology applications, Cytotechnology **62**(1), 1–16 (2010)
38.245 G.A. Di Lullo, S.M. Sweeney, J. Korkko, L. Ala-Kokko, J.D. San Antonio: Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen, J. Biol. Chem. **277**(6), 4223–4231 (2002)
38.246 R. Pankov, K.M. Yamada: Fibronectin at a glance, J. Cell Sci. **115**, 3861–3863 (2002)
38.247 N.M. Green: Avidin and streptavidin, Method Enzymol. **184**, 51–67 (1990)
38.248 K. Hirayama, S. Akashi, M. Furuya, K. Fukuhara: Rapid confirmation and revision of the primary structure of bovine serum albumin by ESIMS and Frit-FAB LC/MS, Biochem. Biophys. Res. Commun. **173**, 639–646 (1990)
38.249 T. Peters: *Albumin: An Overview and Bibliography* (Miles Laboratory Research Products Division, Elkhart 1980)
38.250 S. Hattori, E. Adachi, T. Ebihara, T. Shirai, I. Someki, S. Irie: Alkali-treated collagen retained the triple helical conformation and the ligand activity for the cell adhesion via alpha2beta1 integrin, J. Biochem. **125**, 676–684 (1999)
38.251 R.A. Proctor: Fibronectin: A brief overview of its structure, function, and physiology, Rev. Infect. Dis. **9**, S317–S321 (1987)
38.252 M.M. Browne, G.V. Lubarsky, M.R. Davidson, R.H. Bradley: Protein adsorption onto polystyrene surfaces studied by XPS and AFM, Surf. Sci. **553**, 155–167 (2004)
38.253 E. Pamuła, V. De Cupere, Y.F. Dufrêne, P.G. Rouxhet: Nanoscale organization of adsorbed collagen: Influence of substrate hydrophobicity and adsorption time, J. Colloid Interface Sci. **271**, 80–91 (2004)
38.254 J.D. Andrade, V. Hlady, A.P. Wei: Adsorption of complex proteins at interfaces, Pure Appl. Chem. **64**(11), 1777–1781 (1992)
38.255 M. Kosmulski: *Chemical Properties of Material Surfaces* (Marcel Dekker, New York 2001)
38.256 J. Fan, C.W. Myant, R. Underwood, P.M. Cann, A. Hart: Inlet protein aggregation: A new mechanism for lubricating film formation with model synovial fluids, Proc. Inst. Mech. Eng. H **225**, 696–709 (2011)
38.257 F.H. Silver, J.W. Freeman, G.P. Seehra: Collagen self-assembly and the development of tendon mechanical properties, J. Biomech. **36**, 1529–1553 (2003)
38.258 S. Xu, J. Yamanaka, I. Miyata, M. Yonese: Effects of denaturation and association of collagen on adsorption behavior: Two-dimensional nanostructure and its property, Colloid. Polym. Sci. **8**, 899–906 (2007)
38.259 E. Eteshola, M.T. Keener, M. Elias, J. Shapiro, L.J. Brillson, B. Bhushan, S.C. Lee: Engineering functional protein interfaces for immunologically modified field effect transistor (immunoFET) by molecular genetics means, J. R. Soc. Interf. **5**, 123–127 (2008)
38.260 K. Kallury, P.M. MacDonald, M. Thompson: Effect of surface water and base catalysis on the silanization of silica by (aminopropyl)alkoxysilanes studied by x-ray photoelectron spectroscopy and 13C cross-polarization/magic angle spinning nuclear magnetic resonance, Langmuir **10**, 492–499 (1994)
38.261 J.H. Moon, J.W. Shin, S.Y. Kim, J.W. Park: Formation of uniform aminosilane thin layers: An imine formation to measure relative surface density of

the amine group, Langmuir **12**, 4621–4624 (1996)
38.262 Y. Han, D. Mayer, A. Offenhausser, S. Ingebrandt: Surface activation of thin silicon-oxides by wet cleaning and silanization, Thin Solid Films **510**, 175–180 (2006)
38.263 N.S. Tambe, B. Bhushan: Identifying materials with low friction and adhesion for nanotechnology applications, Appl. Phys. Lett. **86**, 061906-1–061906-3 (2005)
38.264 N.S. Tambe, B. Bhushan: Micro/nanotribological characterization of PDMS and PMMA used for bioMEMS/NEMS applications, Ultramicroscopy **105**, 238–247 (2005)
38.265 B. Bhushan, Z. Burton: Adhesion and friction properties of polymers in microfluidic devices, Nanotechnology **16**, 467–478 (2005)
38.266 D.R. Tokachichu, B. Bhushan: Bioadhesion of polymers for bioMEMS, IEEE Trans. Nanotech. **5**, 228–231 (2005)
38.267 S. Sundararajan, B. Bhushan: Static friction and surface roughness studies of surface micromachined electrostatic micromotors using an atomic force/friction force microscope, J. Vac. Sci. Technol. A **19**, 1777–1785 (2001)
38.268 C.H. Mastrangelo, C.H. Hsu: Mechanical stability and adhesion of microstructures under capillary forces – Part II: Experiments, J. Microelectromech. Syst. **2**, 44–55 (1993)
38.269 R. Maboudian, R.T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B **15**, 1–20 (1997)
38.270 C.H. Mastrangelo: Surface force induced failures in microelectromechanical systems. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 367–395
38.271 M.P. De Boer, T.A. Michalske: Accurate method for determining adhesion of cantilever beams, J. Appl. Phys. **86**, 817 (1999)
38.272 R.L. Alley, G.J. Cuan, R.T. Howe, K. Komvopoulos: The effect of release-etch processing on surface microstructure stiction. In: *Proc. Solid State Sens. Actuator Workshop*, ed. by C.H. Mastrangelo, C.H. Hsu (IEEE, New York 1992) pp. 202–207
38.273 H. Liu, B. Bhushan: Adhesion and friction studies of microelectromechanical systems/nanoelectromechanical systems materials using a novel microtriboapparatus, J. Vac. Sci. Technol. A **21**, 1528–1538 (2003)
38.274 B. Bhushan, H. Liu, S.M. Hsu: Adhesion and friction studies of silicon and hydrophobic and low friction films and investigation of scale effects, ASME J. Tribol. **126**, 583–590 (2004)
38.275 B. Bhushan: *Biomimetics: Bioinspired Hierarchical-Structured Surfaces for Green Science and Technology*, 2nd edn. (Springer International, Cham 2016)
38.276 S.K. Chilamakuri, B. Bhushan: A comprehensive kinetic meniscus model for prediction of long-term static friction, J. Appl. Phys. **15**, 4649–4656 (1999)
38.277 Y.C. Tai, R.S. Muller: Frictional study of IC processed micromotors, Sens. Actuators A **21–23**, 180–183 (1990)
38.278 K.J. Gabriel, F. Behi, R. Mahadevan, M. Mehregany: In situ friction and wear measurement in integrated polysilicon mechanisms, Sens. Actuators A **21–23**, 184–188 (1990)
38.279 M.G. Lim, J.C. Chang, D.P. Schultz, R.T. Howe, R.M. White: Polysilicon microstructures to characterize static friction. In: *Proc. IEEE Micro Electro Mech. Syst.* (IEEE, New York 1990) pp. 82–88
38.280 U. Beerschwinger, S.J. Yang, R.L. Reuben, M.R. Taghizadeh, U. Wallrabe: Friction measurements on LIGA-processed microstructures, J. Micromech. Microeng. **4**, 14–24 (1994)
38.281 D. Matheison, U. Beerschwinger, S.J. Young, R.L. Rueben, M. Taghizadeh, S. Eckert, U. Wallrabe: Effect of progressive wear on the friction characteristics of nickel LIGA processed rotors, Wear **192**, 199–207 (1996)
38.282 G. Wei, B. Bhushan, S.J. Jacobs: Nanomechanical characterization of digital multilayered thin film structures for digital micromirror devices, Ultramicroscopy **100**, 375–389 (2004)
38.283 G. Wei, B. Bhushan, S.J. Jacobs: Nanoscale indentation fatigue and fracture toughness measurements of multilayered thin film structures for digital micromirror devices, J. Vac. Sci. Technol. A **22**, 1397–1405 (2004)
38.284 H. Liu, B. Bhushan: Bending and fatigue study on a nanoscale hinge by an atomic force microscope, Nanotechnology **15**, 1246–1251 (2004)
38.285 B. Bhushan, H. Liu: Characterization of nanomechanical and nanotribological properties of digital micromirror devices, Nanotechnology **15**, 1785–1791 (2004)
38.286 R.C. Jaeger: *Introduction to Microelectronic Fabrication*, Vol. 5 (Addison-Wesley, Reading 1988)
38.287 J. Voldman, M.L. Gray, M.A. Schmidt: Microfabrication in biology and medicine, Annu. Rev. Biomed. Eng. **1**, 401–425 (1999)
38.288 J.W. Judy: Microelectromechanical systems (MEMS): Fabrication, design, and applications, Smart Mater. Struc. **10**, 1115–1134 (2001)
38.289 C. Liu: *Foundations of MEMS* (Pierson Prentice Hall, Upper Saddle River 2006)
38.290 G. Brewer: *Electron-Beam Technology in Microelectronic Fabrication* (Academic, New York 1980)
38.291 K. Valiev: *The Physics of Submicron Lithography* (Plenum, New York 1992)
38.292 E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Munchmeyer: Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA Process), Microelectron. Eng. **4**, 35–56 (1986)
38.293 C.R. Friedrich, R.O. Warrington: Surface characterization of non-lithographic micromachining. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 73–84

38.294 M. Tanaka: Development of desktop machining microfactory, Riken Rev. **34**, 46–49 (2001)
38.295 Y. Xia, G.M. Whitesides: Soft lithography, Angew. Chem. Int. Ed. **37**, 550–575 (1998)
38.296 H. Becker, C. Gaertner: Polymer microfabrication methods for microfluidic analytical applications, Electrophoresis **21**, 12–26 (2000)
38.297 Y. Xia, E. Kim, X.M. Zhao, J.A. Rogers, M. Prentiss, G.M. Whitesides: Complex optical surfaces formed by replica molding against elastomeric masters, Science **273**, 347–349 (1996)
38.298 S.Y. Chou, P.R. Krauss, P.J. Renstrom: Imprint lithography with 25-nanometer resolution, Science **272**, 85–87 (1996)
38.299 A. Kumar, G.M. Whitesides: Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol ink followed by chemical etching, Appl. Phy. Lett. **63**, 2002–2004 (1993)
38.300 J.C. McDonald, D.C. Duffy, J.R. Anderson, D.T. Chiu, H. Wu, O.J.A. Schueller, G.M. Whitesides: Fabrication of microfluidic systems in poly(dimethylsiloxane), Electrophoresis **21**, 27–40 (2000)

39. Friction and Wear in Micro- and Nanomachines

Maarten P. de Boer, Sameer S. Shroff, Frank W. DelRio, W. Robert Ashurst

The prediction and characterization of multi-length-scale tribological phenomena is challenging, yet essential for the advancement of micro- and nanomachine technology. Here, we consider theoretical underpinnings of multiasperity friction, review various approaches to measure micro- and nanoscale friction and discuss the effect of monolayer coatings to reduce friction and wear. We then overview a theoretical framework known as rate-and-state friction (RSF), in which friction is considered to be a continuous function of velocity and interface state. A microscale test platform that is used to measure friction over multiple decades of velocity and normal load is presented and results are reported. Using the RSF framework, we quantitatively predict and validate the transition from stick-slip to steady sliding, enabling the creation of a microscale kinetic phase diagram. Next, we take a brief look at continued progress in spinning micromachine motor technology. Finally, we discuss wear- and tribopolymer-related phenomena in micro- and nanoswitches, which are promising devices to complement transistors due to their low on resistance and steep subthreshold swing. We anticipate great progress towards reliable, contacting micro- and nanomachines by linking theory and experiment to nano- and microscale tribological phenomena and by improving the testing, materials and processing methods used to characterize these phenomena.

39.1 **From Single- to Multiple-Asperity Friction** 1419
39.1.1 Micromachined Test Structures 1420
39.1.2 Monolayer Lubrication in Nano- and MEMS Tribology 1422
39.2 **Rate-State Friction** 1422
39.2.1 MEMS Rate-State Friction Test Platform with Decades of Range-in-Control Parameters 1423
39.2.2 Rate-State Friction Measurements 1425
39.3 **Putting MEMS Friction to Use: Making Motors and Friction Instruments** 1429
39.3.1 Microengines, Pumps and Generators 1429
39.3.2 In-situ TEM for Friction Investigations . 1429
39.4 **Wear and Tribopolymer Evolution in Micro- and Nanoswitches** 1430
39.5 **Concluding Remarks** 1432
References 1432

Friction and wear phenomena present both challenges and opportunities for micro- and nanosystems. In many possible applications, counterfaces rub against each other. Examples of how this can occur are illustrated in Fig. 39.1, where a close-up of features in a complex locking mechanism [39.1] are shown. Here, a large gear ratio is employed to amplify torque. The gear teeth mesh and rotate about hubs that are made by a five structural level polycrystalline silicon (polysilicon) surface micromachining technology at Sandia National Laboratories, known as the SUMMiT V^{TM} process [39.2]. Features known as dimples are attached under the gears to minimize contact and hence adhesion with the substrate. The dimples also prevent the gears from tilting excessively, as does a guide that is shown at the top right of Fig. 39.1. The output gear meshes with a linear rack that is guided via long rails; rubbing contact is also made there. To move from right to left, a pin must correctly traverse a maze inside the linear rack. If an incorrect decision is made, a ratchet and pawl mechanism (not shown), which also requires rubbing contact, prevents resetting.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_39

In the example from Fig. 39.1, friction is generally a parasitic effect and efforts are directed towards minimizing it. Eliminating friction in micromachines has proven difficult, yet therein lies opportunity. A number of researchers have chosen to take advantage of friction, achieving impressive micro- and nanoscale actuation performance. Microscale stepper motors are linear actuators with step sizes on the order of 10–100 nm. Their actuation forces can reach the mN scale and travel ranges up to hundreds of μm have been demonstrated. These characteristics compare favorably with the more commonly used comb drives, which typically move 10–20 μm and deliver forces in the low μN range. Stepper motors are attractive for applications such as positioning for optics [39.3], data storage [39.4, 5] and medicine [39.6]. We also envision that they may be very useful for testing properties of micro- and nanoscale specimens [39.9], and as nanoscale manipulation instruments.

Representative implementations of stepper motors, in which actuation is based on frictional effects, are illustrated in Fig. 39.2. Each of these devices has unique design features, performance tradeoffs and processing requirements. An asymmetric bushing geometry enables the motion of the scratch drive actuator (SDA) in Fig. 39.2a [39.7]. The layout and the drive signal for the SDA are simple, the device area is small ($\approx 30\,\mu m \times 100\,\mu m$), the applied voltage is moderate (30–100 V depending on layer thicknesses), the unidirectional travel distance is 500 μm or more [39.10] and the force output is reasonably large (tens of μN) [39.7]. A model that predicts operating voltages has been developed, but it does not treat frictional forces [39.11]. The shuffle motor [39.12] and nanotractor [39.13] designs place voltage-controlled clamps at the ends of an actuator plate, enabling bidirectional motion. These actuators can be modeled well and the nanotractor, pictured in Fig. 39.2b, achieves a theoretical maximum force value of $F_{max} = 0.5\,mN$ [39.13]. A subsequent design known as the μWalker with two degrees of freedom has been demonstrated [39.8, 14] as shown in Fig. 39.2c. Both the nanotractor and the (2-D) μWalker achieve other theoretical expectations including step size on the order of tens of nm and average velocity proportional to step frequency up to the mm/s range. The 2-D μWalker is produced in a four mask process with two layers of structural polysilicon. It takes advantage of a Si_3N_4 coating for in-plane isolation of voltage applied to the clamps, and also as a mechanical layer to enhance wear characteristics. The nanotractor is fabricated in the SUMMiT V™ process and is also a model friction test structure.

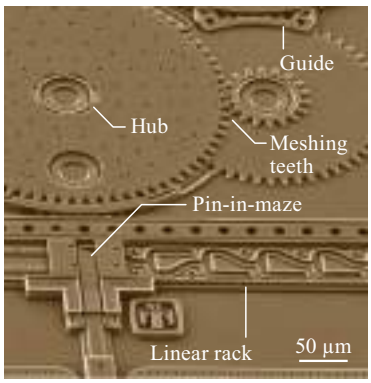


Fig. 39.1 Rubbing surfaces occur in many micromachine applications and pose a reliability challenge. This figure shows an example of complex mechanical logic

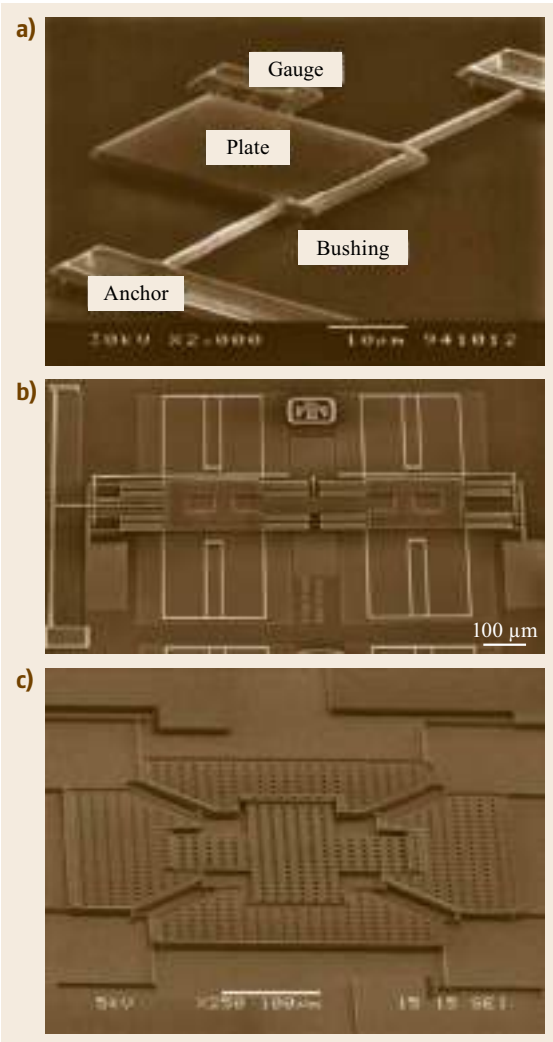


Fig. 39.2a–c Rubbing surfaces can also be used to create powerful actuators such as a **(a)** scratch drive actuator (after [39.7], © IEEE by permission) **(b)** nanotractor and **(c)** 2-D μWalker (after [39.8], © Koninklijke Bibliotheek, Den Haag)

In the following sections, we shall describe the theoretical underpinnings of multiasperity friction, micromachined test structures to measure friction and monolayer coatings that have been used to reduce friction on microelectromechanical systems (MEMS). The major focus of this chapter will then be on testing, metrology, friction and wear.

39.1 From Single- to Multiple-Asperity Friction

According to Amontons' empirically deduced law [39.15], the friction force, F_f, is directly proportional to the applied normal load, F_n, and independent of the apparent contact area, as described by

$$F_f = \mu F_n \,, \tag{39.1}$$

where μ is the coefficient of friction. For surfaces at rest, the coefficient of static friction is denoted as μ_s, while for surfaces in relative motion, the coefficient of dynamic friction is denoted as μ_d.

Upon considering two surfaces in mechanical contact, it becomes clear that real surfaces are rough on the microscale, consisting of many peaks and valleys. Thus, the *real* contact area A_r between two surfaces is actually defined by the highest peaks, also known as asperities, and is much less than the *apparent* contact area A_a. *Bowden* and *Tabor* [39.16] presented the idea that friction is related to the real (as opposed to apparent) area of contact via two basic mechanisms: shearing and plowing. Shearing refers to the force required to break the junctions at contacting asperities, while plowing represents the force needed for a hard asperity to displace softer material. In the absence of plowing (i. e., for two materials with roughly the same hardness as is usually the case in micromachined materials), friction is directly proportional to the real contact area via

$$F_f = \tau A_r \,, \tag{39.2}$$

where τ is the shear strength of the contact. Hence Amontons' law is recovered if A_r is proportional to normal load.

The proportionality between real contact area and applied load, however, seems to disappear for two asperities in *elastic* contact. *Hertz* [39.17] provided the first analysis of the stress distribution and displacement at the interface of two elastic spheres, which can be simplified to a sphere in contact with a flat surface. The total load F_n compressing a sphere of radius R into a flat surface can be related to the contact area A_r by

$$A_r = \pi \left(\frac{3F_n R}{4E^*} \right)^{\frac{2}{3}} , \tag{39.3}$$

where the combined elastic modulus of the two contacting surfaces

$$E^* = \left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right)^{-1} .$$

Johnson, *Kendall*, and *Roberts* (JKR) later observed that the contact area between two rubber spheres was larger than that predicted by the Hertz theory and developed a model to account for the surface forces at the interface [39.18]. The model was based on a balance between the stored elastic energy, the mechanical energy in the applied load, and the surface energy. *Derjaguin*, *Muller*, and *Toporov* (DMT) used a thermodynamic approach to consider the molecular forces in and around the contact zone [39.19]. Unlike the JKR approach, the profile of the sphere outside of the contact area was assumed to be Hertzian (i. e., the surface forces are small enough that their effect on the deformation of the sphere can be neglected). Initially, the models seemed to contradict each other. *Tabor* [39.20] determined that the models represent the extreme ends of a spectrum, and developed a parameter μ_T to span the spectrum. Accordingly, the JKR theory is suitable for elastically compliant materials with a large radius and surface energy w, while the DMT theory is appropriate for elastically stiff materials with a small radius and surface energy. *Maugis* [39.21] later developed a semi-analytical solution using a Dugdale approximation; the transition from DMT to JKR was described in terms of the parameter λ, which is related to μ_T via $\lambda = 1.16\mu_T$. Figure 39.3 illustrates the relationship between A_r and F_n for all of these models. In the presence of attractive forces, the contact area increases at a given load, which increases the friction force, as depicted by (39.2). More importantly, all of the models exhibit a nonlinear relationship between real contact area and applied load, contradicting expectations from Amontons' law.

A number of researchers [39.22–24] have successfully used the aforementioned single-asperity relationships to model friction measurements taken via friction-force microscopy (FFM). However, micro- and nanodevices fabricated using surface micromachining techniques often have contacting surfaces with nanometer-scale surface roughness. As a result, it be-

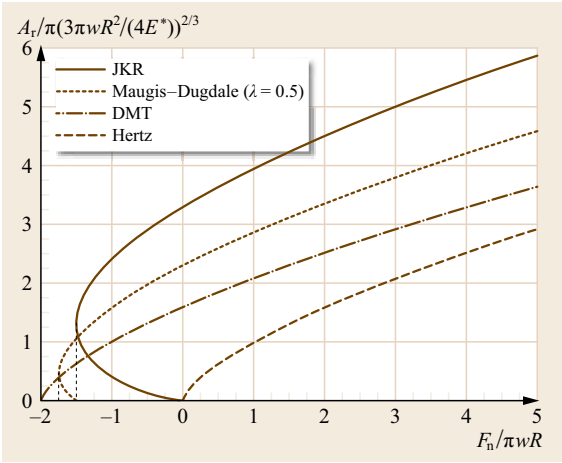


Fig. 39.3 Normalized A_r–F_n curves for the Hertz, DMT, Maugis–Dugdale, and JKR contact models. As the transition parameters μ_T and λ increase from zero to infinity, there is a continuous transition from the DMT to the JKR regime. In the absence of adhesion, all of the models approach the Hertzian case

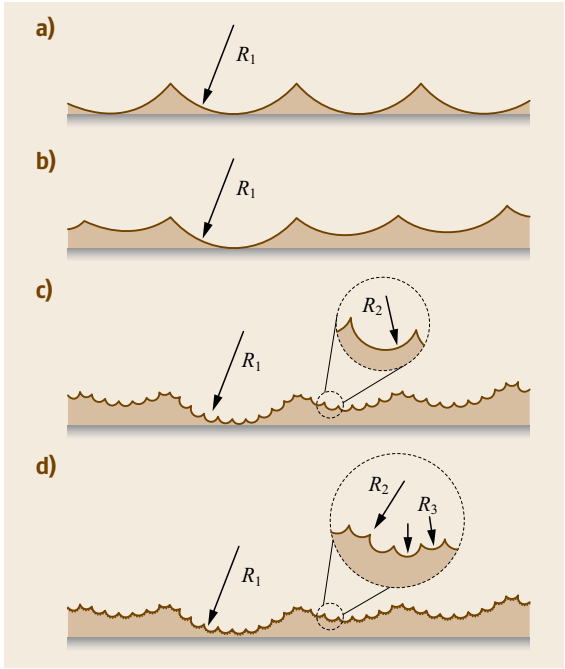


Fig. 39.4a–d Multiple-asperity models (after [39.26]). Assuming elastic deformation, the relationship between the contact area A_r and the applied load F_n are **(a)** $A_r \propto F_n^{2/3}$, **(b)** $A_r \propto F_n^{4/5}$, **(c)** $A_r \propto F_n^{14/15}$, and **(d)** $A_r \propto F_n^{44/45}$. In general, the index n in the relationship $A_r \propto F_n^n$ ranges from $2/3$ for a series of asperities at the same height to ≈ 1 for a multiple-asperity surface with several sets of protuberances

comes important to develop *multiple*-asperity contact models to elucidate the impact of surface topography on the relationship between contact area and applied load. In the simplest case, the rough surface consists of a series of identical asperities all at the same height, as shown in Fig. 39.4a. Here, the applied load is divided evenly among all of the contacting asperities, which results in $A_r \propto F_n^{2/3}$. *Archard* [39.25, 26] improved on this idea by considering a *uniform* distribution of spherical asperities (radii of curvature R_1) in contact with a rigid flat surface as shown in Fig. 39.4b. For elastic deformation, $A_r \propto F_n^{4/5}$. To examine multiscale roughness, a second set of protuberances (radii of curvature R_2) can be evenly distributed over the surface of the existing spheres, such that $R_2 \ll R_1$, as shown in Fig. 39.4c. In this case, $A_r \propto F_n^{14/15}$. With even smaller protuberances of radius R_3, as shown in Fig. 39.4d, the relationship becomes $A_r \propto F_n^{44/45}$. This indicates that proportionality between area and load is not necessarily the result of plastic flow in the contact zones, as originally proposed, but can be due to elastic contact between rough surfaces.

Greenwood and *Williamson* (GW) considered a *statistical* distribution of N asperities with constant radius of curvature R in contact with a flat surface [39.27]. For a Gaussian distribution of asperity heights, A_r is almost exactly proportional to F_n. *Fuller* and *Tabor* [39.28] extended the GW model for JKR contacts to study rough surface adhesion and defined the adhesion parameter θ, which represents the competition between the compressive forces exerted by the higher asperities and the adhesive forces acting between the lower asperities (analogous to the Tabor parameter). In a similar manner, *Maugis* [39.29] extended the GW model for DMT contacts. For infinite θ (work of adhesion $w = 0$), the Maugis theory is equivalent to the GW model and $A_r \propto F_n$. On the other hand, as θ decreases (w increases), the relationship between the contact area and applied load becomes more nonlinear. More advanced theories that extend these ideas are under development [39.30–32].

39.1.1 Micromachined Test Structures

From this survey of multiple-asperity contact models, we see that adhesive forces and asperity distribution can significantly impact the relationships between friction force, contact area and applied load. Therefore, although widely observed, Amontons' law is not necessarily expected. While there are techniques in place to convert topographic data from real surfaces into Gaussian asperity distribution data [39.33], it is unclear as to whether these statistical models accurately represent the surface topography [39.32]. Therefore, it becomes nec-

essary to fabricate micromachined test structures that can be used to measure friction directly. Moreover, micromachined structures are appropriate test vehicles for studying friction and wear because their performance will reflect any effects that are due to details of the fabrication process. These include technologically relevant issues such as the effect of sidewall slope on vertical surfaces, surface roughness and the ability to coat micromachined structures with friction-reducing coatings in spite of sometimes tortuous access paths.

One of the earliest systematic studies of microscale friction was carried out by *Howe* et al. at the University of California, Berkeley [39.34]. Since then, a number of researchers have investigated this critical issue using a variety of techniques and devices. A common approach is to design a moveable beam that can be brought into contact and slid against a counter surface [39.35–41]. An example from [39.39] is shown in Fig. 39.5a,b. The measurements from this study showed that statistical contact models as described above may not apply to lightly loaded, small apparent contact areas in MEMS. This is because such models would predict less than one contact within the contact area! The researchers also found that adhesion between the surfaces must be recognized as an important contribution to the applied normal load. Most microdevice friction work has involved so-called *sidewall* devices such as this one, where the contacting surfaces are the result of an etching procedure and may have vertical striations and significant directional anisotropies. The topography of sidewalls is influenced by lithography, etching *and* grain boundaries, as seen in Fig. 39.5c [39.42], and often exhibits significantly higher roughness than in-plane surfaces where topography is primarily determined by grain growth phenomena.

Timpe and *Komvopoulos* [39.40] developed a sidewall microscale friction device that involved a push drive and a shear drive. Testing of the device consisted of first loading the contact and then shearing the newly formed interface while monitoring the position of the shear drive with a charge-coupled device (CCD) camera, with a spatial resolution of about 0.3 μm. It was shown that the engineering coefficient of friction exhibited a nonlinear dependence on contact pressure that was unique to the testing environment; higher friction was observed at higher humidity levels. Further, the adhesion forces, significant at the microscale, added substantially to the friction when the external load was comparable to the adhesion force [39.40].

Frenken and co-workers recently developed a similar sidewall device, but included an on-device, real-time, electrical capacitance readout mechanism for high-resolution displacement measurement (about 4 nm peak-to-peak noise sensitivity). Their device, called the

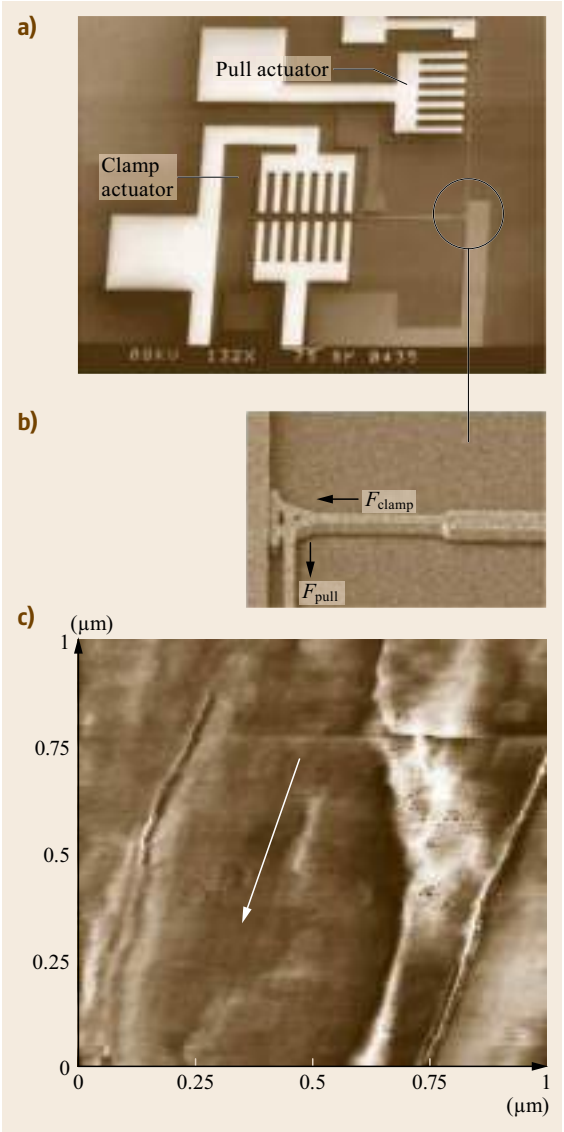


Fig. 39.5 **(a)** A MEMS test structure **(b)** designed to measure sidewall friction (after [39.39], © VSP). **(c)** Topographical image of a sidewall surface, where the RMS roughness is 40 nm (after [39.42], © Xavier by permission)

Leiden MEMS Tribometer, consists of two orthogonally oriented comb drives that are used to position a test slider against a fixed surface. The comb drives are used to generate loading and shearing forces on the contact surface. They indicated two regimes in the friction behavior depending on the load. For the low-loading regime, a wearless stick-slip phenomenon is observed that is repeatable over many cycles but stochastic in position and presumably related to the details of the topography. In the high-loading regime, wear was ob-

served, and although the stick-slip behavior exhibited in the low-load regime disappeared, the motion was not smooth, as ongoing wear processes altered the surface topography and changed the contact mechanics [39.38].

39.1.2 Monolayer Lubrication in Nano- and MEMS Tribology

While the results in the previous section are of interest, the effects of monolayer coatings, which significantly reduce friction, were not studied. It has been well established that monolayer lubricants are effective at reducing friction at the nanoscale. Much of this knowledge has been achieved through nanotribology experiments involving atomic-force microscopy (AFM) [39.43, 44] or a surface-forces apparatus (SFA) [39.45], where ideal systems (i. e., pristine environments, single-crystal substrates, perfect monolayers, etc.) are employed to probe fundamental aspects of friction phenomena using single-asperity contacts.

Nanotribology studies have shown significant (and sometimes conflicting) impact of monolayer thickness (i. e., precursor chain length) and terminal group on friction [39.46–49]. Similar to rough surfaces, there are two basic phenomena involved in single-asperity friction: plowing and shearing [39.44]. With regard to monolayer coatings, plowing refers to the action of the asperity tip deforming the monolayer thereby dissipating energy by introducing chain defects that may elastically relax after the tip has passed, while shearing refers to the localized disruption of intermolecular forces as a result of the shear interaction of the film surface with the tip. Some researchers report an increase in friction as chain length increases [39.50], while others report a decrease in friction [39.48]. These conflicting findings suggest that additional experimental factors contribute to the frictional behavior of monolayer films, and the operating conditions must be considered in the friction study.

As discussed above, the friction force may be proportional to the true contact area, leading to a nonlinear dependence of friction with load. This has been supported by experimental data for inorganic materials. However, in many studies of organic monolayers, a linear relationship is found [39.46, 48, 51–53]. This could be attributed to plastic deformation, to multiple contacting nanoasperities on the single-AFM tip or to viscoelastic deformation during the sliding. It is important to note that many of the fundamental mechanistic studies of nanoscale friction have involved high-modulus inorganic materials with similar bulk and surface properties. Since monolayers are comparatively compliant materials, usually deposited on a stiff material, the validity of applying findings from such fundamental studies to monolayer systems can be questioned. Another aspect to consider is that deviation from linearity as in the JKR model [39.18] may be small and experimentally obscured for low loadings, typical of AFM studies [39.44]. Recently, a continuum thin-coating theory was developed to address these issues [39.54, 55].

Some researchers have successfully applied monolayer films to micromachines for the purpose of studying their effects on friction [39.13, 56–59]. The most convenient monolayers to apply to polysilicon microdevices involve silane-based linking reactions with the oxide layer present on the surface. From the silane class of materials, one of the most commonly studied monolayer film is that produced by octadecyltrichlorosilane (OTS). This particular film has received much attention in the literature due to its self-assembly characteristics [39.60]. Work using AFM has shown that the friction behavior of the film is dependent on the local 2-D phase (liquid expanded or liquid condensed) at the asperity contact [39.61] and that liquid-condensed phases are likely to be associated with grain boundary areas for polysilicon devices [39.62]. It is generally accepted by the MEMS community that OTS and other low-friction monolayer boundary lubricant films reduce the friction on microdevices when properly integrated and applied to the device. For example, a factor of ≈ 4 reduction in friction was reported at 100 nN normal load for self-mated OTS versus self-mated silicon surfaces [39.63].

The findings assembled from uncoated and coated sidewall devices have significant value to the MEMS community and provide guidance on how to more reliably design and operate microdevices. However, statistical contact mechanics models, which rely on numerous contacts, do not apply because of the small contact areas and nonvertical nature of sidewall etching. Rather, an interface with uniform loading over a large area of in-plane (horizontal) surfaces during friction testing is needed.

39.2 Rate–State Friction

Clearly, it would be interesting to link single-asperity measurements made by nanotribologists to multiasperity measurements made by MEMS tribologists. This would involve a detailed understanding of the surface properties, loading characteristics and topography. Indeed, asperity radii of polycrystalline silicon are on

the order of 20–200 nm, in the same range as AFM tips. At the macroscale, friction measurements on kg-scale masses have revealed complex behavior including aging, velocity dependence, and stick-slip to steady-sliding bifurcations. From these, a phenomenological framework known as rate-state friction (RSF) has been advanced [39.64–68]. This development derives from detailed measurements with various combinations of springs, puller velocities and normal loads, and applies to a wide range of materials from rock [39.65] to cardboard [39.66] to plastic [39.69]. The phenomenology invokes the notion of contact rejuvenation and requires the solution of coupled differential equations in terms of rate (instantaneous velocity) and state (interface order or age) variables [39.70, 71]. Quite possibly, microscale friction will have similar dependencies. As an initial effort to explore some of these potential dependencies, we recently developed a platform to study whether RSF applies to this scale, where a typical mass is on the order of 10^{-10} kg.

Amontons' law states that friction is independent of load, while Coulomb's law states that dynamic friction is lower than static friction, but independent of velocity. Long ago, *Desaguliers* [39.72] observed that static friction tends to increase with time. RSF links these empirical observations and postulates that the coefficient of friction μ is a *continuous* function of state and time. Various functional forms have been proposed. The *Ruina–Dieterich* [39.68, 73, 74] *slip* law can be written as

$$\mu(\psi,\dot{x}) = \mu_0 + A\ln\left(1+\frac{|\dot{x}|}{v_c}\right) + B\ln\left(1+\frac{\psi}{\psi_c}\right), \tag{39.4}$$

where ψ represents the interface state, $\dot{x}$ the instantaneous block velocity (the overdot indicates the derivative with respect to time), μ_0 is a constant, A and B are respectively velocity strengthening and aging coefficients, and v_c and ψ_c are, respectively, velocity and time cutoffs. A coupled equation describes how ψ evolves with time. The *Dieterich–Ruina* [39.68, 73] *aging* law is

$$\dot{\psi} = 1 - \frac{|\dot{x}|\psi}{d_c}\,. \tag{39.5}$$

At rest, $\dot{x}=0$, and hence ψ is time. At constant velocity (steady sliding at steady state), $\dot{x}=v_p$, and $\psi = d_c/v_p$, where d_c is a memory length on the order of the diameters of contacting asperities in dry systems. Referring to the (1-D schematic in Fig. 39.6 with a puller at position x_p moving at velocity v_p, a spring with constant k and block of mass m_b subject to normal load N_b, we write

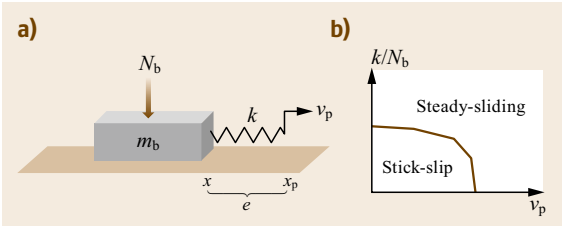


Fig. 39.6 **(a)** A puller moving at velocity v_p causes a block of mass m_b subject to normal load N_b to slide via a spring with constant k. **(b)** A schematic kinetic friction phase diagram indicating the block response as a function of v_p, k and N_b in the stick-slip or steady-sliding regimes

the block's equation of motion as

$$m_b\ddot{x} = k(x_p - x) - F_0\,, \tag{39.6}$$

where the friction $F_0 = \mu N_b + c\dot{x}$, and c is a viscous damping coefficient.

39.2.1 MEMS Rate-State Friction Test Platform with Decades of Range-in-Control Parameters

With a test platform as schematically illustrated in Fig. 39.6a, the coefficients in (39.4) can be found in principle. In order to do so, however, it is necessary that v_p, k and N_b be varied over a wide range as friction varies only slowly with velocity. While this is relatively straightforward to accomplish in macroscale experiments, its design and implementation at the microscale involves many considerations. The overarching objective in creating this apparatus is to determine quantitatively a full response of the system in [v_p, k, N_b] space using (39.4) and (39.5). One important outcome would be prediction of the stick-slip to steady sliding bifurcation line as seen in Fig. 39.6b. This would demonstrate that a microscale interface can have a predictable sliding response. Further, once the coefficients are known, in principle the response to any velocity or normal load stimulus can be known.

Figure 39.7 shows the microscale platform we have developed in pursuit of this objective [39.75]. A thermal actuator (TA) [39.76–78] is chosen as the puller. While its displacement versus voltage characteristic is nonlinear, it has the advantage that when voltage is applied, the time to achieve the final displacement value is governed by a thermal time constant, τ_H. When a voltage V_{TA} as in Fig. 39.7a is initially applied across the TA legs, a nonzero shuttle velocity is achieved almost instantaneously ($\approx 0.5\,\mu$s), but decays exponentially as the final displacement position is reached. In order to achieve a constant velocity, V_{TA} is continuously

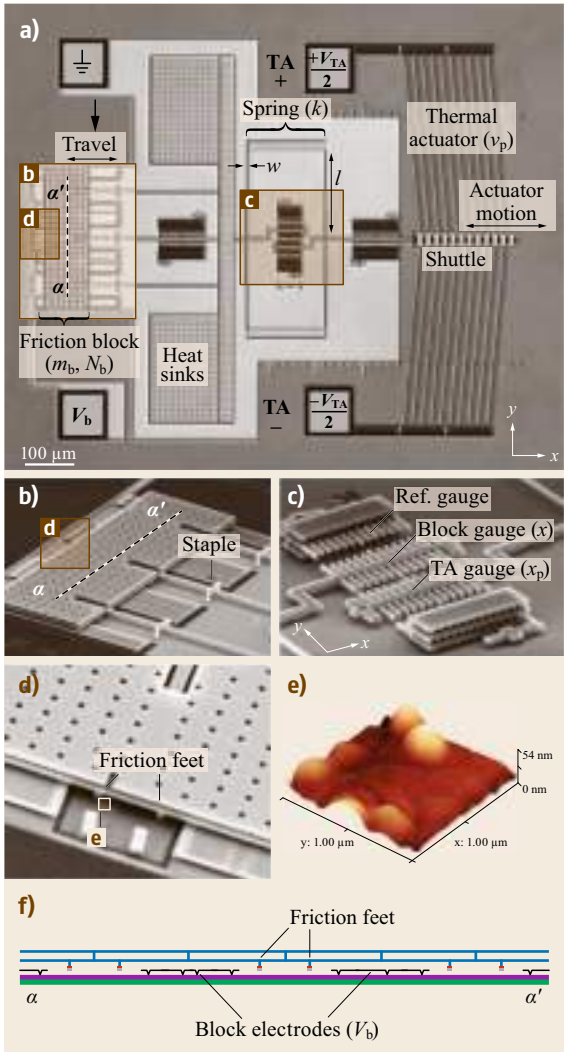


Fig. 39.7 **(a)** Optical image of a fabricated and released test platform with the basic components shown in Fig. 39.6 labeled. Here $k = 1\,\mu N/\mu m$. *Boxes* indicate details shown in the corresponding figure parts **(b–e)**. **(b)** SEM image of the friction block. Cross section α–α' is *labeled* and shown in **(f)**. The staple minimizes deflections during the release process and plays no role during testing. **(c)** SEM image of the displacement gauges used for position measurement. **(d)** SEM image showing two of the six friction feet below the friction block (shown from a different perspective than indicated in the *inset* of **(b)**). **(e)** AFM scan of the countersurface below the friction feet shown in **(d)**. The scan width is approximately half the width of the friction feet. **(f)** Schematic of cross section α–α' showing the electrically grounded friction feet and block actuation electrodes to apply electrostatic normal load

increased after a small fraction of τ_H. In detail, a training algorithm known as iterative learning control (ILC) [39.79] is applied to reach the desired velocity in a time on the order of $1/20$ of the time it takes to travel along the $\approx 10\,\mu m$ travel range. With ILC, we have trained the actuator to move at velocities from 0.1 to $3000\,\mu m/s$ with $\approx \pm 10\%$ control [39.75]. Using ILC, the actuator can also be trained to change velocity instantaneously [39.80].

Another important feature of this design is that the TA legs are much stiffer than k, so the shuttle's motion is independent of whether it is pulling a sliding block or not. The spring itself is labeled in Fig. 39.7. It consists of four fixed-guided beams and its constant k is determined by their width, w. By varying w, designs with $k = 0.1$, 1, 10 and $100\,\mu N/\mu m$ have been fabricated and tested. The TA is thermally isolated from the friction block. Its temperature is $\lesssim 2\,^\circ C$ above the substrate temperature [39.81].

The block's mass is $3.6 \times 10^{-10}\,kg$. Its gravitational normal force is negligible. The applied normal force is due to two main contributions: electrostatic loading F_{el} and adhesion, F_{adh}. Hence

$$N_b = F_{el} + F_{adh} \,. \tag{39.7}$$

The block is designed with nominally flat feet that contact the substrate over a nominal area of $A_{nom} = 1000\,\mu m^2$. Electrodes between the feet are used to apply the voltage to deliver F_{el}. The value for F_{adh} will be seen to be $\approx 1.6\,\mu N$. As measured by tapping mode AFM, the roughness of the combined top and bottom surfaces is about 8 nm RMS while the asperity radius of curvature is 35 nm. According to *Greenwood–Williamson* contact mechanics [39.27], the interface makes at most tens of contacts for the range of N_b values applied here.

The displacement gauge that is used for metrology is in the middle of the spring. Vernier scales are monitored optically through a 50× microscope objective with a high-speed camera at rates as high as 200 000 fps (frames per second). The images are analyzed using subpixel interpolation, with a resulting resolution of ± 2 nm. The spring force is simply $k\Delta e$, where Δe is the spring extension.

The experimental data that follows was measured using the Fig. 39.7 test platform coated by a 1.2 nm-thick monolayer coating (tridecafluoro-1,1,2,2-tetrahydrooctyl-*tris*(dimethylamino)silane, $CF_3C_5F_{10}C_2H_4Si(N(CH_3)_2)_3$, FOTAS) deposited from the vapor phase [39.82]. The 8-carbon chain molecule has a sufficiently high vapor pressure to allow vapor-phase deposition, which makes it highly reproducible in a manufacturing setting. However, it is too short to

form a self-assembled monolayer; rather the chains attached to the surface are believed to be somewhat disordered [39.83]. This lack of order may play an important role in the effects reported below.

A wear test was first conducted. In this test, with $F_e = 5$ or 250 μN, it was seen that $\mu \approx 0.3$ remained stable for 20 000 cycles for FOTAS-coated devices. For uncoated devices, $\mu \approx 0.65$ at $F_e = 5$ μN was also stable up to 10 000 cycles. In comparison, RSF tests require approximately 100 trials from a given device for complete characterization. Therefore, we can be optimistic that wear is not an issue with respect to assessing RSF with this test platform. We attribute the relatively low wear rate to the lightly loaded interface in which different asperities share the load over the sliding travel distance.

39.2.2 Rate-State Friction Measurements

Transition from Steady Sliding to Stick-Slip

We pull the block at puller speed $v_p = 1$ μm/s using a spring constant $k = 10$ μN/μm while applying electrostatic load F_e, and measure elongation e. The camera framerate is 300 fps, corresponding to a frame every 3.3 nm of puller position. The data in Fig. 39.8 begin at $F_e = 0$ μN. There, the extension is approximately constant. At $F_e = 0.3$ μN, the extension exhibits somewhat greater noise. At $F_e = 0.58$ μN, the data exhibit discontinuities in e, corresponding to slip events, and at $F_e = 0.90$ μN, the discontinuities are somewhat larger. Finally at $F_e = 2.01$ μN, the response is similar to a classic stick-slip curve. Dashed black lines indicate perfect stick of the block. Hence at $N_b = 2.01$ μN, there is frictional creep, but at a slower rate than at the lower F_e values of 0.58 and 0.90 μN. This creep is widely observed [39.66, 83, 84] and is indicative of an avalanche process leading to inertial slip.

A metric to deduce the transition between these regimes must be developed. For $N_b = 2.01$ μN in Fig. 39.8, a slip distance δ is indicated. If we consider the distribution of e data for a perfect stick-slip curve, it will be uniform. The standard deviation of the e data multiplied by $\sqrt{12}$ will equal the width of the (uniform) e distribution. We call this product the pseudo-slip distance.

In Fig. 39.9a, we plot the pseudo-slip distance against F_e for FOTAS-coated and oxide coated blocks. For the former, a small but clear abrupt drop in the pseudo slip distance is seen as F_e is reduced. We use this drop to estimate the transition from stick-slip to steady sliding. The uncoated block in Fig. 39.9a displays only stick behavior, and there is no abrupt drop in δ. Choosing the abrupt drop as a metric, the data in Fig. 39.8 for $F_e = 0$ and 0.3 μN correspond to steady sliding. We attribute the small deviations in e to local differences in the topography encountered, which can be important as there are likely only a few contacting asperities at these low F_e values.

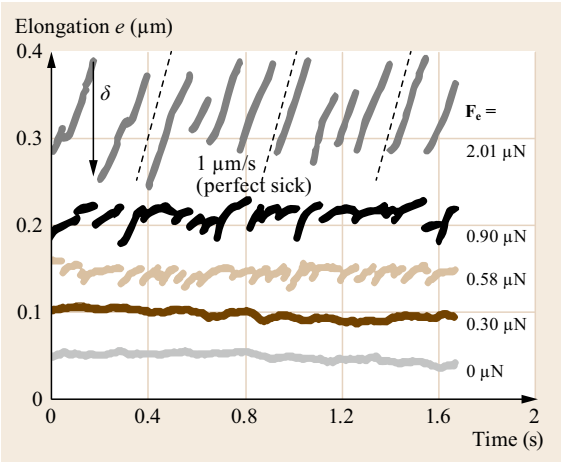


Fig. 39.8 Spring elongation versus time for increasing F_e. Here, $v_p = 1$ μm/s, $k = 10$ μN/μm and the camera framerate is 300 fps

The maximum extension e_{max} for the uncoated and coated blocks is plotted against F_e in Fig. 39.9b. We fit the data to (39.7), and find values $F_{adh} = 2.0$ and 1.6 μN for the two surfaces. With $A_{nom} = 1000$ μm^2, this corresponds to an adhesive pressure of 2.0 and 1.6 kPa, respectively. Such pressures are typical of rough micromachined interfaces, and are attributed not to the contacting asperities but to van der Waals forces across the relatively vast area of surface that are separated by just a few nm [39.85]. These F_{adh} values are not interpreted in the context of RSF, but provide a reasonable correction to N_b. F_{adh} sets a lower bound on N_b. All subsequent measurements discussed are only for FOTAS-coated blocks, and the N_b value including F_{adh} is reported.

We now present five tests: M1–M3, V1 and V2. Here M stands for measurement, while V stands for validation.

Test M1, Friction Versus Velocity. In the results shown in Fig. 39.10, v_p is varied from 0.1 to 3000 μm/s, while $k = 100$ μN/μm and $N_b = 14.1$ μN are held fixed. The stiff spring ensures steady sliding while the relatively high N_b value gives rise to an extension e that can be measured well. A frame was taken every 100 nm of travel, which corresponds to 30 000 fps at $v_p = 3000$ μm/s. The average and standard deviation of the block velocity are shown in the figure. There is a clear trend in which dynamic friction μ_d decreases from 0.1 to ≈ 300 μm/s, and increases thereafter.

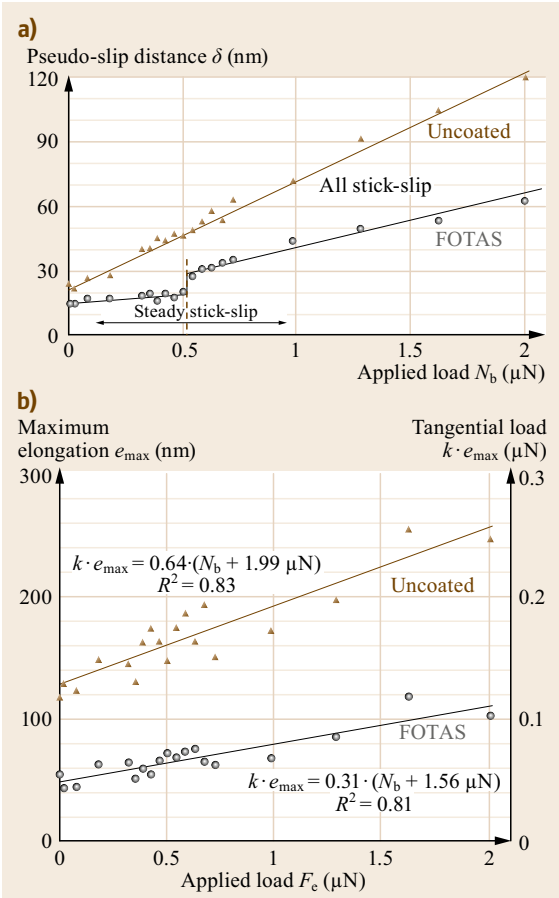


Fig. 39.9 **(a)** δ versus N_b for constant $v_p = 1\,\mu m/s$ and $k = 10\,\mu N/\mu m$ for uncoated (*triangles*) and FOTAS-coated (*circles*) devices. **(b)** e_{max} versus N_b for the same data set

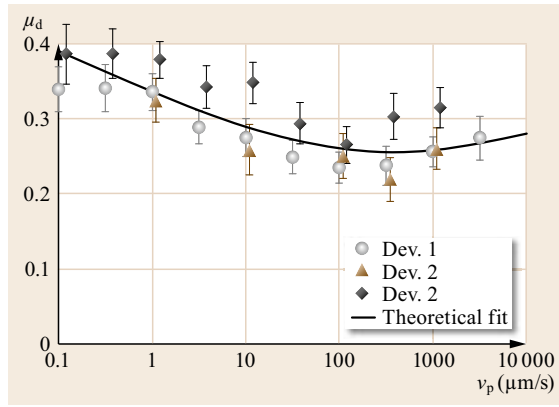


Fig. 39.10 Dynamic friction coefficient $\mu_d(v)$ versus actuator pulling velocity in the steady-sliding regime ($v = v_p$). Three devices are shown. Devices 2 and 3 are offset by $0.1\,v_p$ and $0.2\,v_p$, respectively. Here $k = 100\,\mu N/\mu m$ and $N_b = 14.12\,\mu N$. *Error bars* represent one standard deviation of μ_d by considering the time-weighted variation in spring elongation of one or more trials

Equations (39.4) and (39.5) are used to determine values of $\mu_0 = 0.19$, $A = 0.013$ and $B = 0.024$ from this data, as represented by the best fit solid black curve in Fig. 39.10. We also use values of $v_c = 10\,\mu m/s$, $\psi_c = 10\,\mu s$ and $d_c = 4.3\,nm$. The fit is only weakly dependent of the v_c and ψ_c values, and they are chosen such that v_p/v_c and $(\psi = d_c/v_p)/\psi_c$ are both $\gg 1$ in the relevant ranges. Meanwhile, d_c is determined as shown below. Note that in the fit to (39.4), μ initially decreases because $B > A$. That is, the asperity contact lifetime decreases as v_p increases, and hence less aging is responsible for the friction decrease. As v_p increases, this effect becomes small compared to μ_0 and the velocity strengthening effect due to A becomes significant.

Test M2, Velocity Step Response. Let us again consider (39.4) and (39.5). Upon an instantaneous change in v_p, state ψ remains the same. If $A > 0$, we expect a step increase in μ equal to $A \ln(v_{p2}/v_{p1})$, as shown in Fig. 39.11a. As represented in Fig. 39.11b, v_p was changed in $\approx 1\,\mu s$ from 50 to $1000\,\mu m/s$ with $k = 100\,\mu N/\mu m$ and $N_b = 15.4\,\mu N$. Here, imaging at 200 000 fps was necessary to capture the transient event, as its duration is $\approx 20\,\mu s$. The response data from three different trials is seen in Fig. 39.11c, where it is magnified in the offset on the right. An increase in e of approximately 5 nm occurs, corresponding to an increase in μ of $ke/N_b = 0.032$. Equating this with $A \ln(1000/50)$, we find $A = 0.011$. This is slightly lower than the value $A = 0.013$ reported above. The discrepancy is because in Fig. 39.11a it is assumed that the block velocity can change instantaneously. This is only possible if the spring is rigid, which would then mean that e could not be measured. In the actual situation, the state has had some time to change, so μ begins to decrease. A full numerical simulation incorporating (39.4)–(39.6) better estimates the value of A [39.86]. The final value of $A = 0.013$ from the data in Fig. 39.11c is in good agreement with the value obtained from Test M1.

As indicated by Fig. 39.11a, a fraction of the width of the response pulse can be used to estimate the memory length, d_c. That width gives $d_c \approx 0.5 \times (20\,\mu s) \times (1000\,\mu m/s) = 10\,nm$. The full numerical simulation, which is optimized to capture the response curve, gives $d_c = 4.3\,nm$. The value from GW contact mechanics is $d_c \approx 13\,nm$. However, it is difficult to get a good measure of asperity radius R from AFM because of tip convolution effects – the GW calculation may overestimate R. On the other hand, the monolayer chain length is $\approx 1.2\,nm$. The memory length is associated

with the notion that during sliding, asperities are new when they first make contact, and then age as they slide past each other. The contribution to μ is then due to the effective age d_c/v_p. The implicit assumption is that bonding is instantaneous when new asperities come into contact and subsequently make and break bonds. Poorly ordered monolayers such as FOTAS may reconfigure when brought into contact with a counterface. Hence, the value of $d_c = 4.3\,\text{nm}$ may reflect information on bonding dynamics as well as contact diameter. Work to understand the memory length in the context of microscopic processes is a topic of current interest [39.87].

Test M3, Static Aging. In this test, the block is held at $N_b = 300\,\mu\text{N}$ under zero spring force for a time ranging from 0.1 to 3120 s. After the hold time, the puller travels at $1\,\mu\text{m/s}$ and exerts a force ke (here $k = 100\,\mu\text{N}/\mu\text{m}$) until a slip event is observed at extension e_{crit}. The total time is taken to be the aging time t_{age}, and the static friction coefficient μ_s is then equal to ke_{crit}.

The resulting data is seen in Fig. 39.12, in the area annotated as *static aging trials*. Each data point is the average of about ten trials, except at the longest time of 3120 s where just three points were taken. The spread for each data point represents one standard deviation, and the dashed lines give the 95% confidence intervals for the average from a *t*-test. A logarithmic fit of the form $\mu_s = \mu_0 + B\ln(t_{age}/1\,\text{s})$, with $B = 0.0041 \pm 0.0016$ (95% confidence) and $R^2 = 0.80$. Hence, at long times, μ_s appears to be saturating. This could be indicative of the monolayer chains approaching their equilibrium configurations.

The data from Fig. 39.10 is replotted in Fig. 39.12, where time is taken to be d_c/v_p. The data from Tests M1 and M3 (dynamic and static friction measurements, respectively) extend towards each other, implying that the memory length concept applies during the sliding test. At short times, $B = 0.024$ is a better fit to the data, supporting the idea that the aging is saturating at long times.

Test V1, Crossing the Bifurcation Line with N_b. In Fig. 39.8, including the effect of F_a, the bifurcation line was crossed at $N_b \approx 2\,\mu\text{N}$ with $v_p = 1\,\mu\text{m/s}$. In Fig. 39.13, we see that at $v_p = 50\,\mu\text{m/s}$, the line is crossed at $N_b \approx 4\,\mu\text{N}$. This indicates a negative slope in the bifurcation line, as qualitatively indicated from Fig. 39.6b. The data now approaches the perfect stick lines, as the block has less time to creep.

Test V2, Crossing the Bifurcation Line with v_p. Using $N_b = 3.45\,\mu\text{N}$ and $k = 10\,\mu\text{N}/\mu\text{m}$, we see in Fig. 39.14 that the bifurcation line is also crossed

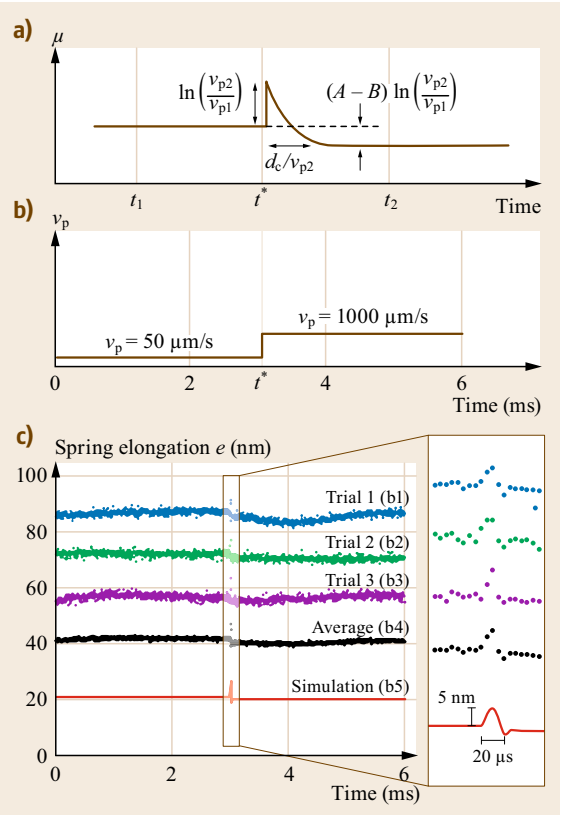


Fig. 39.11 **(a)** Expected friction versus time for a velocity step. **(b)** Pulling velocity versus time showing a velocity step at $t^* = 3\,\text{ms}$ ($k = 100\,\mu\text{N}/\mu\text{m}$, $N_b = 15.42\,\mu\text{N}$). **(c)** Spring elongation versus time for the pulling profile shown in **(b)**. Multiple trials (b1)–(b3) as well as the average (b4) are shown

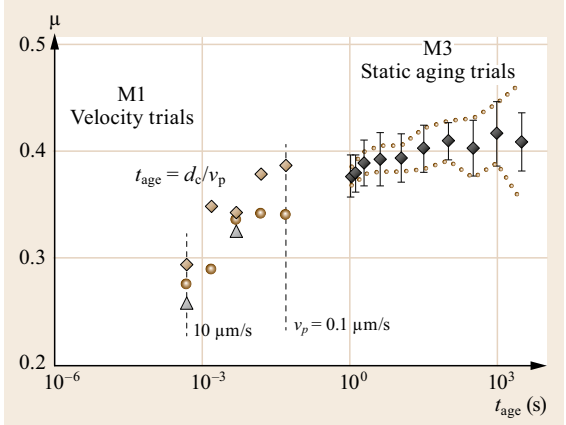


Fig. 39.12 Results of the static aging trials, plotted alongside the results of the velocity trials shown in Fig. 39.10 with $t_{age} = d_c/v_p$

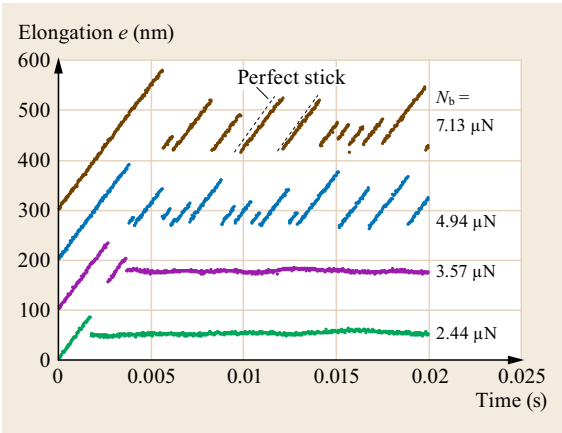


Fig. 39.13 Elongation versus time for a fixed $v_p = 50\,\mu m/s$ and $k = 10\,\mu N/\mu m$ for increasing N_b. The camera framerate was 30 000 fps. At this v_p, a transition between stick-slip and steady sliding is seen between $N_b = 3.57$ and $4.94\,\mu N$

between $v_p = 30$ and $40\,\mu m/s$. The black dashed horizontal lines indicate numerical modeling results of the maximum and minimum extensions taking the full RSF parameters and equations of motion into account. These lines indicate the average expected values. The data does show that the amplitude tends to decrease as v_p increases. The deviations from the average prediction observed in the data likely indicate that the actual topography sampled from point to point varies along the block trajectory.

Quantitative Bifurcation Line Prediction

The equation of motion incorporating the RSF coefficients can now be written. A perturbation on the extension e is applied to a condition where steady sliding exists. An exponential solution is assumed for the subsequent e time dependence. Negative roots mean that there will be an oscillatory decay back to steady sliding, while positive roots indicate a divergent response and that the system will take on a stick-slip response. The bifurcation line divides these two responses. The perturbation theory result is [39.88]

$$k_{crit} = N_b \left(1 + \frac{v_p m_b (v_p + v_c)}{N_b A d_c}\right) \times \left(\left(\frac{B}{d_c + v_p \psi_c}\right) - \left(\frac{A}{v_p + v_c}\right)\frac{v_p}{d_c}\right). \tag{39.8}$$

The values for the parameters A, B, d_c, v_c and ψ_c were determined from Tests M1 and M2 as described above. For a given value of k, $1/N_b$ can be plotted against v_p. This has been done in Fig. 39.15 using further data from Tests V1 and V2 above. Some 20 devices, different from the ones used to determine the parameters, were used to find the bifurcation experimentally. It is seen that the experimental results agree well with the predicted curve. While this is expected from theory, experimentally this means (1) that wear is sufficiently low over the ≈ 100 trials needed to characterize the experimental parameters, and (2) that the parameters apply to devices other than those measured, and therefore that the kinetic phase diagram has been well quantified.

RSF Summary

In this section, we have seen that RSF is a useful phenomenological framework for measuring the time and velocity dependence of friction not only at the scale

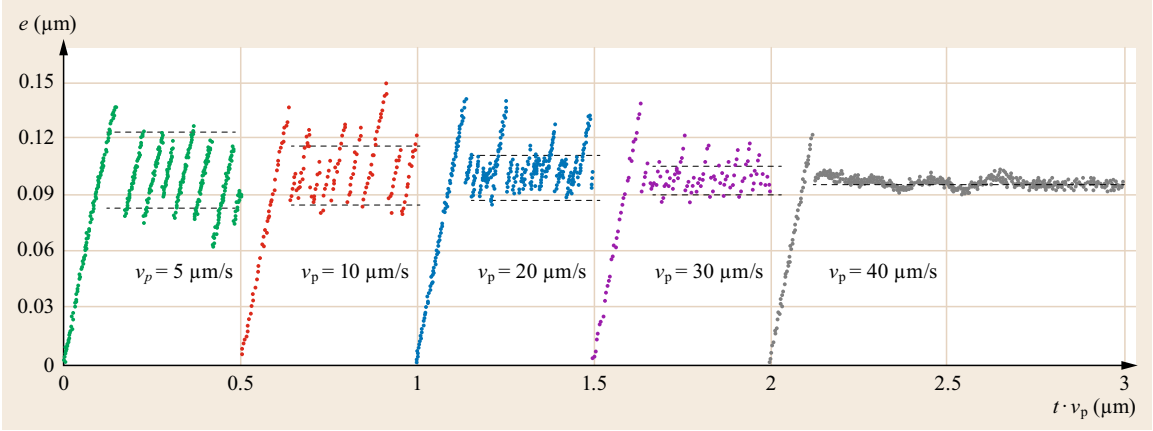


Fig. 39.14 Spring elongation versus actuator position $t\,v_p$ for a representative device at various velocities. Here $k = 10\,\mu N/\mu m$ and $N_b = 3.45\,\mu N$. RSF simulation maximum and minimum elongations are represented by *dashed lines*

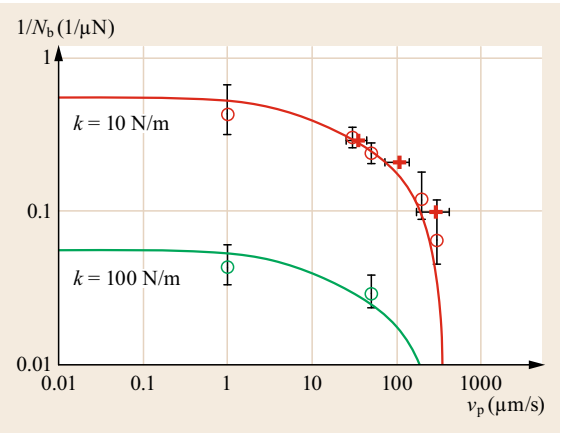


Fig. 39.15 Predicted bifurcation lines and average experimental transition between stick-slip and steady sliding via normal load variation (*circles*, *vertical error bars*) or velocity variation (*crosses*, *horizontal error bars*). *Error bars* represent one standard deviation of the critical velocity or inverse normal force at which the transition occurs ◀

of earthquakes [39.89] or macroscale objects (≈ 1 kg with μm-scale asperities) [39.66], but also at the scale of microscale objects ($\approx 10^{-10}$ kg with nm-scale asperities). Although RSF predicts average friction well at the microscale, very likely only a few asperities are in contact. This gives rise to noise in the friction behavior, which requires further study. Another area of interest is the dependence of the RSF parameters on the particular monolayer chain that is applied to the surface. Although bonding across the interface may lower the free energy, the chain's entropy is also lowered. The dynamical interplay between these factors will be interesting to study. It will also be of interest to see if and how nanotribology results can be linked to RSF.

39.3 Putting MEMS Friction to Use: Making Motors and Friction Instruments

39.3.1 Microengines, Pumps and Generators

A long-held dream is to fabricate reliable gas turbine engines from MEMS technology [39.90]. Such engines would theoretically deliver a thrust to weight ratio 100 times that of their macroscale counterparts. Most of the early work in this area involved high-speed dynamics control, with engines working above 1 million rpm on a cushion of air [39.91]. These engines were fabricated from multiple layers of silicon. The design allowed no taper during very deep etches and hence the manufacturing margin was low.

More recently, *Ghodssi* et al. put ball bearings to use in a six-phase electric micromotor that generated 300 μW at 500 rpm [39.92]. One advantage of the design is that the bearings allow precise control of rotation to within a few degrees, which is important for certain sensor designs [39.93]. Another is that fabrication tolerance need not be controlled as tightly. In principle, such machines could run up to 200 000 rpm [39.93]. Microscale pumps with ball bearings have also been investigated. They could feed hydrogen or alcohol to miniaturized fuel cells for electrical generation or microturbines for torque generation. Electric generators can also be built with this technology.

In the designs, a silicon raceway surrounds stainless steel microballs of 285 μm diameter. Friction is characterized using a *spin-down* test [39.94]. The friction platform is schematically represented in Fig. 39.16a. The rotor spins about the axis shown as driven by gas pressure that flow through stator vanes. The microballs enable the spinning and fill 85% of the racetrack. Normal force is provided by gas pressure through the thrust plenum flow, as shown in Fig. 39.16b. An optical probe measures radial displacement. Figure 39.16c shows spin-down data after the turbine flow is shut off. A simple friction torque model captures the data very well in this case.

Similar spin-down tests after increasing numbers of cycles (up to 200 million cycles) were used to characterize wear with raceways coated by SiC and TiN [39.95]. The SiC coating was quickly removed due to adhesive wear, and the balls adhered to the underlying silicon surface. The TiN coating wore gradually in select areas. Bare silicon raceways experienced high pressure, as evidence by a phase change to an amorphous phase as evidenced by Raman spectroscopy. The US Army is now supporting this research. Future work will address the ball strength and study the effect of molecular monolayers.

39.3.2 In-situ TEM for Friction Investigations

While driven by applications such as accelerometers, gyroscopes and displays, microtechnology evolution has also been heavily leveraged to develop ever more sophisticated scientific instruments. One example i the

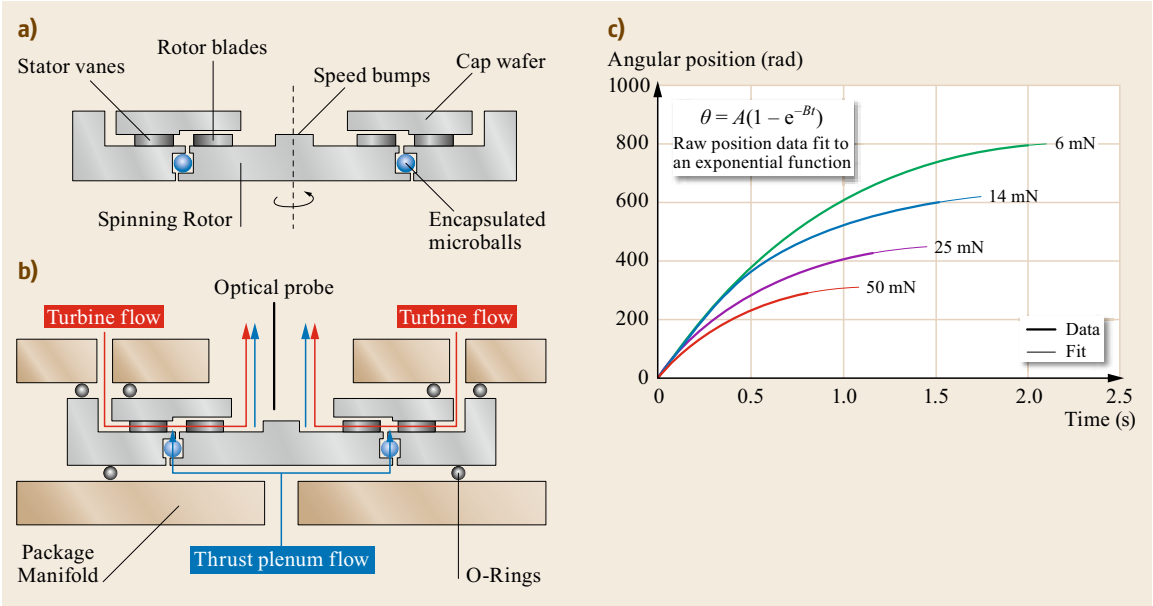


Fig. 39.16a–c Ball bearing friction test platform. **(a)** The rotor spins about the axis shown as driven by gas pressure that flow through stator vanes. The microballs enable the spinning. **(b)** Normal force is provided by gas pressure through the thrust plenum flow. An optical probe measures radial displacement. **(c)** Output data as a function of normal load from 6 to 50 mN (after [39.94])

push-to-pull indenter available from Hysitron [39.96, 97]. Transmission electron microscopy (TEM) has been applied to gain detailed images of the tips that are so widely used in FFM, both before and after use [39.98]. Wear has been analyzed by a transition state theoretical approach [39.99]. In-situ TEM friction force tools have also been recently been constructed [39.100–102]. The MEMS in-situ TEM platform of *Sato* et al. [39.102] has enabled real-time TEM imaging of shearing action on a Ag-Ag nanoasperity junction. This method may allow detailed observations of shear-induced dislocation motion at the interface versus in the bulk.

39.4 Wear and Tribopolymer Evolution in Micro- and Nanoswitches

We now present some closing thoughts on another nanotribology issue of great current interest. Micro- and nanoswitches afford many potential advantages including high switching speeds (ns to μs), low power operation (aJ to nJ per cycle), and high-temperature operation. Microswitch applications envisaged include radio-frequency phased array radars [39.103] and downstream circuit breakers [39.104]. Nanoswitches are attracting attention as candidate devices to complement metal oxide semiconductor field effect transistors (MOSFETs) in digital logic [39.105, 106]. MOSFETs are limited to ≈ 1 V operation, and begin to leak significantly as the voltage is reduced, creating static power losses [39.107]. Some nanoswitch designs employ electrostatic actuation and can switch in 10–20 ns [39.108] at gate voltages < 10 V while other designs use piezoelectric materials [39.109] and can be switched at voltages as low as 1 mV. These devices have shown negligible off-state leakage. Significant dynamic power savings can also be realized with nanoswitches. An important challenge in commercializing *microswitches* is that the electrical contact resistance (ECR) must often remain low (≈ 1 Ω) and stable over 10^8–10^9 cycles [39.103, 110]. This requirement may be relaxed for *nano*switch digital logic applications because even at several kΩ, mechanical delay time is large compared to electrical delay time [39.106]. However, to fulfill industry needs, they must also function reliably up to 10^{16} cycles. Their low-power consumption makes them exciting candidates for sensors in the developing internet of things (IoT) devices [39.111].

Given the daunting challenge of ultrahigh cycles, fundamental insight into the underlying reliability mechanisms must be gained. One important issue is ma-

terial wear [39.113]. We close here with some thoughts on the more poorly understood issue of tribopolymer (TP) development. It is well- known from *Hermance* and *Egan* [39.114] and subsequent work [39.115] that switch resistance is sensitive to trace levels of organic vapors, which can develop into TP when subject to shear. Such vapors are typically present in the atmosphere at the parts per billion (ppb) level. Because of the nN forces and mV levels in nanoswitches, a thin adsorbed layer must be considered as a potential cause of high resistance. TPs can deposit on nearly any material from both short- and long-chain molecules [39.114].

The ambient background may play an important role in their adsorption and evolution as well. *Brand* et al. have performed tests in both N_2 and $N_2 : O_2$ ambience [39.112, 116]. Figure 39.17 shows how the ECR of vacuum-baked switches increases when cycled in 0.25% by volume benzene. For Pt-coated switches, ECR rapidly rises with cycle count in an N_2 environment (Fig. 39.17a), but more slowly in $N_2 : O_2$ (Fig. 39.17b). RuO_2 is reported to be a catalytically *deactivated* surface [39.117] that has been used to enhance reliability in reed relays, but in an N_2 environment, ECR rises quickly (Fig. 39.17c). However, in an $N_2 : O_2$ environment, ECR remains stable at just 30 Ω in spite of the very high benzene concentration. Much less TP was generated in the case of Fig. 39.17d compared to Fig. 39.17a–c as indicated by SEM imaging [39.112]. Bridging oxygens may act to desorb the benzene and are subsequently replenished in the $N_2 : O_2$ atmosphere. Another important observation is that these switches were cycled under cold switching conditions, in which 5 V is applied to the contacts only after they were closed. It appears that the TP breaks down electrically and forms a graphitic layer that conducts [39.118]. In the absence of current, a thick insulating layer builds up even under repetitive normal load cycling.

Materials, environment and cycling conditions each play a significant role in micro- and nanoswitch reliability. This short overview indicates that fundamental understanding of the mechanisms leading to contact degradation may enable a tribological interface to be engineered that will meet device reliability requirements.

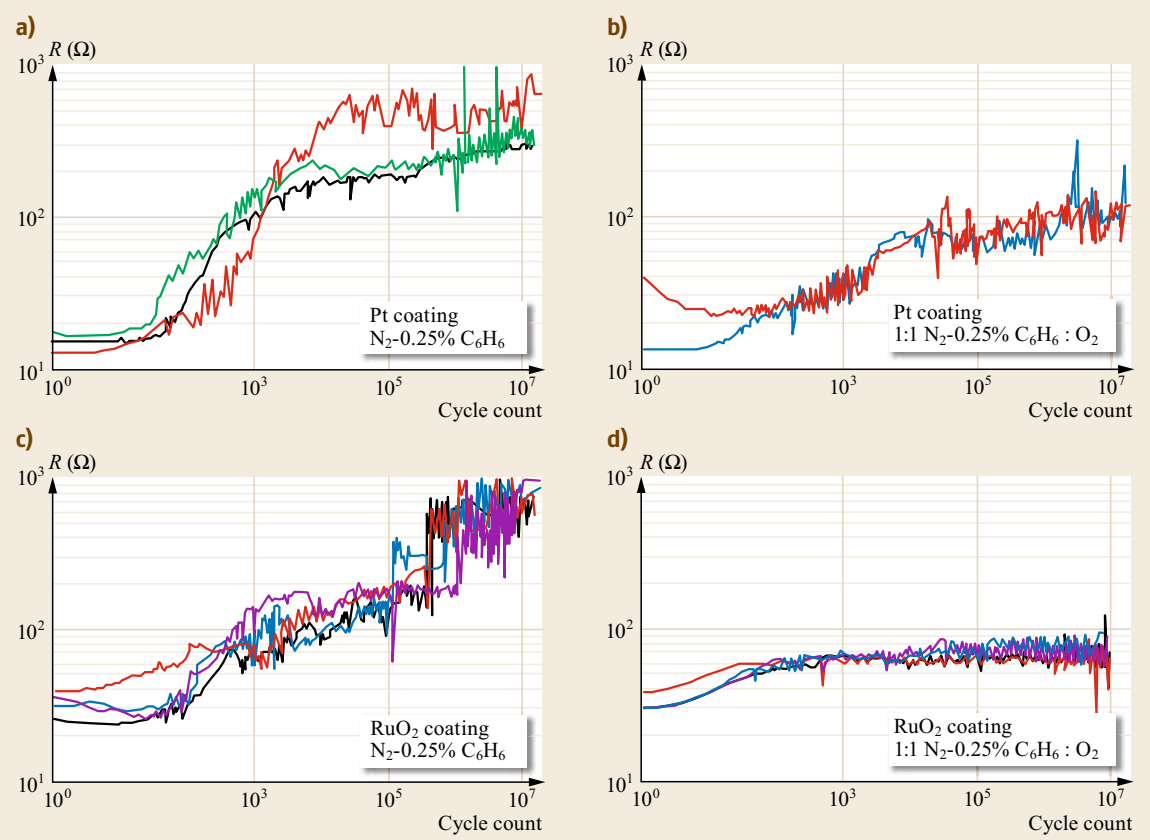


Fig. 39.17 ECR versus cycle count for Pt and RuO_2 coating in N_2-0.25% C_6H_6 and 1 : 1 N_2-0.25% $C_6H_6 : O_2$ atmospheres (after [39.112])

39.5 Concluding Remarks

We have reviewed the background and taken a whirlwind tour of the status and recent progress in micro- and nanotribology. Understanding and solving the challenging issues in this area could enable components such as forceful linear actuators and micropowered pumps, motors and generators to come into use. Low-power switches may become integral to IoT sensors. In order to attain these goals, further progress towards a fundamental understanding of the links between interfacial shear and normal stress and the effect of intended (and unintended) lubricants is needed. Rate-state phenomenology is a promising framework for understanding friction at many length scales.

Acknowledgments. This work was partially supported by NSF CMMI Grants 1030322, 1030278 and 1334572. SSS was supported by the NSF Graduate Research Fellowship Program. Many of the MEMS devices discussed were fabricated by Sandia National Laboratories, Albuquerque, NM, USA. Certain commercial equipment, instruments, or materials are identified in this chapter in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

39.1 M.S. Rodgers, J.J. Sniegowski: Five-level polysilicon surface micromachine technology: Application to complex mechanical systems. In: *Proc. Solid-State Sens. Actuator Workshop, Hilton Head Island* (1998) pp. 144–149

39.2 J.J. Sniegowski, M.P. de Boer: IC-compatible polysilicon surface micromachining, Annu. Rev. Mater. Sci. **30**, 299–333 (2000)

39.3 L. Eldada: Advances in telecom and datacom optical components, Opt. Eng. **40**, 1165 (2001)

39.4 E. Sarajlic, C. Yamahata, E. Berenschot, N. Tas, H. Fujita, G. Krijnen: High-performance shuffle motor fabricated by vertical trench isolation technology, Micromachines **1**, 48–67 (2010)

39.5 P. Vettiger, M. Despont, U. Drechsler, U. Durig, W. Haberle, M.I. Lutwyche, H.E. Rothuizen, R. Stutz, R. Widmer, G.K. Binnig: The millipede – More than one thousand tips for future AFM data storage, IBM J. Res. Dev. **44**, 323 (2000)

39.6 J.M. Zara, S.W. Smith: A micromachine high frequency ultrasound scanner using photolithographic fabrication, IEEE Trans. Ultrason. ferroelectr. Freq. Control **49**, 947–958 (2002)

39.7 T. Akiyama, D. Collard, H. Fujita: Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS, J. Microelectromech. Syst. **6**, 10 (1997)

39.8 E. Sarajlic: *Electrostatic Microactuators Fabricated by Vertical Trench Isolation Technology*, Ph.D. Thesis (University of Twente, Eusclede 2005)

39.9 M. Naraghi, T. Ozkan, I. Chasiotis, S.S. Hazra, M.P. de Boer: MEMS platform for on-chip nanomechanical experiments with strong and highly ductile nanofibers, J. Micromech. Microeng. **20**, 125022 (2010)

39.10 T. Akiyama, K. Shono: Controlled stepwise motion in polysilicon microstructures, J. Microelectromech. Syst. **2**, 106 (1993)

39.11 R.J. Linderman, V.M. Bright: Nanometer precision positioning robots utilizing optimized scratch drive actuators, Sens. Actuators A **91**, 292 (2001)

39.12 N. Tas, J. Wissink, L. Sander, T. Lammerink, M. Elwenspoek: Modeling, design and testing of the electrostatic shuffle motor, Sens. Actuators A **70**, 171 (1998)

39.13 M.P. de Boer, D.L. Luck, W.R. Ashurst, A.D. Corwin, J.A. Walraven, J.M. Redmond: High-performance surface-micromachined inchworm actuator, J. Microelectromech. Syst. **13**, 63 (2004)

39.14 E. Sarajlic, E. Berenschot, H. Fujita, G. Krijnen, M. Elwenspoek: Bidirectional electrostatic linear shuffle motor with two degrees of freedom. In: *Proc. IEEE MEMS* (2005) pp. 391–394

39.15 G. Amontons: On the resistance originating in machines (in French), Mém. Acad. R. A **0**, 206–222 (1699)

39.16 F.P. Bowden, D. Tabor: *Friction and Lubrication of Solids: Part I* (Oxford Univ. Press, Oxford 1950)

39.17 H. Hertz: On the contact of elastic solids, J. Reine Angew. Math. **92**, 156–171 (1881)

39.18 K.L. Johnson, K. Kendall, A.D. Roberts: Surface energy and the contact of elastic solids, Proc. R. Soc. A **324**, 301 (1971)

39.19 B.V. Derjaguin, V.M. Muller, Y.P. Toporov: Effect of contact deformations on the adhesion of particles, J. Colloid Interface Sci. **53**, 314 (1975)

39.20 D. Tabor: Surface forces and surface interactions, J. Colloid Interface Sci. **58**, 2–13 (1977)

39.21 D. Maugis: Adhesion of spheres: The JKR-DMT transition using a Dugdale model, J. Colloid Interface Sci. **150**, 243–269 (1992)

39.22 R.W. Carpick, N. Agrait, D.F. Ogletree, M. Salmeron: Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscope, J. Vac. Sci. Technol. B **14**, 1289–1295 (1996)

39.23 M. Enachescu, R.J.A.V.D. Oetelaar, R.W. Carpick, D.F. Ogletree, C.F.J. Flipse, M. Salmeron: Obser-

vation of proportionality between friction and contact area at the nanometer scale, Tribol. Lett. **7**, 73–78 (1999)
39.24 M.A. Lantz, S.J. O'Shea, M.E. Welland, K.L. Johnson: Atomic force microscope study of contact area and friction on $NbSe_2$, Phys. Rev. B **55**, 10776–10785 (1997)
39.25 J.F. Archard: Contact and rubbing of flat surfaces, J. Appl. Phys. **24**, 981–988 (1953)
39.26 J.F. Archard: Elastic deformation and the laws of friction, Proc. R. Soc. A **243**, 190–205 (1957)
39.27 J.A. Greenwood, J.B.P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. A **295**, 300–319 (1966)
39.28 K.N.G. Fuller, D. Tabor: The effect of surface roughness on the adhesion of elastic solids, Proc. R. Soc. A **345**, 327 (1975)
39.29 D. Maugis: On the contact and adhesion of rough surfaces, J. Adhes. Sci. Technol. **10**, 161 (1996)
39.30 G. Carbone, F. Bottiglione: Asperity contact theories: Do they predict linearity between contact area and load?, J. Mech. Phys. Solids **56**, 2555–2572 (2008)
39.31 L. Pastewka, M.O. Robbins: Contact between rough surfaces and a criterion for macroscopic adhesion, Proc. Natl. Acad. Sci. USA **111**, 3298–3303 (2014)
39.32 B.N.J. Persson: On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion, J. Phys.: Condens. Matter **17**, 1071–1142 (2005)
39.33 J.I. McCool: Comparison of models for the contact of rough surfaces, Wear **107**, 37 (1986)
39.34 M.G. Lim, J.C. Chang, D.P. Schultz, R.T. Howe, R.M. White: Polysilicon microstructures to characterize static friction. In: *Proc. IEEE MEMS Workshop, Napa Valley* (1990) pp. 82–88
39.35 W.R. Ashurst, Y.J. Jang, L. Magagnin, C. Carraro, M.M. Sung, R. Maboudian: Nanometer-thin titania films with SAM-level stiction and superior wear resistance for reliable MEMS performance. In: *Proc. 17th IEEE Int. Conf. MEMS* (2004) pp. 153–156
39.36 W.R. Ashurst, C. Yau, C. Carraro, C. Lee, R.T. Howe, R. Maboudian: Alkene based monolayer films as anti-stiction coatings for polysilicon MEMS, Sens. Actuators A **9**, 239 (2001)
39.37 D.C. Senft, M.T. Dugger: Friction and wear in surface-micromachined tribological test devices, Proc. SPIE **3224**, 31–38 (1997)
39.38 W.M.V. Spengen, J.W.M. Frenken: The Leiden tribometer: Real time dynamic friction loop measurements with an on-chip tribometer, Tribol. Lett. **28**, 149–156 (2007)
39.39 N.R. Tas, C. Gui, M. Elwenspoek: Static friction in elastic adhesion contacts in MEMS, J. Adhes. Sci. Technol. **17**, 547 (2003)
39.40 S.J. Timpe, K. Komvopoulos: The effect of adhesion on the static friction properties of sidewall contact interfaces of microelectromechanical devices, J. Microelectromech. Syst. **15**, 1612 (2006)
39.41 S.J. Timpe, K. Komvopoulos: Microdevice for measuring friction and adhesion properties of sidewall contact interfaces of microelectromechanical systems, Rev. Sci. Instrum. **78**, 065106-1–065106-9 (2007)
39.42 W.R. Ashurst, M.P. de Boer, C. Carraro, R. Maboudian: An investigation of sidewall adhesion in MEMS, Appl. Surf. Sci. **212/213**, 735–741 (2003)
39.43 R.W. Carpick, D.F. Ogletree, M. Salmeron: A general equation for fitting contact area and friction vs load measurements, J. Colloid Interface Sci. **211**, 395–400 (1997)
39.44 G.J. Leggett: Friction force microscopy of self-assembled monolayers: Probing molecular organisation at the nanometre scale, Anal. Chim. Acta **479**, 17–38 (2003)
39.45 Y.L. Chen, J.N. Israelachvili: Effects of ambient conditions on adsorbed surfactant and polymer monolayers, J. Phys. Chem. **96**, 7752 (1992)
39.46 J.B.D. Green, M.T. McDermott, M.D. Porter, L.M. Siperko: Nanometer-scale mapping of chemically distinct domains at well-defined organic interfaces using frictional force microscopy, J. Phys. Chem. B **99**, 10960–10965 (1995)
39.47 G.Y. Liu, P. Fenter, C.E.D. Chidsey, D.F. Ogletree, P. Eisenberger, M. Salmeron: An unexpected packing of fluorinated *n*-alkane thiols on Au(111): A combined atomic-force microscopy and x-ray diffraction study, J. Chem. Phys. **101**, 4301 (1994)
39.48 M.T. McDermott, J.B.D. Green, M.D. Porter: Scanning force microscopic exploration of the lubrication capabilities of *n*-alkanethiolate monolayers chemisorbed at gold: Structural basis of microscopic friction and wear, Langmuir **13**, 2504–2510 (1997)
39.49 E.W.V.D. Vegte, G. Hadziioannou: Scanning force microscopy with chemical specificity: An extensive study of chemically specific tip–surface interactions and the chemical imaging of surface functional groups, Langmuir **13**, 4357–4368 (1997)
39.50 Y. Liu, D.F. Evans, Q. Song, D.W. Grainger: Structure and frictional properties of self-assembled surfactant monolayers, Langmuir **12**, 1235–1244 (1996)
39.51 B.D. Beake, G.J. Leggett: Friction and adhesion of mixed self-assembled monolayers studied by chemical force microscopy, Phys. Chem. Chem. Phys. **1**, 3345–3350 (1999)
39.52 H.I. Kim, T. Koini, T.R. Koini, S.S. Perry: Systematic studies of the frictional properties of fluorinated monolayers with atomic force microscopy: Comparison of CF_3- and CH_3- terminated films, Langmuir **13**, 7192–7196 (1997)
39.53 E.W.V.D. Vegte, A. Subbotin, G. Hadziioannou, P.R. Ashton, J.A. Preece: Nanotribological properties of unsymmetrical *n*-dialkyl sulfide monolayers on gold: Effect of chain length on adhesion, friction, and imaging, Langmuir **16**, 3249–3256 (2000)
39.54 E.D. Reedy: Thin-coating contact mechanics with adhesion, J. Mater. Res. **21**, 2660–2668 (2006)

39.55 E.D. Reedy: Contact mechanics for coated spheres that includes the transition from weak to strong adhesion, J. Mater. Res. **22**, 2617–2622 (2007)

39.56 W.R. Ashurst, C. Yau, C. Carraro, R.M.R. Maboudian, M.T. Dugger: Dichlorodimethylsilane as an anti-stiction monolayer for MEMS: A comparison to the octadecyltrichlosilane self-assembled monolayer, J. Microelectromech. Syst. **10**, 41 (2001)

39.57 A.D. Corwin, M.P. de Boer: Effect of adhesion on dynamic and static friction in surface micromachining, Appl. Phys. Lett. **84**, 2451 (2004)

39.58 A.D. Corwin, M.D. Street, R.W. Carpick, W.R. Ashurst, M.P. de Boer: Pre-sliding tangential deflections can govern the friction of MEMS devices. In: *Proc. ASME/STLE Jt. Int. Tribol. Conf., Long Beach* (2004)

39.59 E.E. Flater, A.D. Corwin, M.P. de Boer, M.J. Shaw, R.W. Carpick: In-situ wear studies of surface micromachined interfaces subject to controlled loading, Wear **260**, 580–593 (2006)

39.60 C. Carraro, O.W. Yauw, M.M. Sung, R. Maboudian: Observation of three growth mechanisms in self-assembled monolayers, J. Phys. Chem. B **102**, 4441 (1998)

39.61 E.E. Flater, W.R. Ashurst, R.W. Carpick: Nanotribology of octadecyltrichlorosilane monolayers and silicon: Self-mated versus unmated interfaces and local packing density effects, Langmuir **23**, 9242–9252 (2007)

39.62 B.G. Bush, F.W. DelRio, J. Opatkiewiez, R. Maboudian, C. Carraro: Effect of formation temperature and roughness on surface potential of octadecyltrichlorosilane self-assembled monolayer on silicon surfaces, J. Phys. Chem. A **111**, 12339–12343 (2007)

39.63 E.E. Flater, W.R. Ashurst, R.W. Carpick: Nanotribology of octadecyltrichlorosilane monolayers and silicon: Self-mated versus unmated interfaces and local packing density effects, Langmuir **23**, 9242–9252 (2007)

39.64 J.H. Dieterich: Time-dependent friction in rocks, J. Geophys. Res. **77**, 3690 (1972)

39.65 J.H. Dieterich, B. Kilgore: Implications of fault constitutive properties for earthquake prediction, Proc. Natl. Acad. Sci. USA **93**, 3787 (1996)

39.66 F. Heslot, T. Baumberger, B. Perrin, B. Caroli, C. Caroli: Creep, stick-slip and dry-friction dynamics: Experiments and a heuristic model, Phys. Rev. E **49**, 4973 (1994)

39.67 J.R. Rice, A.L. Ruina: Stability of frictional slipping, J. Appl. Mech. **50**, 343–349 (1983)

39.68 A.L. Ruina: Slip instability and state variable friction laws, J. Geophys. Res. **88**, 359 (1983)

39.69 P. Berthoud, T. Baumberger, C. G'Sell, J.-M. Hiver: Physical analysis of the state- and rate-dependent friction law: Static friction, Phys. Rev. B **59**, 14313 (1999)

39.70 T. Baumberger, C. Caroli, B. Perrin, O. Ronsin: Nonlinear analysis of the stick-slip bifurcation in the creep-controlled regime of dry friction, Phys. Rev. E **51**, 4005–4010 (1995)

39.71 Y.F. Lim, K. Chen: Dynamics of dry friction: A numerical investigation, Phys. Rev. E **58**, 5637–5642 (1998)

39.72 J.T. Desaguliers: *A Course of Experimental Philosophy*, Vol. 1 (London 1734)

39.73 J.H. Dieterich: Time-dependent friction and the mechanics of stick-slip, Pure Appl. Geophys. **116**, 790–806 (1978)

39.74 J. Rice, A. Ruina: Stability of steady frictional slipping, J. Appl. Mech. **50**, 343–349 (1983)

39.75 S.S. Shroff, M.P. de Boer: Constant velocity high force microactuator for stick-slip testing of micromachined interfaces, J. Microelectromech. Syst. **24**, 1868–1877 (2015)

39.76 E.T. Enikov, S.S. Kedar, K.V. Lazarov: Analytical model for analysis and design of V-shaped thermal microactuators, J. Microelectromech. Syst. **14**, 788–798 (2005)

39.77 R. Hickey, D. Sameoto, T. Hubbard, M. Kujath: Time and frequency response of two-arm micromachined thermal actuators, J. Micromech. Microeng. **13**, 40 (2003)

39.78 L. Que, J.S. Park, Y.B. Gianchandani: Bent-beam electrothermal actuators-Part I: Single beam and cascaded devices, J. Microelectromech. Syst. **10**, 247 (2001)

39.79 D.A. Bristow, M. Tharayil, A.G. Alleyne: A survey of iterative learning control, IEEE Control Syst. **26**, 96–114 (2006)

39.80 S.S. Shroff, M.P. de Boer: Full assessment of micromachine friction within the rate-state framework: Experiments, Tribol. Lett. (2016) doi:10.1007/511249-016-0718-3

39.81 M.E. Saleh, J.L. Beuth, Y.N. Picard, M.P. de Boer: In situ platform for isothermal testing of thin-film mechanical properties using thermal actuators, J. Microelectromech. Syst. **24**, 2008–2018 (2015)

39.82 M.G. Hankins, P.J. Resnick, P.J. Clews, T.M. Mayer, D.R. Wheeler, D.M. Tanner, R.A. Plass: Vapor deposition of amino-functionalized self-assembled monolayers on MEMS, Proc. SPIE **4980**, 238–247 (2003)

39.83 A.D. Corwin, M.P. de Boer: Frictional aging and sliding bifurcation in monolayer-coated micromachines, J. Microelectromech. Syst. **18**, 250–262 (2009)

39.84 J.S. Courtney-Pratt, E. Eisner: The effect of tangential force on the contact of metallic bodies, Proc. R. Soc. A **238**, 529 (1957)

39.85 F.W. DelRio, M.P. de Boer, J.A. Knapp, E.D. Reedy, P.J. Clews, M.L. Dunn: The role of van der Waals forces in adhesion of micromachined surfaces, Nat. Mater. **4**, 629–634 (2005)

39.86 S.S. Shroff, M.P. de Boer: Direct observation of the velocity contribution to friction in monolayer-coated micromachines, Extreme Mech. Lett. **8**, 184–190 (2016)

39.87 T. Hatano: Rate and state friction law as derived from atomistic processes at asperities, arXiv preprint arXiv:1512.05078 (2015)

39.88 S.S. Shroff, M.P. de Boer: Full assessment of micromachine friction within the rate-state frame-

work: Theory and validation, Tribol. Lett. (2016) doi:10.1007/s11249-016-0724-5
39.89 N. Lapusta, J.R. Rice: Nucleation and early seismic propagation of small and large events in a crustal earthquake model, J. Geophys. Res.: Solid Earth (2003) doi:10.1029/2001JB000793
39.90 A. Epstein, S.D. Senturia: Macro power from micro machinery, Science **276**, 1211 (1997)
39.91 L.G. Fréchette, S.A. Jacobson, K.S. Breuer, F.F. Ehrich, R. Ghodssi, R. Khanna, C.W. Wong, X. Zhang, M.A. Schmidt, A.H. Epstein: High-speed microfabricated silicon turbomachinery and fluid film bearings, J. Microelectromech. Syst. **14**, 141–152 (2005)
39.92 N. Ghalichechian, A. Modafe, M.I. Beyaz, R. Ghodssi: Design, fabrication, and characterization of a rotary micromotor supported on microball bearings, J. Microelectromech. Syst. **17**, 632–642 (2008)
39.93 J. Winters: Rolling with it: Researchers are reducing friction in MEMS devices using a very old solution in a new size, Mech. Eng.-CIME **131**, 22–27 (2009)
39.94 M. McCarthy, C.M. Waits, R. Ghodssi: Dynamic friction and wear in a planar-contact encapsulated microball bearing using an integrated microturbine, J. Microelectromech. Syst. **18**, 263–273 (2009)
39.95 B. Hanrahan, S. Misra, C.M. Waits, R. Ghodssi: Wear mechanisms in microfabricated ball bearing systems, Wear **326**, 1–9 (2015)
39.96 C. Chisholm, H. Bei, M. Lowry, J. Oh, S.S. Asif, O. Warren, Z. Shan, E.P. George, A.M. Minor: Dislocation starvation and exhaustion hardening in Mo alloy nanofibers, Acta Mater. **60**, 2258–2264 (2012)
39.97 H. Guo, K. Chen, Y. Oh, K. Wang, C. Dejoie, S.S. Asif, O. Warren, Z. Shan, J. Wu, A. Minor: Mechanics and dynamics of the strain-induced M1–M2 structural phase transition in individual VO_2 nanowires, Nano Lett. **11**, 3207–3213 (2011)
39.98 J. Liu, J.K. Notbohm, R.W. Carpick, K.T. Turner: Method for characterizing nanoscale wear of atomic force microscope tips, ACS Nano **4**, 3763–3772 (2010)
39.99 T.D.B. Jacobs, B. Gotsmann, M.A. Lantz, R.W. Carpick: On the application of transition state theory to atomic-scale wear, Tribol. Lett. **39**, 257–271 (2010)
39.100 A. Desai, M. Haque: A novel MEMS nano-tribometer for dynamic testing in-situ in SEM and TEM, Tribol. Lett. **18**, 13–19 (2005)
39.101 T. Ishida, Y. Nakajima, K. Kakushima, M. Mita, H. Toshiyoshi, H. Fujita: Design and fabrication of MEMS-controlled probes for studying the nano-interface under in situ TEM observation, J. Micromech. Microeng. **20**, 075011 (2010)
39.102 T. Sato, T. Ishida, L. Jalabert, H. Fujita: Development of MEMS-in-TEM setup to observe shear deformation for the study of nano-scale friction, Tribol. Online **6**, 226–229 (2011)
39.103 J.J. Yao: RF MEMS from a device perspective, J. Micromech. Microeng. **10**, R9 (2000)
39.104 C. Keimel, G. Claydon, B. Li, J.N. Park, M.E. Valdes: Microelectromechanical systems-based switches for power applications, IEEE Trans. Ind. Appl. **48**, 1163–1169 (2012)
39.105 T.-J.K. Liu, D. Markovic, V. Stojanovic, E. Alon: The relay reborn, IEEE Spectrum **49**(31), 39–43 (2012)
39.106 M. Spencer, F. Chen, C.C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, H. Kam, V. Pott, J. Jeon, T.J.K. Liu, D. Markovic, E. Alon, V. Stojanovic: Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications, IEEE J. Solid – State Circuits **46**, 308–320 (2011)
39.107 S. Narendra, V. De, S. Borkar, D.A. Antoniadis, A.P. Chandraskasan: Full-chip subthreshold leakage power prediction and reduction techniques for sub-0.18 μm CMOS, IEEE J. Solid – State Circuits **39**, 501–510 (2004)
39.108 D.A. Czaplewski, G.A. Patrizi, G.M. Kraus, J.R. Wendt, C.D. Nordquist, S.L. Wolfley, M.S. Baker, M.P. de Boer: A nanomechanical switch for integration with CMOS logic, J. Micromech. Microeng. **19**, 085003 (2009)
39.109 U. Zaghloul, G. Piazza: Sub-1-volt piezoelectric nanoelectromechanical relays with millivolt switching capability, IEEE Electron Device Lett. **35**, 669–671 (2014)
39.110 G.M. Rebeiz, J.B. Muldavin: RF MEMS switches and switch circuits, IEEE Microw. Mag. **35**, 59–71 (2001)
39.111 J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami: Internet of things (IoT): A vision, architectural elements, and future directions, Future Gener. Comput. Syst. **29**, 1645–1660 (2013)
39.112 V. Brand, M.S. Baker, M.P. de Boer: Impact of contact materials and operating conditions on stability of micromechanical switches, Tribol. Lett. **51**, 341–356 (2013)
39.113 R.A. Coutu Jr., P.E. Kladitis, K.D. Leedy, R.L. Crane: Selecting metal alloy contact materials for MEMS switches, J. Micromech. Microeng. **14**, 1157–1164 (2004)
39.114 H.W. Hermance, T.F. Egan: Organic deposits on precious metal contacts, Bell Syst. Tech. J. **37**, 739–776 (1958)
39.115 H.P. Koidl, W.F. Rieder, Q.R. Salzmann: Parameters influencing the contact compatibility of organic vapours in telecommunication and control switches. In: *Proc. 44th IEEE Holm Conf. Electr. Contacts* (1998) pp. 220–225
39.116 V. Brand, M.S. Baker, M.P. de Boer: Contamination thresholds of Pt- and RuO_2-coated ohmic switches, J. Microelectromech. Syst. **22**, 1248–1250 (2013)
39.117 K. Hinohara: Reed switches. In: *Electrical Contacts: Principles and Applications*, ed. by P.G. Slade (CRC, Boca Raton 1999) pp. 535–572
39.118 V. Brand, M.P. de Boer: Oxygen-induced graphitization of amorphous carbon deposit on ohmic switch contacts improves their electrical conductivity and protects them from wear, J. Micromech. Microeng. **24**, 095029 (2014)

40. Failure Mechanisms in MEMS/NEMS Devices

W. Merlijn van Spengen, Robert Modliński, Robert Puers, Anne Jourdain

The commercialization of MEMS/NEMS devices is proceeding slower than expected, because the reliability problems of microscopic components differ from macroscopically known behavior. In this chapter, we provide an overview of the state of the art in MEMS/NEMS reliability. We discuss the specific, MEMS-related problems caused by stiction due to surface forces and electric charge. Materials issues such as creep and fatigue are treated as well. Nanoscale wear is covered briefly. MEMS packaging is also discussed, because the reliability of MEMS/NEMS components critically depends on the available protection from the environment.

40.1 **Failure Modes and Failure Mechanisms** 1438
40.2 **Stiction and Charge-Related Failure Mechanisms** 1439
40.2.1 Stiction Due to Surface Forces 1439
40.2.2 Stiction Due to Electrostatic Attraction . 1442
40.3 **Creep, Fatigue, Wear, and Packaging-Related Failures** 1445
40.3.1 Creep 1445
40.3.2 Fatigue 1448
40.3.3 Wear 1450
40.3.4 Packaging 1451
40.4 **Conclusions** 1454
References 1454

With the rapid spread of micro-electromechanical systems (MEMSs) technology in the 1990s, it has become clear that the reliability of microscopic mechanical devices is a large concern [40.1]. It is one of the main reasons that much of the micromechanical technology available all over the world in laboratories has not been commercialized yet. The route from a working prototype to a commercial MEMS component should not be underestimated: the pioneering effort of Texas Instruments to bring the digital micromirror device (DMD) MEMS chip to the market took more than 10 years after its invention [40.2]. The same holds for gyroscopes and accelerometers: many years of hard work were spent on the conversion of laboratory prototypes to commercial products [40.3].

The failure mechanisms of MEMS/NEMS (in the rest of this chapter we will use MEMS as a generic name that also includes devices with nanometer-scale features, which are increasingly being referred to as nanoelectromechanical systems (NEMSs)) components are widely differing from those of their macroscopic counterparts, and this is the main reason that this trajectory takes so long. Many MEMS designers still do not take into account the implications that the scale difference has on reliability when designing new devices, although the paradigmatic approach of *design for reliability*, i. e., taking into account reliability issues from the earliest design stage, is gaining more acceptance. A second point is that, even if people are willing to take into account all the available information on the failure modes and mechanisms of MEMS, they soon discover that not so much is known at all. Still, over the last few years, the information available on MEMS failure mechanisms has been steadily growing [40.4]. At present, we are not yet in a position to predict quantitatively when or where most of the different failure mechanisms will occur, but we can describe them in some detail and give rules of thumb to avoid many of them.

In this chapter we will introduce the basic MEMS failure mechanisms and point out when they are important and in which respects they differ from macroscopic failure mechanisms. We will study the most important of them in some detail and give recommendations for procedures that are known to counteract them.

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_40

40.1 Failure Modes and Failure Mechanisms

We have to start with a clear distinction between failure modes and failure mechanisms. The failure mode is the apparent failure. It might be that, e.g., a MEMS beam is stuck to the substrate. If this causes a device to malfunction, the failure mode of the device is a stuck beam. This failure mode may be caused by different failure mechanisms, such as stiction of the beam (stiction is a generic term for surfaces being stuck due to surface forces), it may be a fatigue fracture failure or caused by an overload condition, and so on. The failure mechanism is hence the physical cause of the failure mode. In Fig. 40.1 the failure mode is broken beams, but the failure mechanism is rough handling, causing a mechanical overload condition.

This distinction is important because often the failure mode is reported when something is wrong, while the problem can only be accurately addressed by investigating the failure mechanism. In Table 40.1, the most commonly encountered failure mechanisms are given (most of them taken from [40.5]).

Dedicated equipment to assess failure mechanisms is definitely needed, but only a few set ups have been reported to date. Sandia National Laboratories is famous for their Sandia high-volume measurement of micromachine reliability (SHiMMeR) system, where they can test a significant number of rotating micromachines simultaneously under various environmental conditions [40.6]. The operation of the micromotors is monitored with advanced software tools [40.7]. Inter-University Microelectronics Center (IMEC) has a dedicated system to assess radio-frequency (RF) MEMS switch reliability, where environmental conditions can be altered with in situ optical access and various stimuli can be applied [40.8]. Interferometry- and vibrometry-based optical systems of various kinds are in use to monitor the dynamic behavior of MEMS devices [40.9–12] (Fig. 40.2). Some commercial laboratories also have their dedicated systems but do not disclose their exact set ups.

Other equipment that is used for failure mechanism assessment has not been developed exclusively for MEMS, but the MEMS reliability community is relying heavily on them: scanning electron microscopy (SEM), focused ion beam (FIB) cross-sectioning, atomic force microscopy (AFM), optical microscopy, and various electronic actuation and detection systems are the most widespread among them.

Table 40.1 Common MEMS failure mechanisms

Fracture	Overload fracture Fatigue fracture
Creep	Applied stress Intrinsic stress Thermal stress
Stiction	Capillary forces Van der Waals molecular forces Electrostatic forces Solid bridging
Electromigration	
Wear	Adhesive Abrasive Corrosive
Degradation of dielectrics	Leakage Charging Breakdown
Delamination	
Contamination	
Pitting of contacting surfaces	
Electrostatic discharge (ESD)	

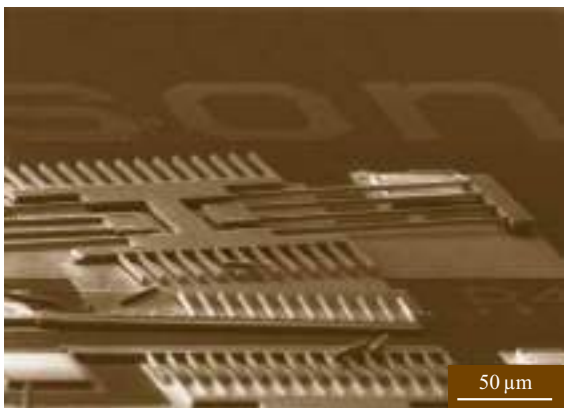


Fig. 40.1 The failure mode of this resonator is broken beams; the failure mechanism is rough handling, causing an overload stress in the beams and support

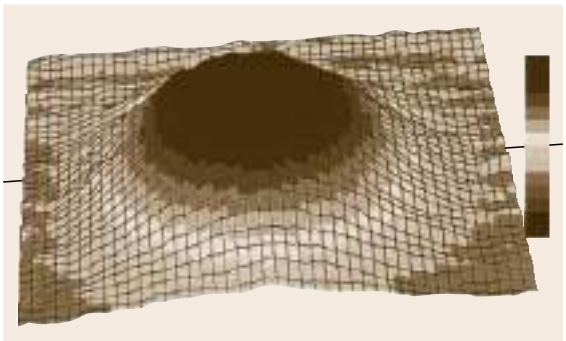


Fig. 40.2 Example of optical assessment of dynamic MEMS behavior: laser television (TV) holography measurement image of the first resonance mode of a MEMS pressure sensor membrane (after [40.8])

40.2 Stiction and Charge-Related Failure Mechanisms

40.2.1 Stiction Due to Surface Forces

One of the most important problems in MEMS/NEMS devices is stiction, i. e., the unintentional sticking of contacting surfaces (Fig. 40.3). Because forces intentionally generated in MEMS by, e.g., electrostatic actuation are generally very small, surface forces can dominate the device behavior. The fact that MEMS surfaces are so smooth that a large part of the surfaces is within the range of the atomic forces of the other surface aggravates the matter further. The most important surfaces forces commonly listed are: forces due to capillary condensation, van der Waals molecular forces, and chemical and hydrogen bonds between the surfaces. Surfaces tend to stick together when they are dried after the release etch, and a good overview of what happens is given by *Mastrangelo* and *Hsu* [40.13, 14]. From a reliability point of view, in-use stiction is more important.

The capillary condensation, van der Waals forces, and bonding force effects mentioned have in common that they are true interface forces. The force depends in this case on the situation at the interface, and not on effects far away from it. All these forces have in common that they are strong for surfaces very close together but fall off relatively fast when the distance is larger than a few nanometers.

As a result, surface roughness is an important parameter governing the stiction force, because the surface roughness of MEMS devices is commonly found to be of the same order of magnitude as the range of the surface forces (Fig. 40.4). This was demonstrated for the first time by *De Boer* et al. [40.15]. They monitored one-side-clamped free-standing cantilever MEMS structures that were stuck over a considerable part of their length, using an interferometer (Fig. 40.5). Fringes in the image indicated which part of the beams was free-standing and which part was stuck. Because the restoring spring force of the beams was known, the length over which the beams were not stuck gave a measure for the surface interaction energy (Fig. 40.6 and Table 40.2).

We have to distinguish for this calculation between beams that adhere only at the tip (*arc-shaped*) and beams that adhere over a considerable length (*S-shaped*). It was argued by De Boer and coworkers that the S-shaped beam gives a more precise measure for the surface interaction energy because the arc-shaped beam introduces a large statistical error; only the properties of the surface at the very small contacting tip produce the measured result. Arc-shaped beams can very easily be used to find the surface interaction energy to first order however, because we only have to look at which beams are stuck and which beams are free-standing.

De Boer et al. [40.15] found that the surface interaction energy varied considerably with the relative humidity and attributed this effect to capillary condensation. Small liquid bridges are supposed to form when two hydrophilic surfaces are in close contact (Fig. 40.7).

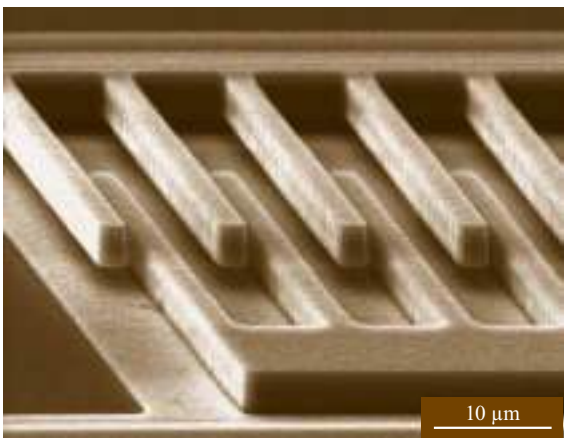


Fig. 40.3 Stiction failure of a comb drive

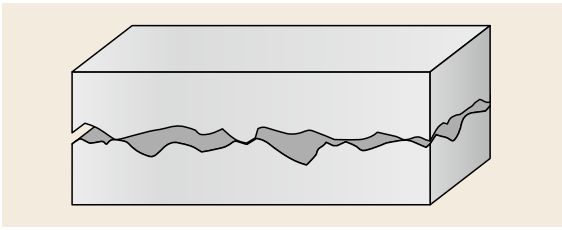


Fig. 40.4 Two rough surfaces in contact

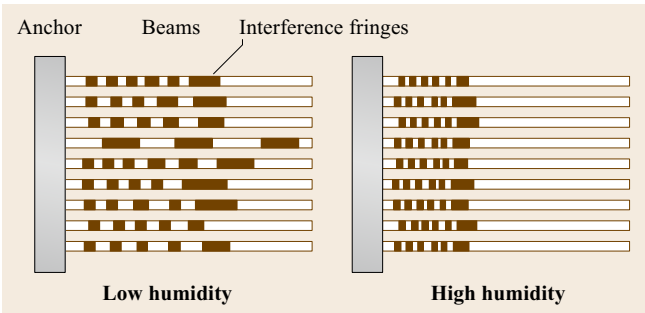


Fig. 40.5 Schematic top view of the measurement of adhesion as a function of relative humidity using interference imaging by *De Boer* et al. [40.15]. The *dark parts* are fringes created with an interferometric objective. The fringe pattern reveals where the beams are bending, and where they are flat, because fringes are only observed where the beams are not running perfectly in plane

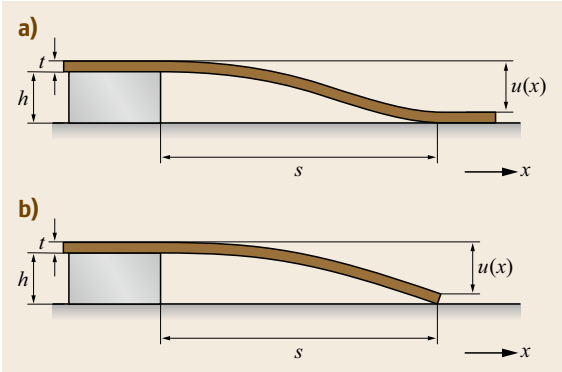


Fig. 40.6 (a) S-shaped beam; (b) arc-shaped beam

This effect had already been reported by *Bowden* and *Tabor* in 1950 [40.16] with a macroscopic measuring apparatus, but only with the development of MEMS technology has capillary condensation become a big technological issue. Because only the surface area that is closer to the other surface than a characteristic distance will be bridged by capillary condensed water, generally only part of the surface area is in contact in this sense.

Maboudian and *Howe* [40.17], *Legtenberg* et al. [40.18] and *Tas* et al. [40.19] provided the first quantitative models of stiction, making use the work of *Israelachvilli* [40.20] on more generic systems. They described the surface forces as a function of the distance between two flat surfaces. The relation between surface interaction energy and distance for capillary forces (a function of humidity) and of van der Waals molecular forces (a function of the Hamaker constant of the molecules at the surface) was presented with very general applicability.

The roughness of the surfaces (some parts are closer together and contribute more to the stiction force than others) was discussed briefly by *Komvopoulos* et al. [40.21] and more quantitatively by *van Spengen* et al. [40.22]. The procedure boils down to the determination of a distance distribution function $h(z)$, giving the amount of surface at every distance z from the other surface (Fig. 40.8). The total surface interaction energy Γ_i due to a certain i-th force $e_i(z)$, e.g., the van der Waals force, is then given by

$$\Gamma_i = \int_0^\infty e_i(z)h(z)\mathrm{d}z\,. \tag{40.1}$$

This principle is generally applicable. As long as the forces do not influence each other, they will be reasonably well described by (40.1) and can be added together.

It is not easy to predict the distance distribution function $h(z)$, as the two surfaces may deform each other at the contacting points, giving rise to both plastic and elastic deformation. If we know the height distribution of the contacting surfaces independently, a rough estimate can be obtained by assuming only plastic deformation. This gives rise to the *bearing area* representation [40.22], from which the distance distribution can be obtained. Unfortunately, the surface interaction energy value obtained in this way is expected to be too low, as elastic deformation is expected to bring larger parts of the surfaces closer together than in the fully plastic case. Purely elastic deformation can be treated by the Greenwood–Williamson model [40.23], but has been found to be useful from an experimental point of view only for very soft, elastic materials like rubber. The technologically more interesting cases (silicon on silicon, metals etc.) are best described with a mixed plastic/elastic model, but unfortunately only numerical simulations can be employed to describe this [40.24, 25].

A more experimental approach to characterize the rough contact between two MEMS surfaces is followed by *Kogut* et al. [40.26], who use electrical current mea-

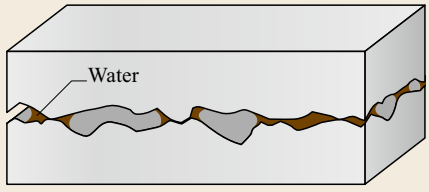


Fig. 40.7 Capillary condensation occurs on all places where the two contacting surfaces are closer together than a characteristic distance

Table 40.2 Deflection and surface interaction energy of S- and arc-shaped beams (after [40.15])

	S-shaped beam	Arc-shaped beam
Deflection $u(x)$	$u(x) = h\left(\frac{x}{s}\right)^2\left(3-2\frac{x}{s}\right)$	$u(x) = h\left(\frac{x}{s}\right)^2\left(3-\frac{x}{s}\right)$
Surface interaction energy Γ [a]	$\Gamma = \frac{3}{2}E\frac{t^3h^2}{s^4}$	$\Gamma = \frac{3}{8}E\frac{t^3h^2}{s^4}$

[a] E = Young's modulus of the beam material, the other parameters are given in Fig. 40.6

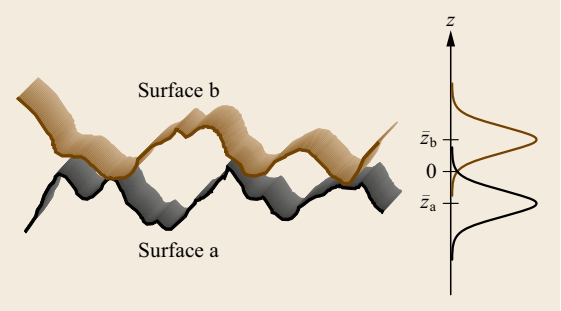


Fig. 40.8 The two surface-height distribution functions give rise to a distance distribution function $h(z)$ describing how large the part of the surface is that has an interaction with the other surface at a certain distance

surements through the contacting points to estimate the real contact area.

Whether the resulting surface interaction energies from the different force contributions result in a stiction failure or not depends on the restoring force. The peel number N_p (first coined by *Mastrangelo* and *Hsu* [40.14] and very well explained by *Zhao* [40.27]) can be used to find out whether a structure will suffer from stiction. For $N_p < 1$ the surfaces will adhere, while for $N_p > 1$ the surfaces will separate again if they are brought into contact.

The most cumbersome stiction force is that caused by capillary condensation. Water will only capillary condense onto hydrophilic surfaces, so a hydrophobic coating can be applied to prevent this. These layers are not unlike the boundary lubricants that have been used for decades to lower friction. Different ways of creating hydrophobic coatings have been applied but there is no industry standard yet. Antistiction coatings are generally made by either plasma deposition or self-assembled monolayer (SAM) deposition. The standard film quality test method is to bring a small water droplet to the surface and monitor the contact angle.

The gas used for the plasma deposition is usually a low-molecular-weight fluorinated organic compound, such as CHF_3 [40.28] or C_4F_{10} [40.29]. The MEMS device is bombarded with all kinds of free radicals and other reactive species generated from these molecules and forms a well-adhered water-repelling coating on the surfaces of the device. A big problem with plasma deposition is obtaining conformal coatings. Tricks such as having a Faraday cage inside the chamber are advantageous [40.29], but even then the conformity is usually not better than 2 : 1 for the top and bottom surfaces of a MEMS device.

The durability of plasma-deposited coatings is relatively good, although they tend to come off in boiling water. *Man* et al. [40.29] tested their films for water contact-angle changes at elevated temperatures. By determining the lifetime at different temperatures and using an Arrhenius relation for the mean time to failure (MTTF)

$$\mathrm{MTTF} \propto \exp\left(\frac{E_a}{k_B T}\right) , \tag{40.2}$$

they calculated a lifetime of more than 10 years at 150 °C. Knowing the Boltzmann constant k_B, one can obtain the activation energy E_a by performing experiments at different temperatures T. Impact wear was also tested.

To overcome the conformity problems associated with plasma deposition, and also because very thin layers are possible in this way, SAM coatings have been investigated since the middle of the 1990s. *Srinivasan* et al. [40.30] reported extensively on the subject. Silicon microstructures are released in HF, and while still in the liquid phase, the HF solution is replaced by a trichlorosilane solution. The most common are octadecyltrichlorosilane (OTS, $CH_3(CH_2)_{17}SiCl_3$) and 1*H*,1*H*,2*H*,2*H*-perfluorodecyltrichlorosilane (FDTS, $CF_3(CF_2)_7(CH_2)\ SiCl_3$). Figure 40.9 shows how such a film is formed on the surface of a MEMS device. A processing problem is that the reaction depends on trace amounts of water in the reaction environment. If the processing conditions are not extremely well controlled, micelles are formed as well as highly interconnected surface films that are easily peeled off [40.31].

Thermal stability tests show that, in air, the fluorocarbon films are able to stand 400 °C while the hydrocarbon films will degrade above 100 °C. In nitrogen no difference was observed [40.31]. *Kim* et al. [40.32] have suggested that the use of dimethyldichlorosilane (DDMS, $(CH_3)_2SiCl_2$) results in a process that is less sensitive to temperature, humidity and coating time. The films thus created withstood temperatures of above 450 °C without degradation.

De Boer et al. [40.15] found that FDTS films were susceptible to *water island formation* in a 90% rela-

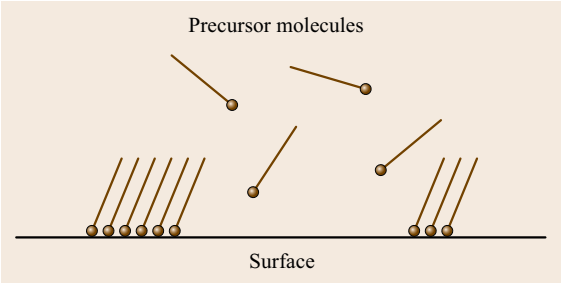


Fig. 40.9 SAM formation on a MEMS surface

tive humidity environment after only a few hours, while OTS films did not show such a dramatic effect. Whether this was a real difference or merely an effect of a difference in surface film quality has not been elucidated.

Another method besides antistiction coatings to reduce stiction problems is roughening the surfaces so that only small parts of the surfaces are in intimate contact. *Maboudian* et al. [40.17] show how this can be accomplished with different etchants (this also decreases van der Waals contributions). An absolutely water-free environment will also prevent capillary condensation problems altogether. The hermeticity of packages to accomplish this is discussed in the packaging section of this chapter (Sect. 40.3.4). The best solution to stiction due to surface interaction energies is the use of very stiff structures and high forces, which will separate the surfaces even under high adhesive contact forces. A typical higher limit for the surface interaction energy due to capillary forces is $0.1\,\mathrm{J/m^2}$ [40.17], which corresponds to water everywhere between the surfaces. Surface interaction energies in rough devices and with antistiction coatings are typically 10–1000 times lower. If these numbers are taken into account, finite element modeling of the structures before they are produced can be used to predict whether stiction due to surface forces will be a reliability threat or not.

40.2.2 Stiction Due to Electrostatic Attraction

Stiction due to dielectric charging is different from the previously discussed forces in the sense that it is not a true interface force. Charges may be present at the interface but may equally well reside in the bulk of a dielectric material. Electrostatic forces are very important on the microscale and hence an in-depth investigation of charge effects is important.

Charges may be generated in microsystems in different ways. It has been found that in devices with rubbing surfaces tribocharging may occur [40.33]. A rotating silicon micromotor was tested until it remained stuck due to the accumulated charge. Irradiation with a focused ion beam (FIB) neutralized the charges, after which the device resumed operation. Another situation where charge may be generated is where sparks in a small gap spray charge onto a surface [40.34]. Contact potentials of materials are usually $< 0.5\,\mathrm{V}$ and are not expected to give any trouble in normal applications.

Charges generated by irradiation with ionizing radiation constitute a considerable challenge in MEMS devices. Accelerometers subjected to nuclear radiation were found to drift with increased dose because the dielectric under the moving mass of such a device was charged [40.35, 36]. Accelerometers without a dielectric under the moving mass were found to be much less susceptible, because no charge can accumulate (Fig. 40.10). One can imagine that excessive radiation would cause the moving mass to snap in and constitute a stiction failure, but this has not yet been reported.

The largest problems with stiction due to electrostatic attraction are found in devices where a large voltage is applied across a thin dielectric, such as in RF MEMS switches [40.37]. The high field causes charge injection into the dielectric by Schottky- or Poole–Frenkel-type conduction and is experimentally found to depend exponentially on the actuation voltage [40.38]. The situation resembles that of the high-k-dielectric gate oxides of the newest generations of metal–oxide–semiconductor (MOS) transistors, where similar charging effects threaten the stability of the threshold voltage [40.39].

The exact charging properties of a dielectric thin film depend on the deposition conditions. A *leaky* dielectric with a low resistivity may have many trapping sites, but can still be good from a technological point of view if the charge can flow away easily [40.40]. The films have to be empirically optimized because no consistent theory to describe good films has been developed yet.

Much more is known about how parasitic charges, when they are present, influence device behavior. For a free-sanding structure with small-scale deflections, *Wibbeler* et al. [40.34] have presented a theory that describes the deflection as a function of the accumulated charge. *Reid* [40.41] has shown finite element model-

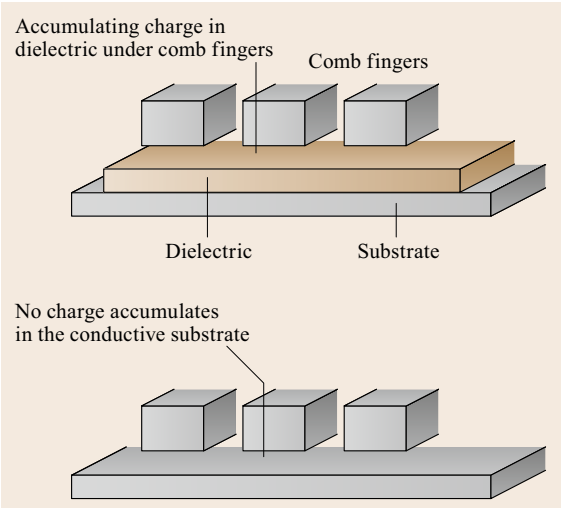


Fig. 40.10 Ionizing radiation can cause charging of the dielectric under the cantilevers (that are part of the moving mass) of an accelerometer and hence cause drift and in extreme cases stiction of the moving part

ing (FEM) simulations for large deflections, pull-in and pull-out effects and the effect of parasitic charge on them.

The most comprehensive efforts, which are very useful from the point of view of failure due to stiction, are the treatises of *van Spengen* et al. [40.42] for uniform charge densities in the dielectric and *Rottenberg* et al. [40.43] for inhomogeneous charge densities. These two papers consider RF MEMS switches, but the principles are generic.

In these models, elementary electrostatics is used to describe the peculiar dynamic effects in MEMS with a structure like that in Fig. 40.11. A free-standing structure (the *bridge*) is located above a dielectric with a lower conductor. The whole assembly behaves as a capacitor. When a voltage is applied between the bridge conductor and the lower conductor, the bridge is attracted. If the bridge is moved over more than one third of the gap, the system becomes unstable and the bridge lands on the dielectric (pull-in). At this point the capacitance of the system becomes much higher and this is the reason that it can be used to affect RF electronic signals. A much lower voltage is required to keep it there, so we have to lower the actuation voltage considerably to let it *pull out* again. This hysteresis is well known from RF MEMS *C–V* curves (Fig. 40.12).

Both positive and negative actuation voltages can cause the bridge to pull in (the *C–V* curve is symmetric for positive and negative voltages). For a uniform parasitic charge, the curve starts to shift. Negative charge causes the curve to shift to the right, while positive charge shifts the whole curve to the left on the voltage axis.

The effect of positive charge is easily understood: when the curve shifts so much to the left that the pull-out voltage crosses the 0 V line, the bridge will remain stuck even if the actuation voltage is removed. A negative charge results in a more tricky behavior. For small charge values, the pull-in and pull-out voltages will shift slightly upward so no stiction is expected. If too much charge accumulates, the pull out of the negative-going actuation will shift through 0 V (Fig. 40.13). For slow, positive-going voltages, this makes no difference. When we increase the voltage on an initially free-standing bridge, it will pull in at the shifted positive pull-in voltage, and return to its upward position when the voltage is again slowly decreased to below the pull-out voltage. Only when we go negative in actuation voltage will the bridge be pulled in again at the shifted negative pull-in voltage.

But when we change the actuation voltage quickly, as is often the case when devices are actuated, e.g., by a square wave, the behavior is completely different. We start in the same condition as before, with the bridge up, and increase the voltage quickly to above the pull-in voltage, the bridge will be pulled in after a characteristic time that is related to the mechanical resonance frequency of the device. When we quickly reduce the voltage to 0 V, the bridge stays behind again and is still down when the voltage is 0 V. However, now we are in a completely different situation. Because the pull-out voltage of the negative branch of the curve is positive,

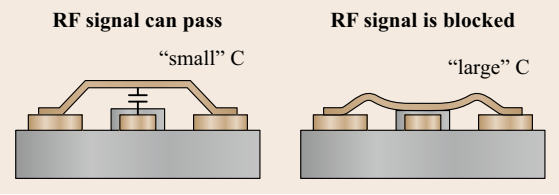


Fig. 40.11 RF MEMS switch cross section

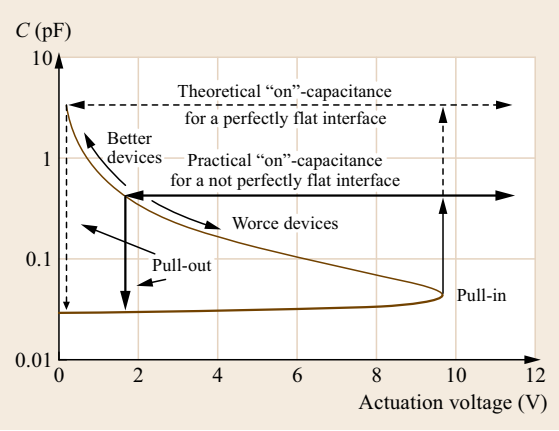


Fig. 40.12 *C–V* curve

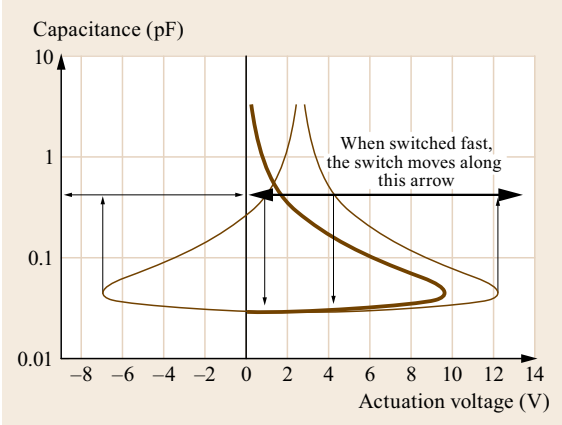


Fig. 40.13 With square-wave actuation, after the negative pull-out voltage has crossed 0 V due to a parasitic charge, the device moves only along the *bold arrow* and remains stuck

the bridge will remain stuck. The bridge does not know whether it came from a negative voltage or from a positive voltage. For a negative voltage it would only pull out above 0 V at the negative-branch pull-out voltage, so it will remain stuck. In this way both positive and negative charges can cause a MEMS device to remain stuck. It is even possible to calculate the approximate magnitude of the charge that is required to cause stiction in either case. The important case in which the negative pull-out voltage crosses zero and causes a device failure has also been verified experimentally [40.42].

Not only charge injection due to high voltages across the dielectric may cause this kind of reliability problems to emerge; ionizing radiation can also play a similar role. It has to be kept in mind however that, if the bridge or other moving structure does not intentionally or accidentally hit the surface below, stiction can only take place when the charge causes a shift that is so large that the much higher pull-in voltage crosses 0 V from either side.

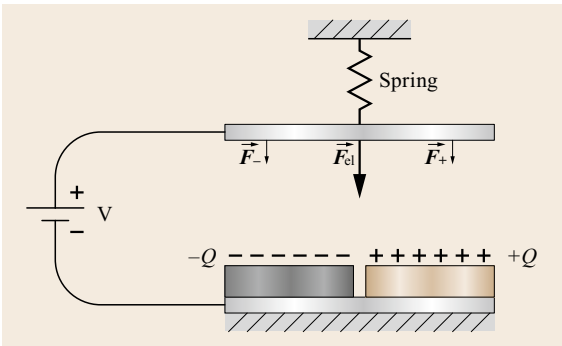


Fig. 40.14 A nonuniform charge can generate attraction of the upper conductor even when the net charge is zero (after [40.43])

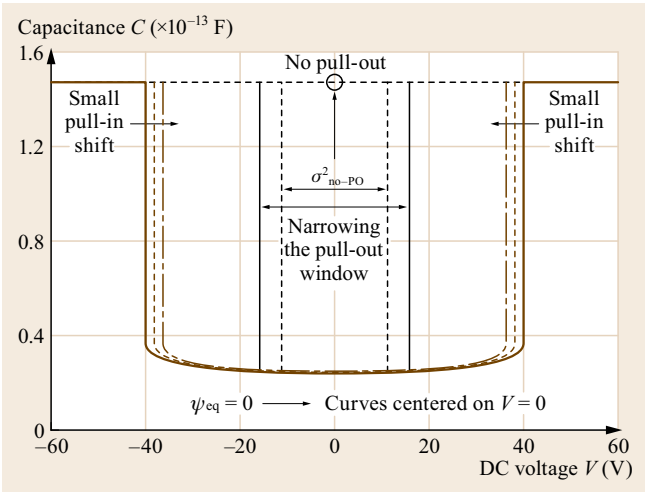


Fig. 40.15 The nonuniform charge causes the pull-out voltage to move to 0 V; in extreme cases the pull-out voltage disappears (after [40.43])

The previous explanation is valid only for uniform charge distributions (either bulk or surface charge). *Rottenberg* et al. [40.43] have extended the theory to deal with nonuniform charge distributions that give rise to effects that have also been experimentally observed. The nonuniform distribution can cause a nonzero force even when the net charge on the dielectric is zero (Fig. 40.14). This causes a narrowing of the C–V curve so that the positive and negative pull-out and pull-in are closer together (Fig. 40.15).

In conclusion, even relatively small charge values of both polarities that are present in or on a dielectric may affect the device reliability when the moving part intentionally or accidentally hits the lower surface. If this is not the case, stiction can only occur when the C–V curve is shifted so much that the pull-in voltage crosses 0 V.

Reduction of the sensitivity for charging phenomena in MEMS devices can be accomplished in different ways. The most obvious way is the reduction of charging by optimizing the dielectrics for low charging. This can be done either by making a very-high-quality dielectric with few trapping sites, or by increasing the trap density but making sure that the average time a charge remains in a trap is very short (the *leaky dielectric*). Another option is to design the dielectrics that are prone to charging out of the device concept. In RF MEMS this can be done with separate actuation pads where no dielectric is located so that no voltage acts across the high-k dielectric. In the accelerometer example of Fig. 40.10, this was done by removing the dielectric altogether. The third option, but probably the least satisfactory because of its added complexity and device cost, is to use advanced actuation schemes to reduce charging. *Goldsmith* et al. [40.38] use an advanced actuation waveform with a high initial voltage to pull the bridge down and a lower *holding* voltage to keep it down, so that charging is reduced. Polarity reversal is also used to sweep out the charges that are accumulating [40.44]. Both of these are useful with actuation signals that are known in advance and do not depend on the particular application in mind.

40.3 Creep, Fatigue, Wear, and Packaging-Related Failures

40.3.1 Creep

Creep is the permanent and irreversible deformation of a ductile material occurring below the yield strength. Its importance depends on temperature, mechanical stress, time, and material composition. It is easy to think of a solid metal as having a well-defined yield strength below which it does not flow and above which flow is rapid. This is unfortunately true only at absolute zero. Above this temperature, plastic flow is a dynamic process, where the strength of solids depends on strain and temperature. The irreversible time-dependent part of a strain ε is called creep.

The strain rate $\partial\varepsilon/\partial t$ (elongation speed) is extremely temperature-sensitive. It can often be approximated by a rate equation

$$\frac{\partial\varepsilon}{\partial t} = A\exp\left(-\frac{q}{k_\mathrm{B}T}\right), \tag{40.3}$$

where A and q are the fit parameters, k_B is Boltzmann's constant and T is the absolute temperature.

Thermal vibrations aid the applied stress in overcoming specific barriers (e.g., obstacles) to induce plastic flow in the metal. Slip, climb, shear on grain boundaries, and vacancy diffusion inside the grain (the so-called Nabarro–Herring effect) are the mechanisms responsible for deformation in metals [40.45]. The last of these is significant only at very high temperatures and under very small stresses. The normal operating temperatures in most practical engineering applications are too low and usually the operating stresses are too high for the Nabarro–Herring mechanism to occur, so we can conclude that creep deformation in metals depends primarily on dislocation movement (slip, climb and grain-boundary shearing).

Three creep stages can be distinguished: primary, secondary (so-called steady-state creep) and tertiary creep, which ends up with fracture (Fig. 40.16). Especially devices made out of low-melting-point metals are sensitive to creep. This is the case because the melting temperature of a material has a major influence on its creep behavior. There exists a threshold temperature, which divides the deformation of a material by creep into two parts: high-temperature and low-temperature creep (Fig. 40.17). In the regime of high temperature, creep is a rather fast deformation and depends strongly on temperature, whereas in the low-temperature regime, the creep rate decreases logarithmically with time. It is commonly accepted that this threshold temperature T_c is given by $T_\mathrm{c} = 0.5T_\mathrm{m}$, where T_m is the absolute melting temperature, although each material must be evaluated separately [40.46]. For instance, T_c is around 200 °C for aluminum and its alloys and 315 °C for titanium alloys.

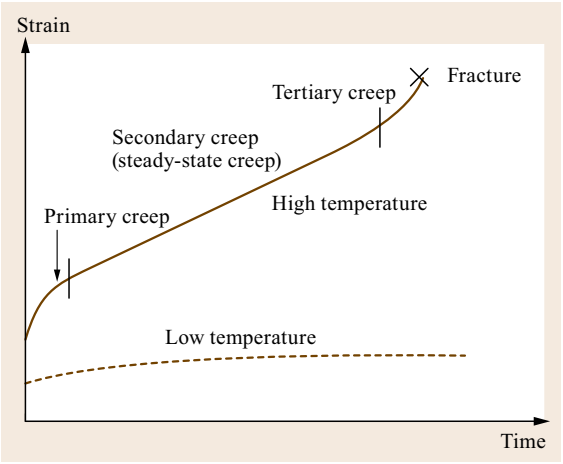


Fig. 40.16 Low-temperature and high-temperature creep under a constant stress (after [40.46])

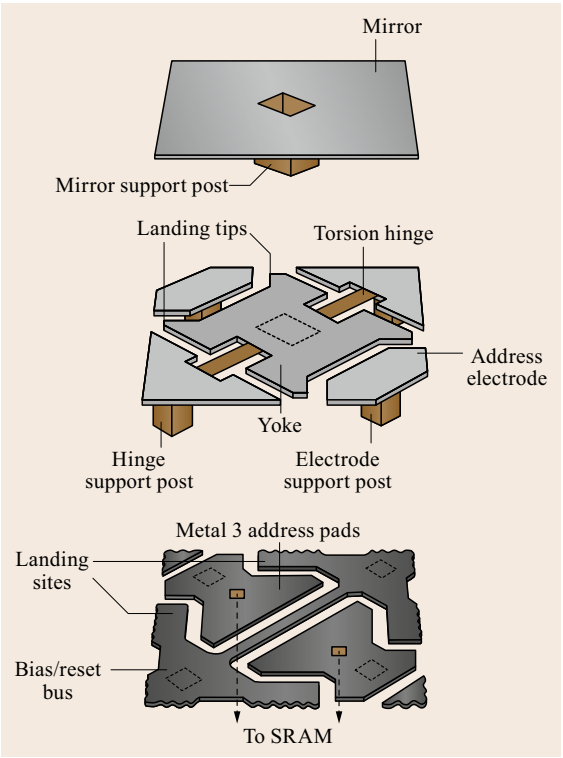


Fig. 40.17 Digital micromirror device for projection displays (© Texas Instruments)

Shape changes caused by creep can be undesirable and represent a limiting factor in the lifetime of a device. For instance, blades on the spinning rotors in turbine engines are endangered by creep. They slowly grow in length during operation and must be replaced before they touch the housing.

Some MEMS, like for instance RF MEMS switches can be affected by creep, and the creep properties of thin metal films are not necessarily the same as those of the bulk material.

Creep in MEMS

Creep is well characterized at the macroscopic level (usually under constant-stress/constant-temperature conditions), but has not been studied extensively in microsystems and therefore it can represent a reliability problem in MEMS. Many microdevices are fabricated by adopting low-melting-point metals (in particular aluminum). In most recent research work, long-term or accelerated reliability tests (i. e., creep tests) seldom attract much attention. In fact, feasibility studies are rarely conducted when starting research work (e.g., academic research). However, reliability became an issue in applications, where durability and reproducibility are a key element for commercial success. Creep can form a danger for commercial devices fabricated from low-melting-point metals, especially when operating at higher temperatures and stresses. Therefore, at an early stage of the design of metal MEMS, creep should be taken into account.

An important source of information on creep in MEMS is the work done by Texas Instruments on the digital micromirror device (DMD). *Van Kassel* et al. [40.47] have shown that creep was a problem in their micromirrors. This reliability problem was termed the hinge memory effect. Figure 40.18 depicts the sketch of a DMD. A single mirror is mounted on hinges and electrostatic attraction between the mirror structure and a semiconductor memory cell below is responsible for a $\pm 10^{\circ}$ tilt. Hinge memory occurs when a mirror is operated in the same direction for a long period of time (high duty cycle). A residual tilt when the voltage is removed appears over time. The lifetime of the mirror is reached when the residual tilt becomes excessive (dead pixels). *Van Kassel* et al. [40.47] related this effect to creep of the hinge material. The lifetime was increased by selecting more creep-resistant materials and by improving the fabrication process. The hinge memory effect was thermally activated. At 65 °C and high duty cycle, mirrors could operate for about 5000 h continuously, while at 45 °C they remained functional for more than 100 000 h even at high duty cycles.

Lee et al. [40.48] presented a test specimen for measuring fatigue and creep in microscale structures. They



Fig. 40.18 SEM image of a test specimen (after [40.48])

patterned wafers in the shape of common (macroscopic) tensile test specimens with 50 µm wide 500 µm long metal beams (Fig. 40.19). Al and Al-1.5 at.% Ti beams were tested in a piezoactuator-driven test apparatus. In stress relaxation tests (the samples were quickly loaded to a certain stress and then held at this strain while monitoring the change of stress), *Lee* et al. [40.48] observed a load drop of 56 and 15% for Al and AlTi beams, respectively. The measurements were carried out for 10 min at room temperature. Figure 40.20 shows these results. *Lee* et al. believe that the difference in the amount of relaxation was from Al_3Ti precipitates that are formed at the Al grain boundaries in the alloy samples. They showed that the microstructure accounted for the relaxation difference in the investigated films.

Cho et al. [40.49] presented creep data at elevated temperatures of electrodeposited lithography, electroforming and molding (LIGA is a German acronym for Lithographie, Galvanoformung, Abformung) nickel dog-bone-shaped microsamples. Figure 40.21 presents creep tests conducted at 265 and 290 °C at stresses significantly below the yield strength. The results clearly show the primary-creep regime, where the creep rate decreases with time, and a steady-state regime, where strain rate is constant in time. *Cho* et al. showed that the time needed to attain 0.5% creep strain (considered as being a significant deformation) was more that 6 h at 265 °C, 1 h at 290 °C and only about 4 min at 400 °C.

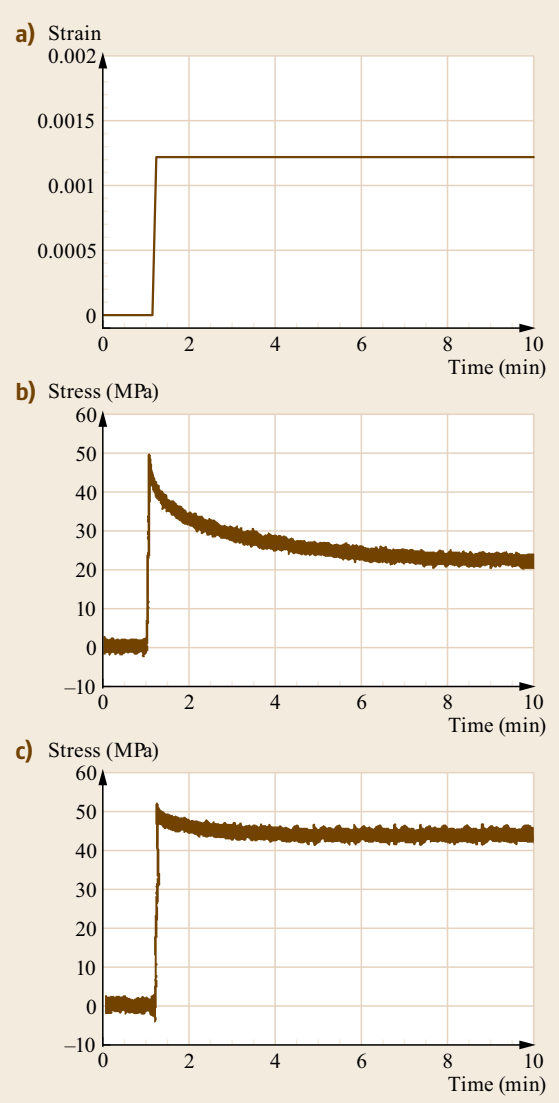


Fig. 40.19a–c Strain change with time **(a)** and corresponding stress change with time for Al beams **(b)** and AlTi beams **(c)** (after [40.48])

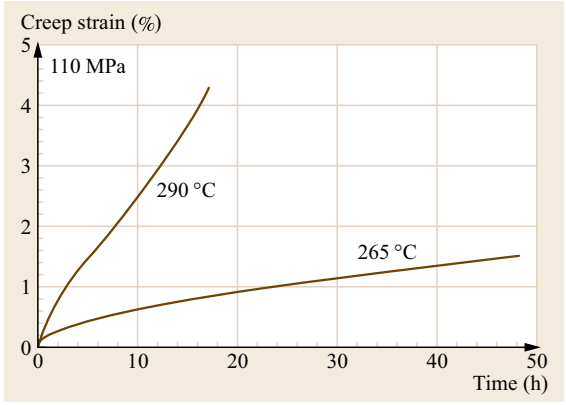


Fig. 40.20 Creep curves at 265 and 290 °C under the stress of 110 MPa (after [40.49])

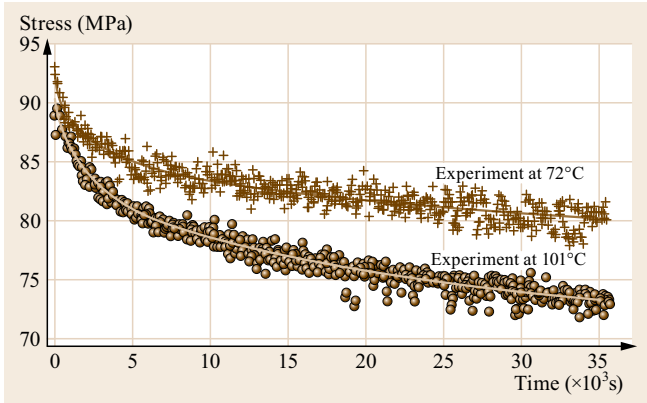


Fig. 40.21 Stress relaxations for Al-0.4 at.% Cu 5 µm thick film at 72 and 101 °C. *Solid lines* represent the fitted creep mechanism: dislocation glide limited by obstacles (after [40.50])

It is clear that the creep phenomenon is thermally activated and much more harmful to devices at elevated temperatures. However, *Larson* et al. [40.51] and *Yin* et al. [40.52] reported that some creep activity of nanocrystalline Ni was observed even at room and low temperatures. This means that for those materials care must be taken under almost any temperature conditions.

Hardened Pt alloys for MEMS have been presented by *Brazzle* et al. [40.53] to serve as creep-resistant materials. Solution-hardened Pt alloys for improved reliability of electromagnetically actuated micromirror devices exhibited precision angular displacement without observable fatigue and creep when used in flexures. *Brazzle* et al. [40.53] found that alloy 851 (79%Pt-15%Rh-6%Ru) and Pt(10%Ir) were very hard flexure materials. The alloying components (rhodium, ruthenium and iridium) impede dislocation movement, thereby increasing the yield strength and reducing fatigue and creep.

Modlinski et al. [40.50] have presented a simple way of selecting creep-resistant aluminum alloys for use in RF MEMS switches, but the technique is generically applicable. They propose substrate curvature measurements to study creep properties of Al alloy films using isothermal tensile stress relaxation (Fig. 40.22). The tensile stress relaxation data fitted well to a known deformation mechanism (dislocation glide limited by obstacles) in the temperature range 60–110 °C. The Al alloys studied were characterized by two creep parameters: the activation energy ΔF and the athermal flow stress τ. The higher these param-

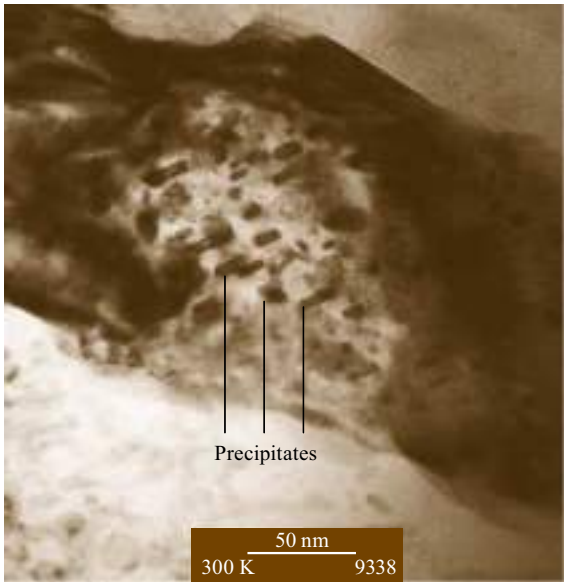


Fig. 40.22 TEM image of AlCuMgMn. Strong and highly dense precipitates hinder the dislocation movement in the grain interior

eters are, the more creep-resistant the alloy is. They selected a very promising Al alloy with the composition $Al_{93.5}Cu_{4.4}Mg_{1.5}Mn_{0.6}$ which has a very high ΔF (6.4 ± 2.5 eV) and τ (920 ± 420 MPa). It was also shown that there is a relation between the creep parameters and the microstructure of the Al alloys in more detail. The activation energy describes the strength of the precipitates, and also gives information about the coherency of the second-phase particles as well. The athermal flow stress, on the other hand, gives information on the density and arrangement of the obstacles. A transmission electron microscope (TEM) image of the AlCuMgMn alloy revealed very small, strong, coherent and/or semi-coherent (high-ΔF) and highly dense (high-τ) precipitates.

Creep is caused by the motion of dislocations inside grains in this temperature range. Impeding the dislocation motion inside the grains is a very effective way of increasing creep resistivity of an $Al_{99.6}Cu_{0.4}$ alloy thin film. It was done by a macroscopically well-known hardening mechanism: precipitation hardening. An increase of both ΔF (by 50%) and τ was observed.

RF MEMS switches are starting to be commercialized, but reliability is still a problem. Hot switching and the impact of RF power on the reliability of the MEMS switches has been extensively investigated, mainly by simulations. *Jensen* et al. [40.54], *Rottenberg* et al. [40.55] and many others reported that temperatures due to RF power present in some parts of a movable metal bridge are expected to be as high as 250 °C [40.54] and can easily exceed 100 °C. The high residual stresses (50–200 MPa) present in an RF MEMS bridge structure and the high temperatures caused by RF power may lead to creep deformation, deteriorating the functionality of a switch over time.

Creep Conclusion

At the macroscale level, creep tests are performed using tensile specimens subjected to constant stress and temperature. These tests are rather easy to prepare and perform. In microscale experiments, the situation is very different. There are no standardized specimens, and there is no standard way of investigating creep in microdevices. Some perform the tests on real MEMS devices (DMD); some try to mimic macroscopic tensile test specimens and use them in the microworld. A very good idea, although not often used, is to implement well-known macroscopic techniques to render a material more creep-resistant. For example, hardening mechanisms such as solution hardening or precipitation hardening are interesting in this respect. These treatments produce stronger obstacles and hence decrease dislocation motion.

The microstructure (grain sizes, precipitates in case of alloys) of deposited films is very process-dependent. There are no standard deposited films. For instance, Al alloy films will exhibit different creep behavior when sputtered under different conditions. Their properties depend on the deposition methods as well as conditions. This variation is a considerable challenge for microscopic creep research.

A human factor is certainly an issue in understanding the state of the art of creep in MEMS. Most engineers and scientists dealing with MEMS are silicon-or, in general semiconductor-oriented. Creep is not really an issue in monocrystalline silicon, polysilicon or silicon-germanium. There is probably a lack of know-how on plastic deformation in the MEMS society.

Research work in a field of metal plasticity in MEMS has been started. Increasing numbers of publications on understanding and investigating creep phenomenon on the microscale are being issued. MEMS engineers and scientists are facing new problems and challenges in metal MEMS, but fundamental research programs, discussions and exchange of views and knowledge of the plastic deformation of metals between specialists will speed up research in this field.

40.3.2 Fatigue

Fatigue of Metals

The only large source of information on fatigue of metal MEMS devices is again the research carried out by Texas Instruments of the digital micromirror device

(DMD). Macroscopically, the lifetime of many metal systems is limited by fatigue. A large, repetitive mechanical stress causes the accumulation of dislocation defects at the surface, resulting in fatigue cracks. In microscopic systems, the structural parts are only one, or a few grains thick, with very little dislocations. Therefore, not enough damage can accumulate at the surface to cause fatigue failures. It has been found indeed that MEMS devices suffer much less from fatigue than their macroscopic counterparts [40.56].

Fatigue behavior of materials is often shown in a so-called S-curve, in which the number of stress cycles to fracture is shown as a function of the applied maximum stress level. We usually find that at very high stresses only a small number of cycles is required to fracture the device, the low-cycle fatigue regime (Fig. 40.23). For lower stresses, the number of cycles increases. For very long low-stress cycling times, we find that different materials have different properties. Most steels, e.g., have a fatigue limit (Fig. 40.23a). Below this limit, the structure survives forever. Other materials, like aluminum, do not have such a threshold. The curve decreases slowly, even for very large numbers of cycles (Fig. 40.23b). Brittle materials such as ceramics do not have a cycling fatigue effect; below the fracture strength the lifetime is infinite, above it, they immediately break without any cyclic stress effect (Fig. 40.23c).

It has been found that high-cycle fatigue is thin films is very unlikely, as explained above. Aluminum RF MEMS switches have likewise never been reported to suffer from high-cycle fatigue [40.57]. An interesting feature of microscopic structures however is that low-cycle fatigue has been found in aluminum [40.58, 59]. This raises the question whether in thin films at very high cyclic stresses new defects are being formed, while at lower stresses only existing defects can move to the surface (and there are too few of them to do any harm). More research is definitely needed to figure out how the process really behaves, and where the turning point is between microscopic and macroscopic fatigue behavior. For the moment we can only say qualitatively that metals that macroscopically tend to have no threshold for fatigue do have such a threshold when used as thin films. Therefore the S-curve for thin-film aluminum resembles that of bulk steel (Fig. 40.23a) instead of bulk aluminum (Fig. 40.23b).

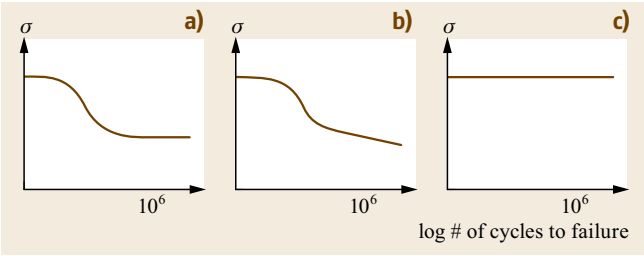


Fig. 40.23 (a) S-curve with threshold; (b) S-curve without threshold; (c) S-curve for ceramic

Fatigue of Silicon

From the early days of MEMS technology on, silicon in its mono- and polycrystalline form has been extensively used as a structural material. Silicon is a brittle material, and as such, should not intrinsically suffer degradation from fatigue (Fig. 40.23c). However, in air it is always covered with a native oxide, and concern was raised that this oxide might be susceptible to a stress corrosion cracking (SCC) phenomenon that is well known in glass technology [40.60]. SCC is the stress-assisted hydrolysis of the SiO_2 bonds at the crack tip under the influence of tensile stress and ambient humidity, causing crack propagation because atomic bonds are broken.

The phenomenon is thought to happen in the following way (Fig. 40.24). A small crack at the surface in the SiO_2 native oxide grows slightly due to SCC under the influence of a tensile stress, so that the silicon below the crack is slightly closer to the surface and oxidizes. Now the crack can propagate through the extended oxide layer, causing a new oxidation step in the silicon below. This continues until the whole structure fractures.

Connally and others [40.61–65] initiated research into SCC in silicon microstructures, and indeed found a large humidity influence on the crack propagation rate. *Muhlstein* et al. [40.66] measured a more realistic configuration in which they did not only monitor crack propagation but also took into account crack initiation. Another advantage over the earlier measurements was that the stress intensity factor at the tip was not a required parameter to have a meaningful description. Similar measurements were also reported by *Kapels* et al. [40.67] and *Kahn* et al. [40.68].

It is likely that there is a critical stress in the silicon device, above which SCC may occur. The native oxide of silicon is normally under a compressive stress [40.5] and the applied external stress should be so large that this stress has to become tensile for SCC to occur at all.

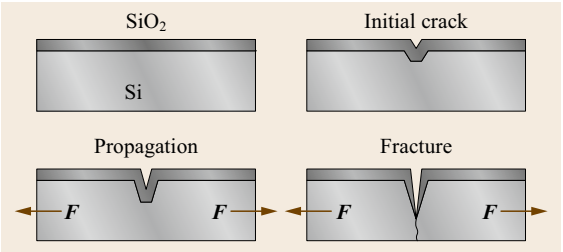


Fig. 40.24 SSC failure mechanism in silicon

If the stress in the silicon is kept below this transition point, where the oxide surface layer turns from its intrinsic compressive stress to tensile, no problems due to SCC are expected.

Bagdahn and *Sharpe* [40.69] put all the data available, including their own, in one table and showed that SCC cannot explain all the properties of the behavior that has been found. Being a chemical, stress-assisted process, SCC should not have a pronounced cycling-speed dependency; only the absolute time that the stress is applied counts towards the reaction rate. *Bagdahn* and *Sharpe* [40.69] showed that the data available provide a definite trend of the number of cycles to failure as a function of the applied stress, which contradicts the SCC assumption. We see in Fig. 40.25 that these effects only show up at very high stresses, near the fracture strength of silicon. Therefore in practical MEMS devices the fatigue of silicon may not be a big issue because it is good design practice to keep some margin to the fracture strength for brittle materials anyway. However, in harsh environments it may turn out to be a problem.

40.3.3 Wear

Wear is one of the major limiting factors in MEMS technology. After its occurrence had been shown by *Gabriel* et al. [40.70] and *Mehregany* et al. [40.71] in silicon micromotors in 1990, it was quickly recognized that wear is a severe limiting factor for almost all MEMS devices incorporating rubbing/sliding surfaces. At Sandia National Laboratories, a large effort was spent on understanding this effect, with experiments on a dedicated micromotor test vehicle (Fig. 40.26). This resulted in a quantitative empirical model to predict wear in certain types of micromotors [40.72]. Beautiful focused ion beam (FIB) cross sections in this work show that wear is indeed severe (Fig. 40.27). The wearing mechanism proposed is a kind of *adhesive wear*, where the surfaces adhere at their contacting points and shear off parts of the other surface (Fig. 40.28). Abrasive and corrosive wear take place at higher contact forces, or when a chemical reaction assists the wear process. A comprehensive overview of their work can be found in their MEMS reliability report [40.73].

A good overview from the middle of the 1990s is given by *Komvopoulos* [40.21]. Studying wear at a more fundamental level is difficult, because typical analytical measurement methods do not cover the whole range of speeds and contact forces in MEMS devices (Fig. 40.29) [40.74]. Atomic force microscopy/friction force microscopy (AFM/FFM) systems are very accurate in mimicking a single asperity, but only at, for MEMS devices, very low speeds [40.75]. A dedicated *tribo-apparatus* measurement system [40.76] fills part of the gap, but a large area remains outside the scope of the different (microscopic) measurement methods. Atomistic simulations can usually only be applied at much higher speeds, because otherwise computing time

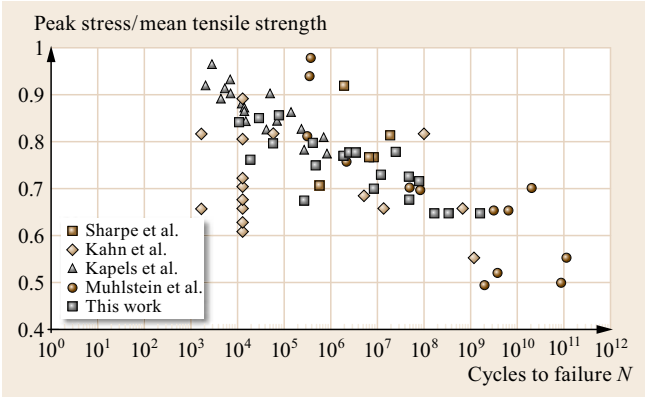


Fig. 40.25 *Bagdahn* and *Sharpe* [40.69] show that a definite trend appears if the applied stress is plotted as a function of the number of cycles to failure, contradicting the basic assumption of SCC

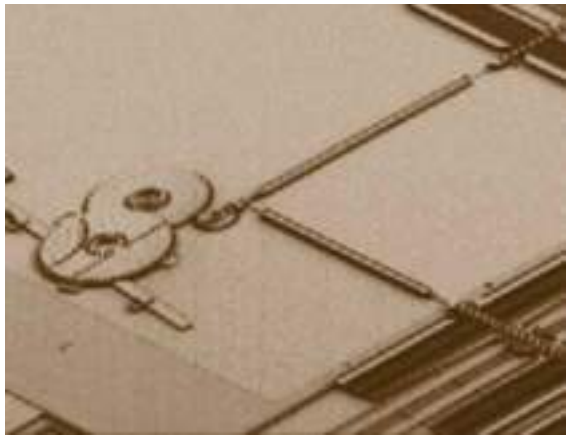


Fig. 40.26 Sandia's micromotor test vehicle to study wear

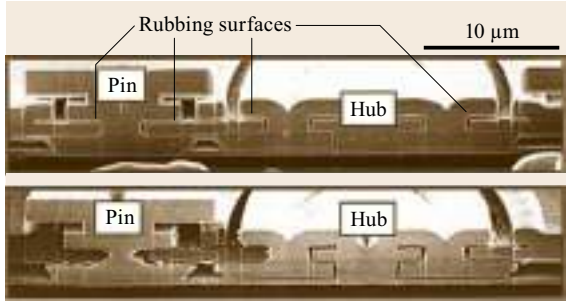


Fig. 40.27 FIB cross section of a new and a worn micromotor

becomes prohibitively long. As a result, wear is often studied using MEMS devices themselves as the measurement vehicle [40.77–80], but, of course, this is generally not as accurate as one would like.

Rymuza [40.81, 82] proposed the use of graphitelike materials embedded in the wear layer of a MEMS device. As the device loses its surface layer due to wear, small amounts of graphite are released to lubricate the device.

Recently, a lot of activity has been going on with respect to coating silicon MEMS devices with a friction-reducing and/or wear-resistant coating after the device has been released. This coating usually takes the form of a hard, chemically inert layer with a thickness of nanometers. Experiments have been performed using tungsten [40.83], SiN_x [40.84], diamondlike carbon (DLC) [40.85–88] ultrananocrystalline diamond (UNCD) [40.89], and SiC [40.90–92], among others. SAM coatings (Sect. 40.2.1) have also been investigated for this purpose but tend to wear off during rubbing [40.93].

The results of these studies show that especially DLC, UNCD and SiC are very promising coating materials, but also that their properties severely depend on the exact composition, roughness and surface termination (passivation) of the film. At this moment, there is no quantitative theory that is able to predict which coating compositions/properties are the best. Therefore optimization of these films is performed mainly using an experimental approach.

One finds that commercial applications of devices that have continuously rubbing/sliding surfaces, such as rotary micromotors, currently do not exist. This means that wear problems in MEMS devices are still so large that designers of commercial products avoid designs where rubbing surfaces have to survive more than a couple of million cycles.

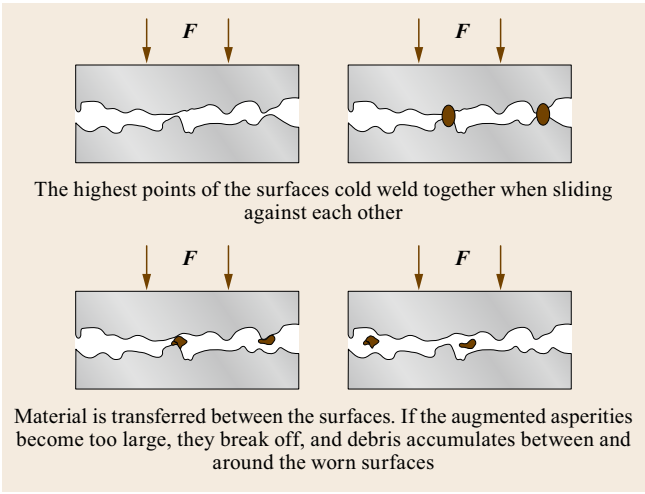


Fig. 40.28 Principle of adhesive wear

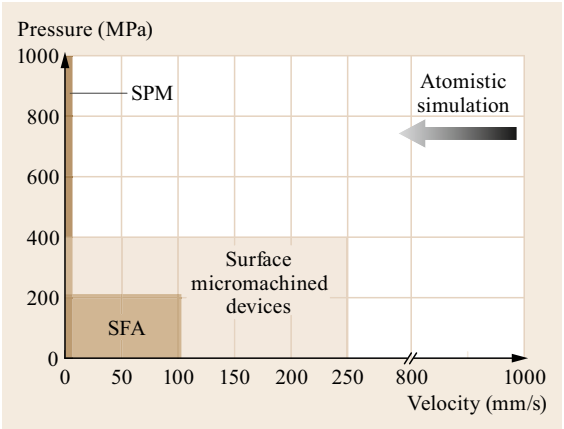


Fig. 40.29 MEMS surfaces have sliding speeds and contact forces not easily assessed by the available analysis methods: scanning probe microscopy (SPM) and surface force apparatus (SFA) (after [40.74])

40.3.4 Packaging

Introduction

Many MEMS contain movable and fragile parts. A specific packaging approach is required for protection during fabrication as well as during operation. For instance, standard wafer dicing will, if not destroy the MEMS device, at least introduce some contamination and particles that can lead to failure. Hence packaging must be carried out on the wafer during wafer processing, prior to die singulation. This packaging step is referred to as wafer-level or 0-level packaging. It defines the first protective interface for the MEMS device, achieved through on-wafer encapsulation of the movable parts in a sealed cavity. The cavity must be strong, for many applications hermetic, and equipped with electrical signal feedthroughs. An example of a 0-level packaging scheme for RF MEMS switches, which relies on wafer or chip stacking techniques using benzocyclobutene (BCB) as the bonding and sealing material, is shown in Fig. 40.30 [40.94].

Failure Modes and Failure Mode Investigation

Important requirements of MEMS packaging are resistance (against shear stress, pull stress, mechanical, thermal loads, etc.), and hermeticity against moisture, liquids, particles or gas contaminants, but also hermeticity to keep a controlled ambient (gas and pressure) inside the package. The most common failure modes that can occur are mainly related to the sealing material itself. For instance, a package can lose its shear

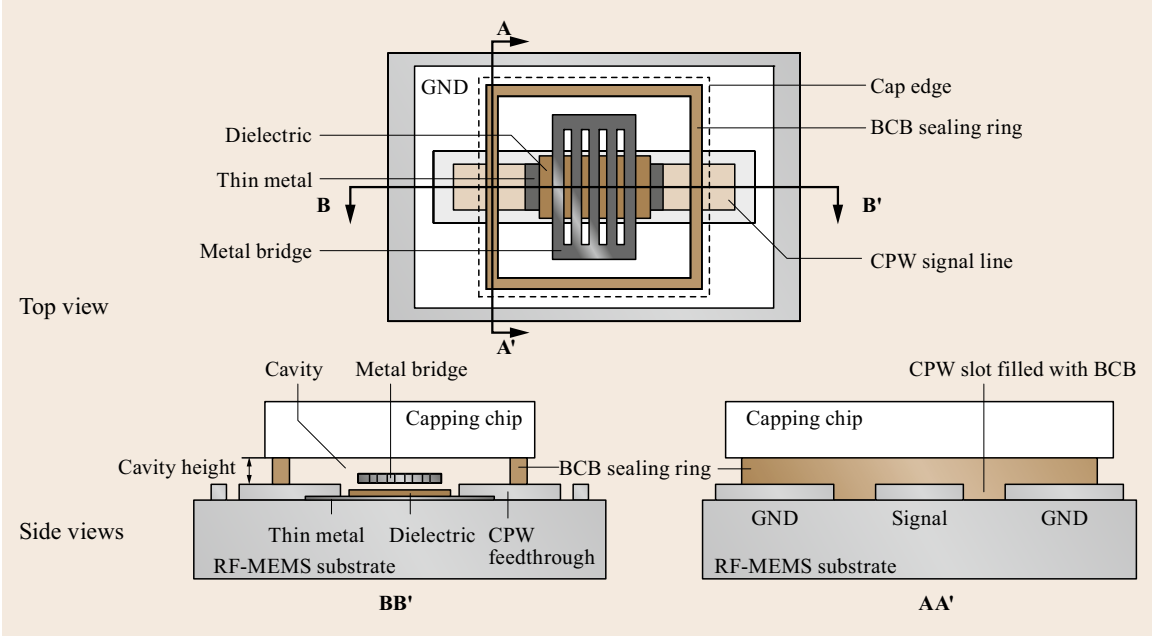


Fig. 40.30 Illustration of a wafer or 0-level packaged RF MEMS switch. The fragile MEMS part is housed in a cavity formed by a capping chip, bonded to the RF MEMS substrate using a layer of BCB

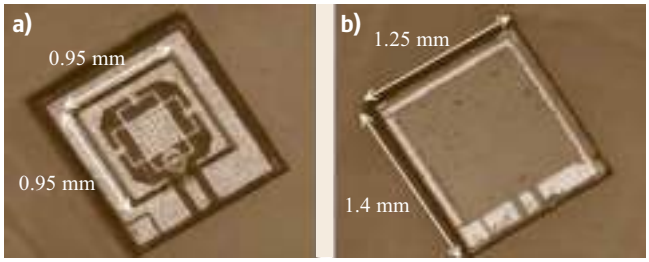


Fig. 40.31a,b Picture of a laser-diced *naked* **(a)** and packaged **(b)** RF tunable capacitor using solder (SnPb) flip-chip assembly

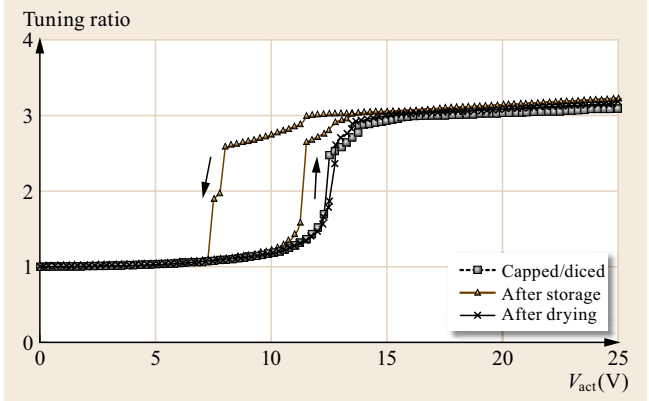


Fig. 40.32 Tuning ratio of the packaged tunable capacitor of Fig. 40.20 right after dicing (*squares*), after storage in ambient air (*triangles*) and after drying in dry N_2 at 75 °C (*crosses*)

or pull resistance in the case of bad adhesion between the sealing material and the MEMS die, although this is rarely the case and is process-related. More important is the loss of hermeticity in the case of an hermetic package. This can happen with the appearance of cracks in the sealing layer or at the bonding interfaces when the package device is subject to thermal shocks, for instance. The difference between the thermal coefficients of expansion of the different materials can induce a very high stress at the bonding interfaces and create some cracks, thus inducing gross leaks.

The importance of this can be seen in the following example. The package of Fig. 40.31b is made of a Si capping chip flip-chip bonded on top of an RF tunable capacitor using a solder seal (SnPb). Figure 40.32 shows the tuning ratio of the capacitor as a function of the actuation voltage, right after packaging and laser dicing, after storage in air, and after drying in dry N_2 at 75 °C, respectively. Compared to the curve obtained right after dicing, the tuning curve after storage in air is shifted towards lower actuation voltages and shows nonreversible tuning. This phenomenon could be explained by a moisture-related charging effect of the RF capacitor when the package is stored in ambient air. This is confirmed by the tuning curve obtained after drying the packaged device in dry N_2 at 75 °C: the initial reversible tuning curve is restored. This clearly shows that, in this example, the package is not hermetic to moisture and that this influences the electrical

performance of the MEMS device. Possible causes of packaging failure here could be microcracks in the solder sealing ring, bad adhesion between the solder and the MEMS die, particle contamination in the solder, or a local scarcity of solder.

For the same reasons, in the case of a (supposedly) hermetic package containing a controlled ambient (pressure and gas), the loss of hermeticity due to a leak in the sealing material will fill the package with air after a determined period of time, depending on the leak rate. For instance, a package of 100 nl (a few tens of nl are typically encountered in MEMS packaging) with an initial internal pressure of 1 mbar and displaying a leak rate as low as 10^{-10} mbar l/s will be filled with air after only 11 days.

Different testing methods exist or have been developed to first test the package (e.g., shear and pull strength, hermeticity, etc.) and second, to investigate the failure mechanisms. One can always refer to the MIL-STD-883 procedure to obtain the testing conditions and the rejection limits for the test [40.95]. This is the case for instance for the pull and shear testing (MIL-STD-883, method 2019.5). Hermeticity (fine leak) testing as specified by the MIL-STD-883 (method 1014.9) is still a problem because the fine-leak test conditions and specifications are clearly not applicable to MEMS packages due to the small cavity volumes they deal with [40.96]. Alternative methods are needed to investigate the leak rate of MEMS packages. However, gross-leak testing using fluorocarbon fluids is still applicable and is a good method to localize a leak path (e.g., a crack or a bad bonding) in the sealing material or the package itself. Similar to gross-leak testing is the dye penetrant method, as illustrated in Fig. 40.33. The package is immersed in a pressurized yellow dye (type Zyglo ZL-56 from Magnaflux). Next, the sample is rinsed with water, dried, and sprayed with a developer (type Zyglo ZP-9F). This developer has the property of sucking out the dye from the package if a leak exists, and some yellow fringes are then clearly visible under UV light.



Fig. 40.33 UV inspection of a leaky package using the dye-penetrant method

Another method to localize a gross leak in a package is scanning acoustic microscopy (SAM). A picture of a polymer-sealed leaky package is shown in Fig. 40.34 [40.97]. The black areas indicate that there is no adhesion between the sealing layer and the Si chip.

Other methods to look at the bonding interface are x-ray imaging and cross sectioning. Figure 40.35 is an x-ray image of a solder-sealed Si package. Irregularities

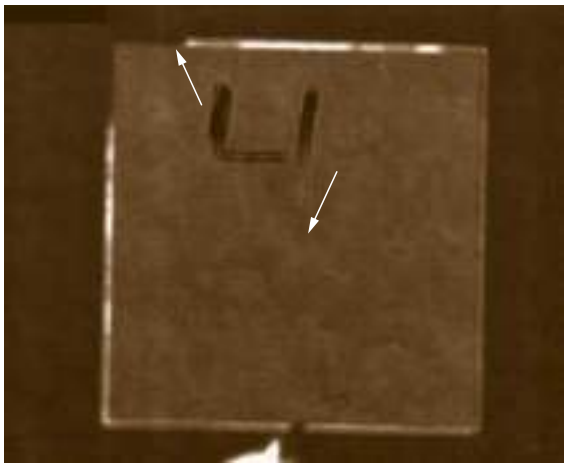


Fig. 40.34 SAM picture of a Si cap bonded to a Si wafer using polymer. The *black areas* indicate positions where the bond failed (*arrows*) (after [40.97])

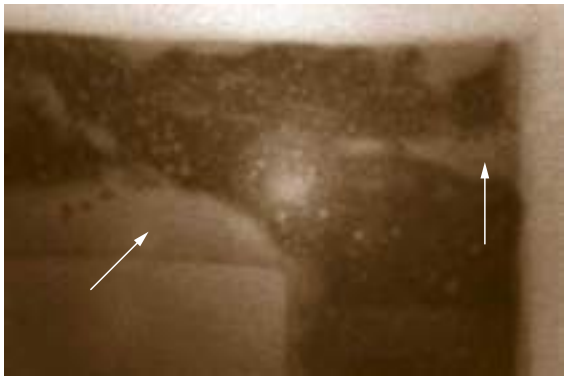


Fig. 40.35 X-ray image of a solder-sealed Si package. *White areas* are assumed to be solder oxide (unbonded areas)

in the solder ring are visible, and the lighter areas are assumed to be oxidized solder areas, and thus potential leaky paths.

Cross sectioning is of course destructive for the package, but it can also be used to investigate the bond quality, as shown in the example of Fig. 40.36. In this picture, voids in the sealing layer are clearly visible, and are potential source of gross leaks.

Conclusion

Investigating the failure mechanisms of the package is at least as important as for the MEMS itself, because without a good and robust package, most MEMS devices cannot survive operating or storage conditions. Understanding why a package failed will help to improve the next generation of MEMS. A lot of development still has to be done in packaging reliability testing, and this requires dedicated methods of investigation, which is now slowly emerging.

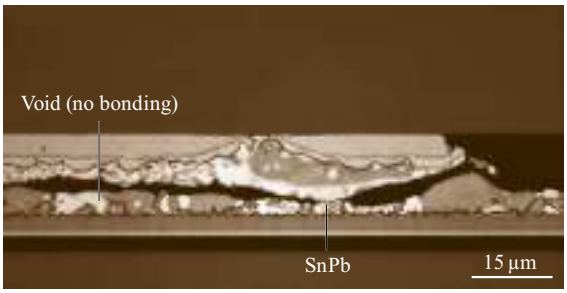


Fig. 40.36 Optical microscopy image of a cross section of a solder-sealed Si package

40.4 Conclusions

We have seen that a lot about the reliability of MEMS/NEMS is already known, and that in some cases even macroscopic principles can be applied to improve the reliability of micromachines over the state of the art. The fact that reliability is such a problem is not because MEMS/NEMS are intrinsically unreliable. They are not, and in fact – due to the low mass and high strength of materials on a microscale – the reliability that can be attained is even higher than in many macroscopic situations. The reason that reliability is hampering the commercialization of MEMS has already briefly been addressed in the creep section: to produce any MEMS at all is already a considerable feat, so that reliability issues are often not taken into account in the early stages of development. In the macroscopic world, instead, reliability is one of the key design issues of almost any product.

This sense still has to pervade the MEMS society, although things are changing. Especially in commercial laboratories, *design for reliability* is the new mantra, and independent research and development labs have set up dedicated groups to study fundamental issues in microscale reliability, such as stiction, charging, creep, fatigue, wear and packaging. If this trend continues over the coming years, MEMS may become just as pervasive as microelectronics is now.

References

40.1 W.M. Miller, D.M. Tanner, S.L. Miller, K.A. Peterson: MEMS reliability. The challenge and the promise. In: *Proc. 4th Annu. Reliab. Chall., Dublin* (Sandia National Labs, Albuquerque 1998) pp. 4-1–4-7

40.2 M.R. Douglass: DMD reliability. A MEMS success story, Proc. SPIE **4980**, 1–11 (2003)

40.3 J. Bienstman: From product to production in automotive MEMS. In: *Proc. MicroMech. Eur* (IOP, Leuven 2004) pp. 107–108

40.4 W.M. van Spengen: MEMS reliability from a failure mechanisms perspective, Microelectron. Reliab. **43**, 1049–1060 (2003)

40.5 B. Stark (Ed.): *MEMS Reliability Assurance Guidelines for Space Applications* (National Aeronautics Space Administration (NASA) and Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena 1999)

40.6 D.M. Tanner, N.F. Smith, D.J. Bowman, W.P. Eaton, K.A. Peterson: First reliability test of a surface micromachined microengine using SHiMMeR, Proc. SPIE **3224**, 14–23 (1997)

40.7 G.F. LaVigne, S.L. Miller: A performance analysis system for MEMS using automated imaging methods. In: *Proc. IEEE Int. Test Conf.* (IEEE, Washington 1998) pp. 442–447

40.8 W.M. van Spengen, R. Puers, R. Mertens, I. De Wolf: High resolution optical investigation of small out-of-plane movements and fast vibrations; characterization and failure analysis of MEMS, Microsyst. Technol. **10**, 89–96 (2004)

40.9 P. Krehl, S. Engemann, C. Rembe, E.P. Hofer: High-speed visualization, a powerful diagnostic tool for microactuators – Retrospect and prospect, Microsyst. Technol. **5**, 113–132 (1999)

40.10 C. Rembe, L. Muller, R.S. Muller, R.T. Howe: Full three-dimensional motion characterization of a gimballed electrostatic microactuator. In: *Proc. IEEE Annu. Int. Reliab. Phys. Symp. (IRPS)*, ed. by E.S. Sneyder (IEEE, Orlando 2001) pp. 91–98

40.11 J.S. Burdess, A.J. Harris, D. Wood, R.J. Pitcher, D. Glennie: A system for the dynamic characterization of microstructures, J. MEMS **6**, 322–328 (1997)

40.12 M.R. Hart, R.A. Conant, K.Y. Lau, R.S. Muller: Stroboscopic interferometer system for dynamic MEMS characterization, J. MEMS **9**, 409–418 (2000)

40.13 C.H. Mastrangelo, C.H. Hsu: Mechanical stability and adhesion of microstructures under capillary forces – Part I: Basic theory, J. MEMS **2**, 33–43 (1993)

40.14 C.H. Mastrangelo, C.H. Hsu: Mechanical stability and adhesion of microstructures under capillary forces – Part II: Experiments, J. MEMS **2**, 44–55 (1993)

40.15 M.P. de Boer, J.A. Knapp, T.M. Mayer, T.A. Michalske: The role of interfacial properties on MEMS performance and reliability, Proc. SPIE **3825**, 2–15 (1999)

40.16 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids* (Clarendon, Oxford 1950)

40.17 R. Maboudian, R.T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B **15**, 1–20 (1997)

40.18 R. Legtenberg, H.A.C. Tilmans, J. Elders, M. Elwenspoek: Stiction of surface micromachined structures after rinsing and drying: Model and investigation of adhesion mechanisms, Sens. Actuators A **43**, 230–238 (1994)

40.19 N. Tas, T. Sonnenberg, H. Jansen, R. Legtenberg, M. Elwenspoek: Stiction in surface micromachining, J. Micromech. Microeng. **6**, 385–397 (1996)

40.20 J. Israelachvili: *Intermolecular and Surface Forces* (Academic, London 1991)

40.21 K. Komvopoulos: Surface engineering and microtribology for microelectro-mechanical systems, Wear **200**, 305–327 (1996)

40.22 W.M. van Spengen, W.M.R. Puers, I. De Wolf: A physical model to predict stiction in MEMS, J. Micromech. Microeng. **12**, 702–713 (2002)

40.23 J.A. Greenwood, J.B.P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. A **295**, 300–319 (1966)

40.24 B. Bhushan: Methodology for roughness measurement and contact analysis for optimization of interface roughness, IEEE Trans. Mag. **32**, 1819–1825 (1996)

40.25 B. Bhushan: *Principles and Applications of Tribology* (Wiley, New York 1999)

40.26 L. Kogut, K. Komvopoulos: Analysis of interfacial adhesion based on electrical contact resistance, J. Appl. Phys. **94**, 6386–6390 (2003)

40.27 Y.-P. Zhao, L.S. Wang, T.X. Yu: Mechanics of adhesion in MEMS – A review, J. Adhes. Sci. Technol. **17**, 519–546 (2003)

40.28 J. Elders, H.V. Jansen, M. Elwenspoek: Materials analysis of fluorocarbon films for MEMS applications. In: *Proc. Investig. Micro Struct. Sens. Actuators Mach. Robot. Syst.* (IEEE, New York 1994) pp. 170–175

40.29 P.F. Man, B.P. Gogoi, H. Mastrangelo: Elimination of post-release adhesion in microstructures using conformal fluorocarbon coatings, J. MEMS **6**, 25–34 (1997)

40.30 U. Srinivasan, M.R. Houston, R.T. Howe, R. Maboudian: Alkyltrichlorosilane-based self-assembled monolayer films for stiction reduction in silicon micromachines, J. MEMS **7**, 252–260 (1998)

40.31 R. Maboudian, W.R. Ashurst, C. Carraro: Self-assembled monoayers as anti-stiction coatings for MEMS: Characteristics and recent developments, Sens. Actuators A **82**, 219–223 (2000)

40.32 B.-H. Kim, C.-H. Oh, K. Chun, T.-D. Chung, J.-W. Byun, Y.-S. Lee: A new class of surface modifiers for stiction reduction. In: *Proc. 12th Int. Conf. Micro Electro Mech. Syst.* (IEEE, Piscataway 1999) pp. 189–193

40.33 K.A. Peterson, P. Tangyunyong, A.A. Pimentel: Failure analysis of surface micromachined microengines, Proc. SPIE **3512**, 190–200 (1998)

40.34 J. Wibbeler, G. Pfeifer, M. Hietschold: Parasitic charging of dielectric surfaces in capacitive microelectromechanical systems (MEMS), Sens. Actuators A **71**, 74–80 (1998)

40.35 A.R. Knudson, S. Buchner, P. McDonald, W.J. Stapor, A.B. Campbell, K.S. Grabowski, D.L. Knies: The effects of radiation on MEMS accelerometers, IEEE Trans. Nucl. Sci. **43**, 3122–3126 (1996)

40.36 C.I. Lee, A.H. Johnston, W.C. Tang, C.E. Barnes: Total dose effects on microelectromechanical systems (MEMS): Accelerometers, IEEE Trans. Nucl. Sci. **43**, 3127–3132 (1996)

40.37 G.M. Rebeiz: *RF MEMS: Theory, Design and Technology* (Wiley, Hoboken 2003)

40.38 C.L. Goldsmith, J. Ehmke, A. Malczewski, B. Pillans, S. Eshelman, Z. Yao, J. Brank, M. Eberly: Lifetime characterization of capacitive RF MEMS switches. In: *Proc. IEEE MTT-S Int. Microw. Symp.* (IEEE, New York 2001) pp. 227–230

40.39 S. Zafar, A. Callegari, E. Gusev, M.V.P. Fischetti: Charge trapping in high k gate dielectric stacks. In: *Proc. Int. Electron Devices Meet. (IEDM)*, ed. by S. Ikeda (IEEE, San Francisco 2002) pp. 517–520

40.40 J.C. Ehmke, C.L. Goldsmith, Z.J. Yao, S.M. Eshelman: Method and apparatus for switching high frequency signals, US Patent 6391675 (1999)

40.41 J.R. Reid: Simulation and measurement of dielectric charging in electrostatically actuated capacitive microwave switches. In: *Proc. Model. Simul. Microsyst. (MSM)*, ed. by M. Laudon, B. Romanowicz (NSTI, San Juan 2002) pp. 250–253

40.42 W.M. van Spengen, R. Puers, R. Mertens, I. De Wolf: A comprehensive model to predict the charging and reliability of capacitive RF MEMS switches, Micromech. Microeng. **14**, 514–521 (2004)

40.43 X. Rottenberg, B. Nauwelaers, W. De Raedt, H.A.C. Tilmans: Distributed dielectric charging and its impact on RF MEMS devices. In: *Proc. 34th Eur. Microw. Conf.* (Artech House, Amsterdam 2004) pp. 77–80

40.44 W.M. van Spengen, R. Puers, I. De Wolf: RF MEMS reliability – The challenge, the physics, and the reward. In: *Proc. MME*, ed. by R. Puers (IOP, Leuven 2004) pp. 319–325

40.45 F.N.R. Nabarro, H.L. de Villiers: *The Physics of Creep* (Taylor Francis, New York 1995)

40.46 H.E. Boyer: *Atlas of Creep and Stress-Rupture Curves* (ASM International, Metals Park 1988)

40.47 P.F. Van Kessel, L.J. Hornbeck, R.E. Meier, M.R. Douglass: A MEMS-based projection display, Proc. IEEE **86**, 1687–1704 (1998)

40.48 H.-J. Lee, G. Cornella, J.C. Bravman: Stress relaxation of free-standing aluminum beams for microelectromechanical systems applications, Appl. Phys. Lett. **76**, 3415–3417 (2000)

40.49 H.S. Cho, K.J. Hemker, K. Lian, J. Goettert, G. Dirras: Measured mechanical properties of LIGA Ni structures, Sens. Actuators A **103**, 59–63 (2003)

40.50 R. Modlinski, A. Witvrouw, P. Ratchev, V. Simons, A. Jourdain, H.A.C. Tilmans, R. Puers, J. den Toonder, I. De Wolf: Creep as a reliability problem in MEMS, J. Microelectron. Reliab. **44**, 1733–1738 (2004)

40.51 K.P. Larsen, A.A. Rasmussen, J.T. Ravnkilde, M. Ginnerup, O. Hansen: MEMS devices for bending test: Measurements of fatigue and creep of electroplated nickel, Sens. Actuators A **103**, 156–164 (2003)

40.52 W.M. Yin, S.H. Whang, R. Morshams, C.H.P. Xiao: Creep behavior of nanocrystalline nickel at 290 and 373 K, Mater. Sci. Eng. A **2301**, 18–22 (2001)

40.53 J.D. Brazzle, W.P. Taylor, B. Ganesh, J.J. Price, J.J. Bernstein: Solution hardened platinum alloy flexure materials for improved performance and reliability of MEMS devices, J. Micromech. Microeng. **15**, 43–48 (2005)

40.54 B.D. Jensen, J.L. Volakis, K. Saitou, K. Kurabayashi: Impact of skin effect on thermal behavior of RF-MEMS switches. In: *6th ASME-JSME Conf.*, ed. by S. Nishio, A. Lavine (Kona, Hawaii 2003), TED-AJ03-420

40.55 X. Rottenberg, B. Nauwelaers, W. De Raedt, H.A.C. Tilmans: RF current and power handling of RF-MEMS shunt switches. In: *Proc. MEMSWAVE* (2004) pp. C1–C4

40.56 M.R. Douglass: Lifetime estimates and unique failure mechanisms of the digital micromirror device. In: *36th Int. Reliab. Phys. Symp. (IRPS)*, ed. by A.N. Campbell (IEEE, Reno 1998) pp. 9–16

40.57 J.J. Yao: RF MEMS from a device perspective, J. Micromech. Microeng. **10**, R9–R38 (2000)

40.58 D.T. Read, J.W. Dally: Fatigue of microlithographically patterned free-standing aluminum thin film under axial stresses, J. Electron. Packag. **117**, 1–6 (1995)

40.59 G. Cornella, R.P. Vinci, R. Suryanarayanan Iyer, R.H. Dauskardt, J.C. Bravman: Observations of low-cycle fatigue of Al thin films for MEMS applications, Mater. Res. Soc. Symp. Proc. **518**, 81–86 (1998)

40.60 S.M. Wiederhorn, E.R. Fuller Jr., R. Thomson: Micromechanisms of crack growth in ceramics and glasses in corrosive environments, Met. Sci. **14**, 450–458 (1980)

40.61 W.W. van Arsdell, S.B. Brown: Subcritical crack growth in silicon MEMS, J. MEMS **8**, 319–327 (1999)

40.62 J.A. Connally, S.B. Brown: Micromechanical fatigue testing. In: *TRANSDUCERS '91, Int. Conf. Solid-State Sens. Actuators Dig.* (IEEE, New York 1991) pp. 953–956

40.63 S.B. Brown, W. van Arsdell, C.L. Muhlstein: Materials reliability in MEMS devices. In: *TRANSDUCERS '97, Int. Conf. Solid-State Sens. Actuators Dig.*, Vol. 1 (IEEE, New York 1997) pp. 591–594

40.64 S.B. Brown, E. Jansen: Reliability and long term stability of MEMS. In: *Summer Top. Meet. Dig. Opt. MEMS Their Appl.* (IEEE, New York 1996) pp. 9–10

40.65 S.B. Brown, G. Povirk, J. Connally: Measurement of slow crack growth in silicon and nickel micromechanical devices. In: *Proc. Micro Electro Mech. Syst. Investig. Micro Struct. Sens. Actuators Mach. Syst.* (IEEE, New York 1993) pp. 99–102

40.66 C.L. Muhlstein, S.B. Brown, R.O. Ritchie: High cycle fatigue and durability of polycrystalline silicon thin films in ambient air, Sens. Actuators A **94**, 177–188 (2001)

40.67 H. Kapels, R. Aigner, J. Binder: Fracture strength and fatigue of polysilicon determined by a novel thermal actuator, Trans. Electron. Dev. **47**, 1522–1528 (2000)

40.68 H. Kahn, N. Tayebi, R. Ballerini, R.L. Mullen, A.H. Heuer: Fracture toughness of polysilicon MEMS devices, Sens. Actuators A **82**, 274–280 (2000)

40.69 J. Bagdahn, W.N. Sharpe Jr.: Fatigue of polysilicon silicon under long-term cyclic loading, Sens. Actuators A **103**, 9–15 (2003)

40.70 K.J. Gabriel, F. Behi, R. Mahadevan: In-situ friction and wear measurements in integrated polysilicon mechanisms, Sens. Actuators A **A21–A23**, 184–188 (1990)

40.71 M. Mehregany, S.D. Senturia, J.H. Lang: Friction and wear in microfabricated harmonic side-drive motors. In: *Tech. Dig. IEEE Solid-State Sens. Actuator Workshop, Hilton Head Island* (IEEE, New York 1990) pp. 17–22

40.72 D.M. Tanner, W.M. Miller, W.P. Eaton, L.W. Irwin, K.A. Peterson, M.T. Dugger, D.C. Senft, N.F. Smith, P. Tanyunyong, S.L. Miller: The effect of frequency on the lifetime of a surface micromachined microengine driving a load. In: *Int. Reliab. Phys. Symp. Proc. (IRPS)*, ed. by A.N. Campbell (IEEE, Reno 1998) pp. 26–35

40.73 D.M. Tanner: MEMS Reliability: Infrastructure, Test Structures, Experiments and Failure Modes, Sandia Rep. (Sandia National Laboratories, Livermore 2000) available from National Technical Information Service, US Department of Commerce, Springfield or http://www.sandia.gov/mstc/technologies/micromachines/tech-info/bibliography/biblog_char.html

40.74 A.D. Romig Jr., M.T. Dugger, P.J. McWorther: Materials issues in microelectromechanical devices: Science, engineering, manufacturability and reliability, Acta Mater. **51**, 5837–5866 (2003)

40.75 J.A. Ruan, B. Bhushan: Atomic-scale friction measurement using friction force microscopy. 1. Gen-

eral – Principles and new measurement techniques, J. Tribol. **116**, 378–388 (1994)
40.76 H. Lui, B. Bhushan: Adhesion and friction studies of microelectromechanical systems/nanoelectromechanical systems materials using a novel triboapparatus, J. Vac. Sci. Technol. A **21**, 1528–1538 (2003)
40.77 M.P. de Boer: A hinged-pad test structure for sliding friction measurement in micromachining, Proc. SPIE **3512**, 241–250 (1998)
40.78 S.L. Miller, J.J. Sniegowski, G. LaVigne, P.J. McWorther: Friction in surface micromachined microengines, Proc. SPIE **2722**, 197–204 (1996)
40.79 D.C. Senft, M.T. Dugger: Friction and wear in surface micromachined tribological test devices, Proc. SPIE **3224**, 31–38 (1997)
40.80 M.P. de Boer, T.M. Mayer: Tribology of MEMS, MSR Bulletin **26**, 302–304 (2001)
40.81 Z. Rymuza: Control tribological and mechanical properties of MEMS surfaces. Part 1: Critical review, Microsyst. Technol. **5**, 173–180 (1999)
40.82 Z. Rymuza, M. Misiak, L. Kuhn, K. Schmidt-Szalowski, Z. Ranek-Boroch: Control tribological and mechanical properties of MEMS surfaces. Part 2: Nanomechanical behavior of self-lubricating ultrathin films, Microsyst. Technol. **5**, 181–188 (1999)
40.83 J.G. Fleming, S.S. Mani, J.J. Sniegowski, R.S. Blewer: Tungsten coating for improved wear resistance and reliability of microelectromechanical devices, US Patent 6290859 (2001)
40.84 U. Beerschwinger, D. Mathieson, R.L. Reuben, S.J. Yang: A study of wear on MEMS contact morphologies, J. Micromech. Microeng. **4**, 95–105 (1994)
40.85 R. Bandorf, H. Lüthje, T. Staedler: Influencing factors on microtribology of DLC films for MEMS and microactuators, Diam. Relat. Mater. **13**, 1491–1493 (2004)
40.86 A.P. Musinho, R.D. Mansano, M. Massi, J.M. Jaramillo: Micro-machine fabrication using diamond-like carbon films, Diam. Relat. Mater. **12**, 1041–1044 (2003)
40.87 A. Erdemir: Superlubricity and wearless sliding in diamondlike carbon films, Proc. Mater. Res. Soc. **697**, 391–403 (2002)
40.88 A. Erdemir: Design criteria for superlubricity in carbon films and related microstructures, Tribol. Int. **37**, 577–583 (2004)
40.89 A.R. Krauss, O. Auciello, D.M. Gruen, A. Jayatissa, A. Sumant, J. Tucek, D.C. Mancini, N. Moldovan, A. Erdemir, D. Ersoy, M.N. Gardos, H.G. Busmann, E.M. Meyer, M.Q. Ding: Ultrananocrystalline diamond thin films for MEMS and moving mechanical assembly devices, Diam. Relat. Mater. **10**, 1952–1962 (2001)
40.90 W.R. Ashurst, M.B.J. Wijesundra, C. Carraro, R. Maboudian: Tribological impact of SiC encapsulation of released polycrystalline silicon microstructures, Tribol. Lett. **71**, 195–198 (2004)
40.91 S. Sundararajan, B. Bhushan: Micro/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear **217**, 251–261 (1998)
40.92 X. Li, B. Bhushan: Micro/nanotribological characterization of ceramic films for microdevices, Thin Solid Films **340**, 210–217 (1999)
40.93 R. Maboudian, W.R. Ashurst, C. Carraro: Tribological challenges in microelectromechanical systems, Tribol. Lett. **12**, 95–100 (2002)
40.94 H.A.C. Tilmans, H. Ziad, H. Jansen, O. Di Monaco, A. Jourdain, W. De Raedt, X. Rottenberg, E. De Backer, A. De Caussemaeker, K. Baert: Wafer-level packaged RF-MEMS switches fabricated in a CMOS fab. In: *Proc. IEDM 2001* (IEEE, Washington 2001) pp. 921–924
40.95 Military Standard (MIL-STD-883): Test Methods and Procedures for Microelectronics
40.96 A. Jourdain, P. De Moor, S. Pamidighantam, H.A.C. Tilmans: Investigation of the hermeticity of BCB-sealed cavities for housing (RF-)MEMS devices. In: *Proc. MEMS 2002* (IEEE, Las Vegas 2002) pp. 677–680
40.97 I. De Wolf, W.M. van Spengen, R. Modlinski, A. Jourdain, A. Witvrouw, P. Fiorini, H.A.C. Tilmans: Reliability and failure analysis of RF MEMS switches. In: *Proc. ISTFA 2002* (ASM, Phoenix 2002) pp. 275–281

41. Mechanical Properties of Micromachined Structures

Harold Kahn

To be able to accurately design structures and make reliability predictions, it is first necessary to know the mechanical properties of the materials that make up the structural components. In the fields of microelectromechanical systems (MEMS) and nanoelectromechanical systems (NEMS), the devices are necessarily very small. The processing techniques and microstructures of the materials in these devices may differ significantly from bulk structures. Also, the surface area to volume ratio in these structures is much higher than in bulk samples, and so the surface properties become much more important. In short, it cannot be assumed that mechanical properties measured using bulk specimens will apply to the same materials when used in MEMS and NEMS. This chapter will review the techniques that have been used to determine the mechanical properties of micromachined structures, especially residual stress, strength, and Young's modulus. The experimental measurements that have been performed will then be summarized, in particular the values obtained for polycrystalline silicon (polysilicon).

41.1 **Measuring Mechanical Properties of Films on Substrates** 1459
41.1.1 Residual Stress Measurements 1459
41.1.2 Mechanical Measurements Using Nanoindentation 1460
41.2 **Micromachined Structures for Measuring Mechanical Properties** . 1460
41.2.1 Passive Structures 1461
41.2.2 Active Structures 1464
41.3 **Measurements of Mechanical Properties** 1471
41.3.1 Mechanical Properties of Silicon and Polysilicon 1471
41.3.2 Mechanical Properties of Other Materials 1474
41.4 **Summary** 1475
References 1476

41.1 Measuring Mechanical Properties of Films on Substrates

In order to accurately determine the mechanical properties of very small structures, it is necessary to test specimens made from the same materials, processed in the same way, and of the same approximate size. Not surprisingly, it is often difficult to handle specimens this small. One solution is to test the properties of films remaining on substrates. Micro- and nanomachined structures are typically fabricated from films that are initially deposited onto a substrate and are subsequently patterned and etched into the appropriate shapes, and then finally released from the substrate. If the testing is performed on the continuous film, before patterning and release, the substrate can be used as an effective *handle* for the specimen (in this case, the film). Of course, since the films are adhered to the substrate, the types of tests that are possible are severely limited.

41.1.1 Residual Stress Measurements

One common measurement that is easily performed on films attached to substrates is that of residual film stress. The curvature of the substrate is measured prior to and after film deposition. Curvature can be measured in a number of ways. The most common technique is to scan a laser across the surface (or scan the substrate beneath the laser) and detect the angle of the reflected signal. Alternatively, profilometry, optical interferometry, or even atomic force microscopy (AFM) can be used. As expected, tools that map a surface or perform multiple linear scans can give more accurate readings than tools that measure only a single scan.

Assuming that the film is thin compared to the substrate, the average residual stress in the film, σ_f, is given

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_41

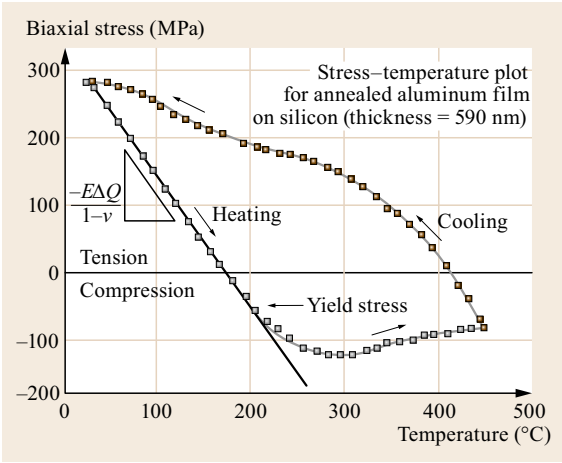


Fig. 41.1 Typical results for residual stress as a function of temperature for an aluminum film on a silicon substrate (after [41.1]). The stresses were determined by measuring the curvature of the substrate before and after film deposition, using the reflected signal of a laser scanned across the substrate surface

by the Stoney equation,

$$\sigma_f = \frac{1}{6}\frac{E_s}{(1-\nu_s)}\frac{t_s^2}{t_f}\left(\frac{1}{R_1}-\frac{1}{R_2}\right), \qquad (41.1)$$

where the subscripts f and s refer to the film and substrate, respectively; t is thickness, E is Young's modulus, ν is Poisson's ratio, and R is radius of curvature before (R_1) and after (R_2) film deposition [41.2]. For the typical (100)-oriented silicon substrate, $E/(1-\nu)$ (also known as the biaxial modulus) is equal to 180.5 GPa, independent of in-plane rotation [41.3]. This investigation can be performed on the as-deposited film or after any subsequent annealing step, provided no changes occur to the substrate.

This measurement will reveal the average residual stress of the film. Typically however, the residual stresses of deposited films will vary through the film thickness. One way to detect this, using substrate curvature techniques, is to etch away a fraction of the film and repeat the curvature measurement. This can be iterated any number of times to obtain a residual stress profile for the film [41.4]. Alternatively, tools have been designed to measure the substrate curvature during the deposition process itself, to obtain information on how the stresses evolve [41.5].

An additional feature of some of these tools is the ability to heat the substrates while performing the stress measurement. An example of the results obtained in such an experiment is shown in Fig. 41.1 [41.1], for an aluminum film on a silicon substrate. The slope of the heating curve gives the difference in thermal expansion between the film and the substrate. When the heating curve changes slope and becomes nearly horizontal, this indicates that the yield strength of the film has been reached.

41.1.2 Mechanical Measurements Using Nanoindentation

Besides residual stress, it is difficult to measure mechanical properties of films adhered to substrates, without the measurement being affected by the presence of the substrate. Recent developments in nanoindentation equipment have allowed this technique to be used in some cases. With specially designed tools, indentation can be performed using very low loads. If the films being investigated are thick enough and rigid enough, measurements can be made that are not influenced by the presence of the substrate. This can be verified by depositing the same film onto different substrates. By continuously monitoring the displacement as well as the load during indentation, a variety of properties can be measured, including hardness and Young's modulus [41.6]. This area is covered in more detail in a separate chapter.

For brittle materials, cracks can be generated by indentation, and strength information can be gathered. However, the exact stress fields created during the indentation process are not exactly known, and therefore, quantitative values for strength are difficult to determine. Anisotropic etching of single-crystal silicon has been done to create 30 μm tall structures which were then indented to examine fracture toughness [41.7], but this is not possible with most materials.

41.2 Micromachined Structures for Measuring Mechanical Properties

Certainly, the most direct way to measure the mechanical properties of small structures is to fabricate structures that would be conducive to such tests. Fabrication techniques are sufficiently advanced that virtually any design can be realized, at least in two dimensions. There are two basic types of devices used for mechanical property testing: *passive* structures and *active* structures.

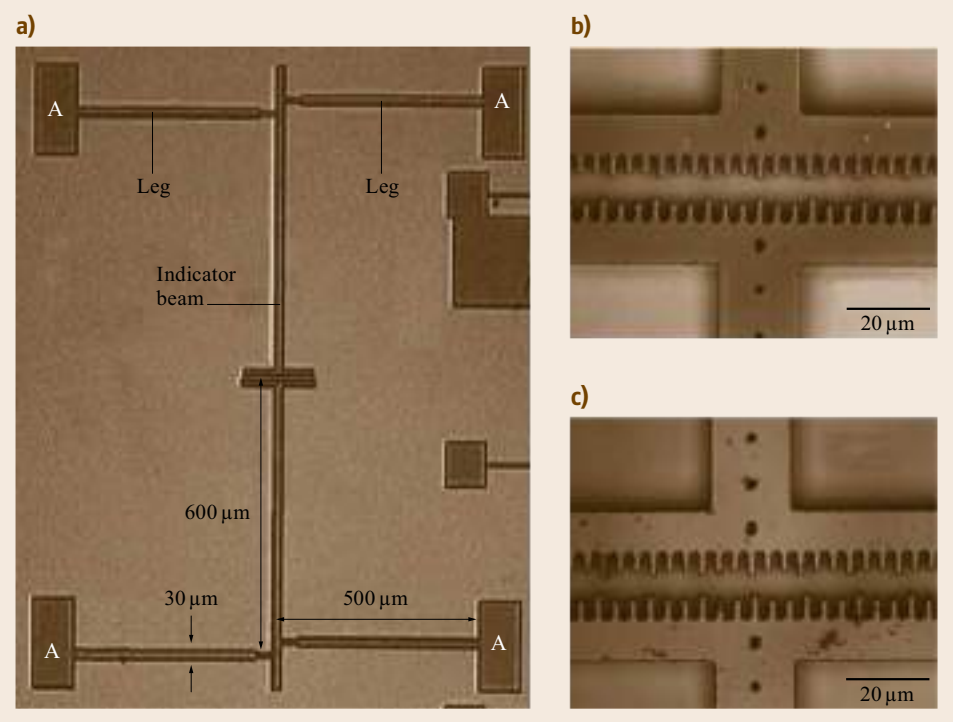


Fig. 41.2 (a) Microstrain gauge fabricated from polysilicon. (b) shows a close-up of the Vernier scale before release, and (c) shows the same area after release

41.2.1 Passive Structures

As mentioned previously, the main difficulty in testing very small specimens is in handling. One way to circumvent this problem is by using passive structures. These structures are designed to act as soon as they are released from the substrate and to provide whatever information they are designed to supply without further manipulation. For all of these passive structures, the forces acting on them come from the structural materials' own internal residual stresses. For devices on the micron scale or smaller, gravitational forces can be neglected, and therefore internal stresses are the only source of actuation force.

Stress Measurements

Since the internal residual stresses act upon the passive devices when they are released, it is natural to design a device that can be used to measure residual stresses. One such device, a rotating microstrain gauge, is shown in Fig. 41.2. There are many different microstrain gauge designs, but they all operate under the same principle. In Fig. 41.2a, the large pads, labeled A, will remain anchored to the substrate when the rest of the device is released. Upon release, the device will expand or contract in order to relieve its internal residual stresses. A structure under tension will contract, and a structure under compression will expand. For the structure in Fig. 41.2, a compressive stress will cause the legs to lengthen. Since the two opposing legs are not attached to the central beam at the same point, but are offset, when these legs expand, they will cause the central beam to rotate. The device in Fig. 41.2 contains two independent gauges that point at one another. At the ends of the two central beams are two parts of a Vernier scale. By observing this scale, the rotation of the beams can be measured.

If the connections between the legs and the central beams were simple pin connections, the strain ε of the legs (the fraction of expansion or contraction) could be determined simply by the measured rotation and the geometry of the device, namely

$$\varepsilon = \frac{(d_{\text{beam}} d_{\text{offset}})}{(2L_{\text{central}} L_{\text{leg}})} , \tag{41.2}$$

where d_{beam} is the lateral deflection of the end of one central beam, d_{offset} is the distance between the connections of the opposing legs, L_{central} is the length of the central beam (measured to the center point between the leg connections), and L_{leg} is the length of the leg. However, since the entire device was fabricated from a single polysilicon film, this cannot be the case; there must be some bending occurring at the connections. As a result, to get an accurate determination of the strain relieved upon release, finite element analysis (FEA) of

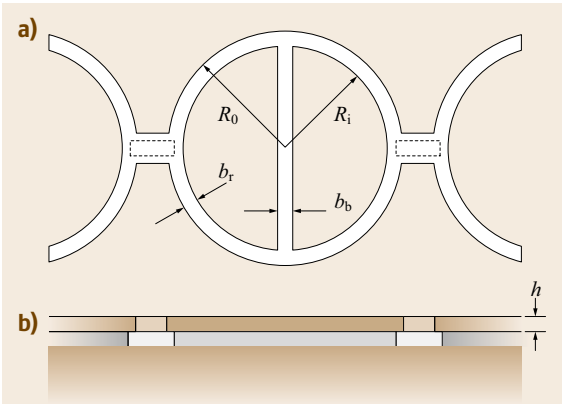


Fig. 41.3a,b Schematic **(a)** top view and **(b)** side view of Guckel ring structures (after [41.8]). The *dashed lines* in **(a)** indicate the anchors

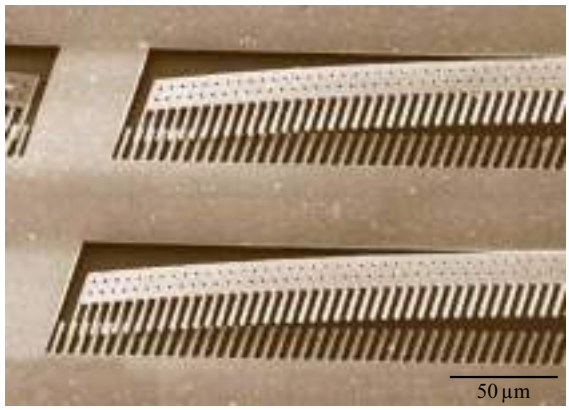


Fig. 41.4 Scanning electron micrograph (SEM) of a portion of a silicon microactuator. Residual stress gradients in the silicon cause the structure to bend

the structure must be performed. This is a common situation for microdevices. FEA is a powerful tool for determining the displacements and stresses of nonideal geometries. One of the drawbacks is that Young's modulus of the material must be known in order to do the FEA, as well as to convert the measured strain into a stress value. However, Young's moduli for many micromachined materials are known, or can be measured using other techniques.

Besides rotating strain gauges there are other devices designed to measure residual stresses. One of the most simple is a doubly clamped beam, that is, a long, narrow beam of constant width and thickness which is anchored to the substrate at both ends. If the beam contains a tensile stress, it will remain straight. However, if the beam contains a compressive stress, it will buckle if its length exceeds a critical value l_{cr} according to the Euler buckling criterion [41.8],

$$\varepsilon_r = -\frac{\pi^2}{3}\left(\frac{h}{l_{cr}}\right)^2 , \qquad (41.3)$$

where ε_r is the residual strain in the beam, and h is the width or thickness of the beam, whichever is less. To determine the residual strain, a series of doubly clamped beams are fabricated of varying lengths. In this way, the critical length l_{cr} for buckling can be deduced after release. One problem with this technique is that during the release process, any turbulence in the solution will lead to enhanced buckling of beams, and a low value for l_{cr} will be obtained.

For films with tensile stresses, a similar analysis can be done using ring-and-beam structures, also called Guckel rings after their inventor, Henry Guckel. A schematic of this design is shown in Fig. 41.3 [41.8].

A tensile stress in the outer ring will cause it to contract. This will lead to a compressive stress in the central beam, even though the material was originally tensile before release. The amount of compression in the central beam can be determined analytically from the geometry of the device and the residual strain of the material. Again, by changing the length of the central beam, the l_{cr} will be determined, and then the residual strain can be deduced.

Stress Gradient Measurements

For structures fabricated from thin deposited films, the stress gradient can be just as important as the stress itself. Figure 41.4 shows a portion of a silicon microactuator. The device is designed to be completely planar. However, stress gradients in the film cause the structures to bend. This figure illustrates the importance of characterizing and controlling stress gradients, and it also demonstrates that the easiest structure for measuring stress gradients is a simple cantilever beam. By measuring the end deflection δ, of a cantilever beam of length l, and thickness t, the stress gradient, $d\sigma/dt$ is determined by [41.9]

$$\frac{d\sigma}{dt} = \frac{2\delta}{l^2}\frac{E}{(1-\nu)} . \qquad (41.4)$$

The magnitude of the end deflection can be measured by microscopy, optical interferometry, or any other technique.

Another useful structure for measuring stress gradients is a spiral. For this structure, the end of the spiral that is not anchored to the substrate will move out-of-plane. Also, the diameter of the spiral will contract, and the free end of the spiral will rotate when released [41.10].

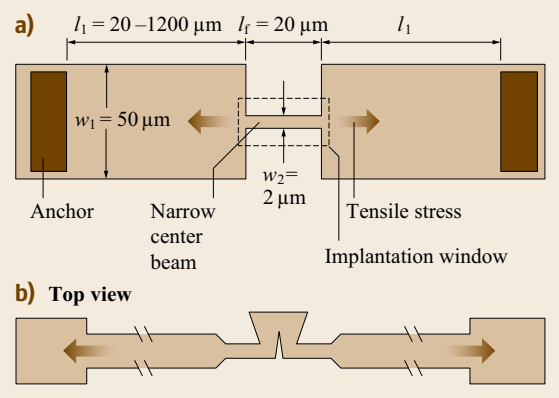


Fig. 41.5a,b Schematic designs of doubly clamped beams with stress concentrations for measuring strength. **(a)** was fabricated from polysilicon (after [41.11]) and **(b)** was fabricated from Si_xN_y (after [41.12])

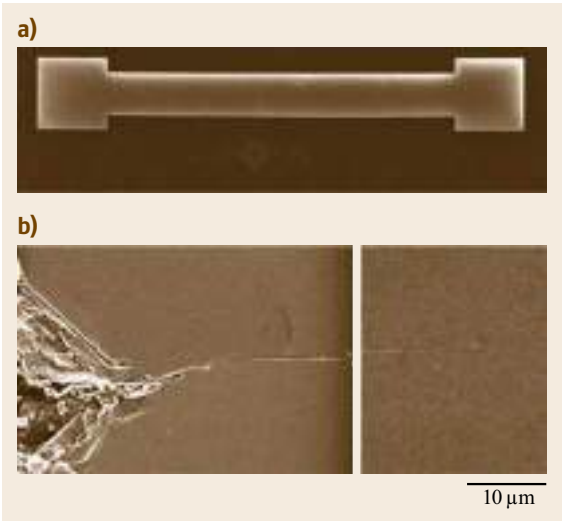


Fig. 41.6 **(a)** SEM of a 500 μm long polysilicon beam with a Vickers indent placed near its center; **(b)** higher magnification SEM of the area near the indent showing the precrack traveling from the substrate into the beam (after [41.14])

Strength and Fracture Toughness Measurements

As mentioned above, if a doubly clamped beam contains a tensile stress, it will remain taut when released, because it cannot relieve any of its stress by contracting. This tensile stress can be thought of as a tensile load that is being applied at the ends of the beam. If this tensile load exceeds the tensile strength of the material, the beam will break. Therefore, since the tensile stress can be measured, as discussed above, this technique can be used to gather information on the strength of materials. Figure 41.5 shows two different beam designs that have been used to measure strength. The device shown in Fig. 41.5a was fabricated from a tensile polysilicon film [41.11]. Different beams were designed with varying lengths of the wider regions (marked l_1 in the figure). In this manner, the load applied to the narrow center beam was varied, even though the entire film contained a uniform residual tensile stress. For l_1 greater than a critical value, the narrow center beam fractured, giving a measurement for the tensile strength of polysilicon.

The design shown in Fig. 41.5b was fabricated from a tensile Si_xN_y film [41.12]. As seen in the figure, a stress concentration was designed into the beam, to ensure the fracture strength would be exceeded. In this case, a notch was etched into one side of the beam. Since the stress concentration is not symmetric with regard to the beam axis, this results in a large bending moment at that position, and the test measures the bend strength of the material. Again, like the beams shown in Fig. 41.5a, the geometry of various beams fabricated from the same film were varied, to vary the maximum stress seen at the notch. By seeing which beams fracture at the stress concentration after release, the strength can be determined.

Through a similar technique, but by using an atomically sharp precrack instead of a stress concentration, the fracture toughness of a material can be determined. Sharp precracks can be introduced into micromachined structures before release by placing a Vickers indent on the substrate, near the device; the radial crack formed by the indent will propagate into the overlying structure [41.13]. Accordingly, the beam with a sharp precrack shown in Fig. 41.6, was fabricated using polysilicon [41.14]. Because of the stochastic nature of indentation, the initial precrack length will vary from beam to beam. Because of this, even though the geometry of the beams remains identical, the stress intensity, K, at the precrack tip will vary. Upon release, only those precracks whose K exceeds the fracture toughness of the material, K_{Ic}, will propagate, and in this way upper and lower bounds for K_{Ic} for the material can be determined.

For materials that do not normally contain residual tensile stresses, composite beams can be fabricated. In this technique, a portion of the beam length is fabricated from a highly tensile material, and the rest of the beam from the material to be tested. The two materials must exhibit perfect adhesion. When released, the tensile material will tend to contract, putting the rest of the beam in tension. This has been demonstrated for evaporated Al films with highly tensile CVD (chemical vapor deposition) SiN [41.15, 16].

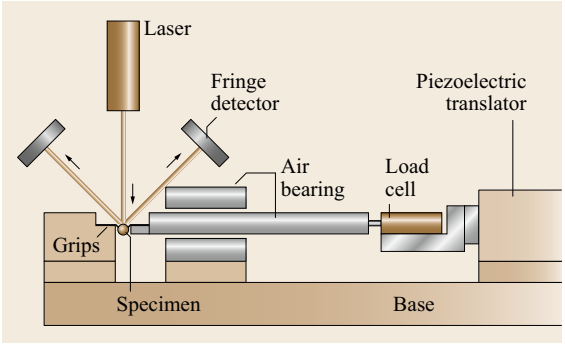


Fig. 41.7 Schematic of a measurement system for tensile loading of micromachined specimens (after [41.17])

For all of the beams discussed in this section, finite element analysis is required to determine the stress concentrations and stress intensities. Even though approximate analytical solutions may exist for these designs, the actual fabricated structure will not have idealized geometries. For example, corners will never be perfectly sharp, and cracks will never be perfectly straight. This reinforces the idea that FEA is a powerful tool for determining mechanical properties of very small structures.

41.2.2 Active Structures

As discussed above, it is very convenient to design structures that act upon release to provide information on the mechanical properties of the structural materials. However, this is not always possible. For example, those passive devices just discussed rely on residual stresses to create the changes (rotation or fracture) that occur upon release, but many materials do not contain high residual stresses as-deposited, or else the processing scheme of the device precludes the generation of residual stresses. Also, some mechanical properties, such as fatigue resistance, require motion in order to be studied. Therefore, active devices are used. These are acted upon by a force (the source of this force can be integrated into the device itself, or can be external to the device) in order to create a change, and by the response to the force, the mechanical properties are studied.

Young's Modulus Measurements

Young's modulus E is a material property that is critical to any structural device design. It describes the elastic response of a material and relates stress σ and strain ε by

$$\sigma = E\varepsilon \,. \tag{41.5}$$

In bulk samples, E is often measured by loading a specimen in tension and measuring displacement as a function of stress for a given length. While this is considerably more difficult for small structures, such as those fabricated from thin deposited films, with careful experimental techniques, it can be achieved. Figure 41.7 shows the schematic of one such measurement system [41.17]. The fringe detectors in the figure detect the reflected laser signal from two gold lines that are deposited onto the polysilicon specimen, which act as gauge markers. In this manner, the strain in the specimen during loading can be monitored. Besides gold lines, Vickers indents placed in a nickel specimen can also serve as gauge markers [41.19], or else a speckle interferometry technique [41.20] can be used to determine strain in the specimen. Once the stress versus strain behavior is measured, the slope of the curve is equal to E. By use of a constant load, such as a dead weight, and resistive heating, high temperature creep can also be investigated with this method [41.21].

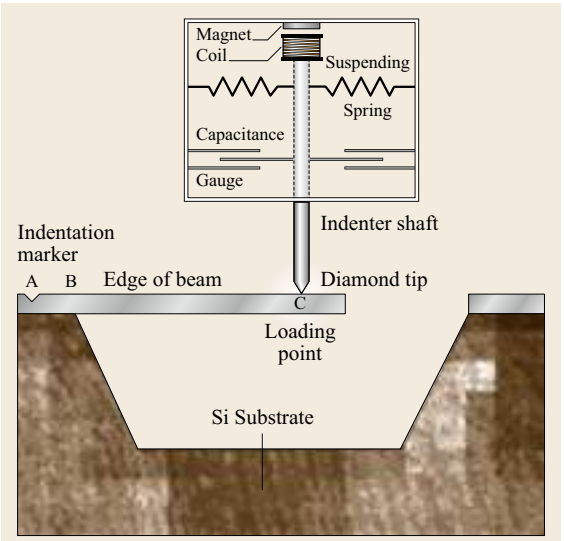


Fig. 41.8 Schematic of a nanoindenter loading mechanism pushing on the end of a cantilever beam (after [41.18])

Besides the tensile test, Young's modulus can be determined by other measures of stress–strain behavior. As seen in Fig. 41.8, a cantilever beam can be bent by pushing on the free end with a nanoindenter [41.18] or profilometer [41.22] or AFM probe [41.23]. The nanoindenter can monitor the force applied and the displacement, and simple beam theory can convert the displacement into strain, to obtain E. A similar technique, shown schematically in Fig. 41.9 [41.24], involves pulling downwards on a cantilever beam by means of an electrostatic force. An electrode is fabricated into the substrate beneath the cantilever beam, and a voltage is applied between the beam and the bottom electrode. The force acting on the beam is equal

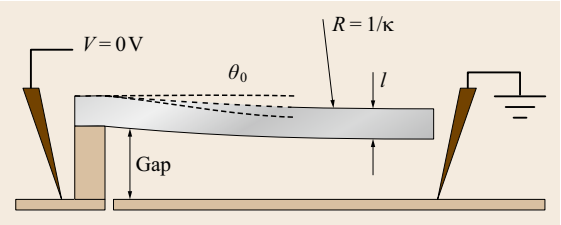


Fig. 41.9 Schematic of a cantilever beam bending test using an electrostatic voltage to pull the beam toward the substrate (after [41.24])

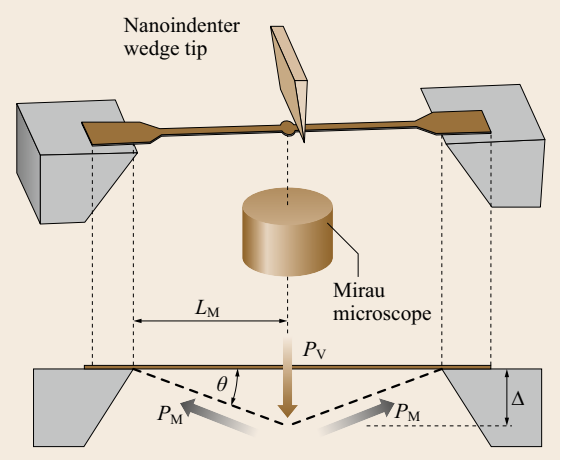


Fig. 41.10 Schematic of an externally loaded doubly clamped beam (after [41.25])

to the electrostatic force corrected to include the effects of fringing fields acting on the sides of the beam, namely

$$F(x) = \frac{\varepsilon_0}{2}\left(\frac{V}{g+z(x)}\right)^2\left(1+\frac{0.65[g+z(x)]}{w}\right), \tag{41.6}$$

where $F(x)$ is the electrostatic force at x, ε_0 is the dielectric constant of air, g is the gap between the beam and the bottom electrode, $z(x)$ is the out-of-plane deflection of the beam, w is the beam width, and V is the applied voltage [41.24]. In this work, the deflection of the beam as a function of position is measured using optical interferometry. These measurements combine to give stress–strain behavior for the cantilever beam. An extension of this technique uses doubly clamped beams instead of cantilever beams. In this case, the deflection of the beam at a given electrostatic force depends on the residual stress in the material as well as Young's modulus. Therefore, this method can also be used to measure residual stresses in doubly clamped beams.

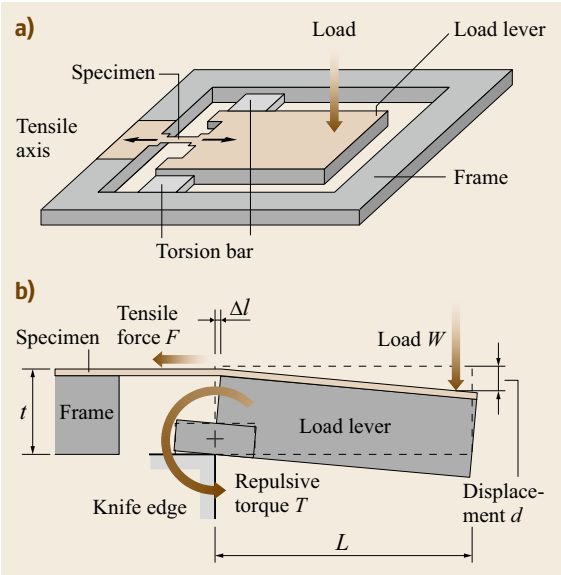


Fig. 41.11a,b Schematic of a device that transfers external vertical loads to in-plane tensile loads in deposited films (after [41.27])

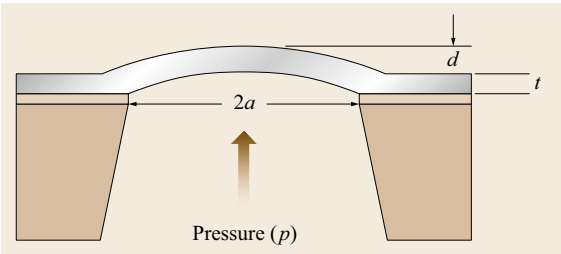


Fig. 41.12 Schematic cross section of a microfabricated membrane (after [41.28])

Similarly, an external load can be applied to the center of the doubly clamped beam, instead of using an electrostatic force, as shown in Fig. 41.10 [41.25]. The beam deflection is monitored using optical interferometry. If a profilometer is used to generate the downward force, the deflection can be measured along the entire length [41.26]. Yet another variation is illustrated in Fig. 41.11 [41.27]. In this technique, the externally applied downward force is transferred to a film of any material deposited onto the device.

Another device that is possible to fabricate from a thin film and also to use for investigating stress–strain behavior is a suspended membrane, seen in Fig. 41.12 [41.28]. As shown in the schematic figure, the membrane is exposed to an elevated pressure on one side, causing it to bulge out in the opposite direction. The deflection of the membrane is measured by optical or other techniques and related to the strain in the mem-

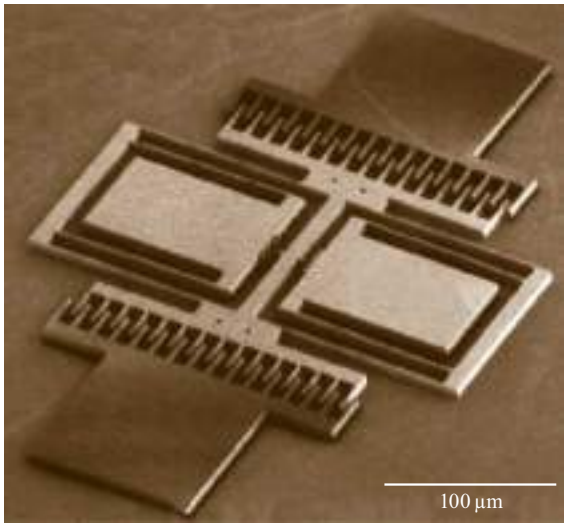


Fig. 41.13 SEM of a polysilicon lateral resonator (after [41.32])

brane. These membranes can be fabricated in any shape, typically square or circular. Both analytical solutions and finite element analyses have been done to relate the deflection to the strain. Like the doubly clamped beams, both Young's modulus and residual stress play a role in the deflected shape. Therefore, both of these mechanical properties can be determined by the pressure versus deflection performance of the membrane.

Besides stress–strain behavior, another measurement that can reveal Young's modulus of a material is to determine the natural resonance frequency. For a cantilever, the resonance frequency f_r for free undamped vibration is given by

$$f_r = \frac{\lambda_i^2 t}{4\pi l^2}\left(\frac{E}{3\rho}\right)^{1/2}, \tag{41.7}$$

where ρ, l, and t are density, length, and thickness of the cantilever; λ_i is the eigenvalue, where i is an integer that describes the resonance mode number; for the first mode $\lambda_1 = 1.875$ [41.29]. Therefore, by measuring f_r and knowing the geometry and density, E can be determined. The cantilever can be vibrated by a number of techniques, including a laser, loudspeaker, or piezoelectric shaker [41.30, 31]. The frequency that produces the highest amplitude of vibration is the resonance frequency.

A micromachined device that uses an electrostatic comb drive and an AC signal to generate the vibration of the structure is known as a lateral resonator [41.33]. One example is shown in Fig. 41.13 [41.32]. When a voltage is applied across either set of interdigitated comb fingers shown in Fig. 41.13, an electrostatic attraction will be generated due to the increase in capacitance as the overlap between the comb fingers increases. The force F generated by the comb drive is given by

$$F = \frac{1}{2}\frac{\partial C}{\partial x}V^2 = n\varepsilon\frac{h}{g}V^2, \tag{41.8}$$

where C is capacitance, x is distance traveled by one comb-drive toward the other, n is the number of pairs of comb fingers in one drive, ε is the permittivity of the fluid between the fingers, h is the height of the fingers, g is the gap spacing between the fingers, and V is the applied voltage [41.33]. When an AC voltage at the resonance frequency is applied across either of the two comb drives, the central portion of the device will vibrate. In fact, since force depends on the square of the voltage for electrostatic actuation, for a time t, dependent drive voltage $v_D(t)$ of

$$v_D(t) = V_P + v_d\sin(\omega t), \tag{41.9}$$

where V_P is the DC bias and v_d is the AC drive amplitude, the time-dependent portion of the force will scale with [41.33]

$$2\omega V_P v_d\cos(\omega t) + \omega v_d^2\sin(2\omega t). \tag{41.10}$$

Therefore, if an AC drive signal is used with no DC bias, at resonance, the frequency of the AC drive signal will be one half the resonance frequency. For this device, the resonance frequency f_r will be

$$f_r = \frac{1}{(2\pi)}\left(\frac{k_{sys}}{M}\right)^{1/2}, \tag{41.11}$$

where k_{sys} is the spring constant of the support beams and M is the mass of the portion of the device that vibrates. The spring constant is given by

$$k_{sys} = 24\frac{EI}{L^3}, \tag{41.12}$$

$$I = \frac{hw^3}{12}, \tag{41.13}$$

where I is the moment of inertia of the beams, and L, h, and w are the length, thickness, and width of the beams. Therefore, by combining these equations and measuring f_r, E can be determined.

One distinct advantage of the lateral resonator technique and the electrostatically pulled cantilever technique for measuring Young's modulus is that they require no external loading sources. Portions of the devices are electrically contacted, and a voltage is applied. For the pure tension tests, such as shown in Fig. 41.7,

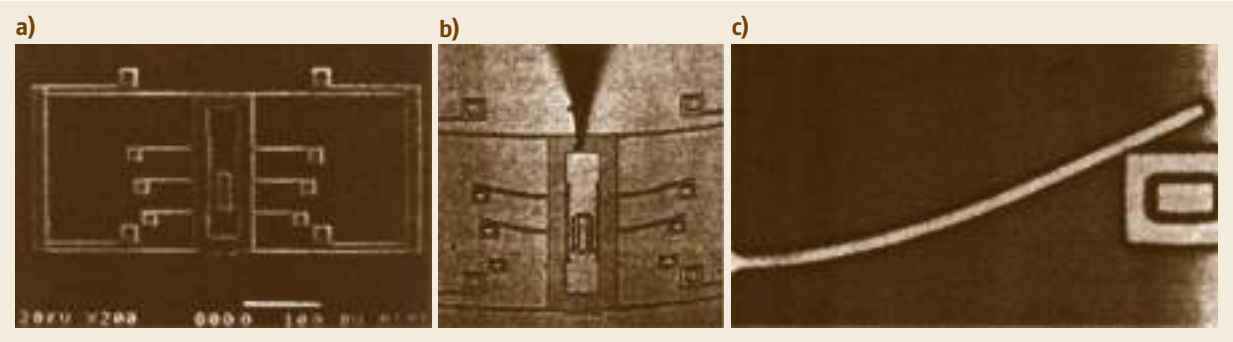


Fig. 41.14 **(a)** SEM of a device for measuring bend strength of polysilicon beams; **(b)** image of a test in process; **(c)** higher magnification view of one beam shortly before breaking (after [41.34])

the specimen must be attached to a loading system, and for the very small specimens discussed here, that can be extremely difficult and any misalignment or eccentricity in the test could lead to unreliable results. However, the advantage of the externally loaded technique is that there are no limitations on the type of materials that can be tested. Conductivity is not a requirement, nor is any compatibility with electrical actuation.

Strength and Fracture Toughness Measurements

As might be expected, any of the techniques discussed in the previous section that strain specimens in order to measure Young's modulus, can also be used to measure fracture strength. Simply, the load is increased until the specimen breaks. As long as either the load or the strain is measured at fracture, and the geometry of the specimen is known, the maximum stress required for fracture σ_{crit} can be determined, either through analytical analysis or FEA. Depending on the geometry of the test, σ_{crit} will represent the tensile or bend strength of the material.

If the available force is limited, or if it is desired to have a localized fracture site, stress concentrations can be added to the specimens. These are typically notches that are micromachined into the edges of specimens. Focused ion beams have also been used to carve stress concentrations into fracture specimens.

All of the external loading schemes, such as those shown in Figs. 41.7 and 41.8, have been used to measure fracture strength. Also, the electrostatically loaded doubly clamped beams can be pulled until they fracture. In this case, there is one complication. The electrostatic force is inversely proportional to the distance between the electrodes, and at a certain voltage, called the *pull-in voltage*, the attraction between the beam and the substrate will become so great that the beam will immediately be pulled into contact with the bottom electrode. As long as the fracture takes place before the pull-in voltage is reached, the experiment will give valid results.

There are other loading techniques that have been used to generate fracture of microspecimens. Figure 41.14 [41.34] shows one device which is designed to be pushed by the end of a micromanipulated needle. The long beams that extend from the sides of the central shuttle come into contact with anchored posts, and at a critical degree of bending, the beams will break off. Since the applied force cannot be measured in this technique, the experiment is optically monitored continuously during the test, and the image of the beams just before fracture is analyzed to determine σ_{crit}.

Another loading scheme that has been demonstrated for micromachined specimens utilizes scratch drive actuators to load the specimens [41.35]. These types of actuators work like inchworms, traveling across a substrate in discrete advances as an electrostatic force is repeatedly applied between the actuator and the substrate. The stepping motion can be made in the nm scale, depending on the frequency of the applied voltage, and so it can be an acceptable approximation to continuous loading. One advantage of this scheme is that very large forces can be generated by relatively small devices. The exact forces generated cannot be measured, so like the technique that used micromanipulated pushing, the test is continuously observed to determine the strain at fracture. Another advantage of this technique is that, like the lateral resonator and the electrically pulled cantilever, the loading takes place on-chip, and therefore, the difficulties associated with attaching and aligning an external loading source are eliminated.

Another on-chip actuator that is used to load microspecimens is shown in Fig. 41.15, along with three different microspecimens [41.13]. Devices have been fabricated with each of the three microspecimens integrated with the same electrostatic comb-drive actuator. In all three cases, when a DC voltage is applied to the

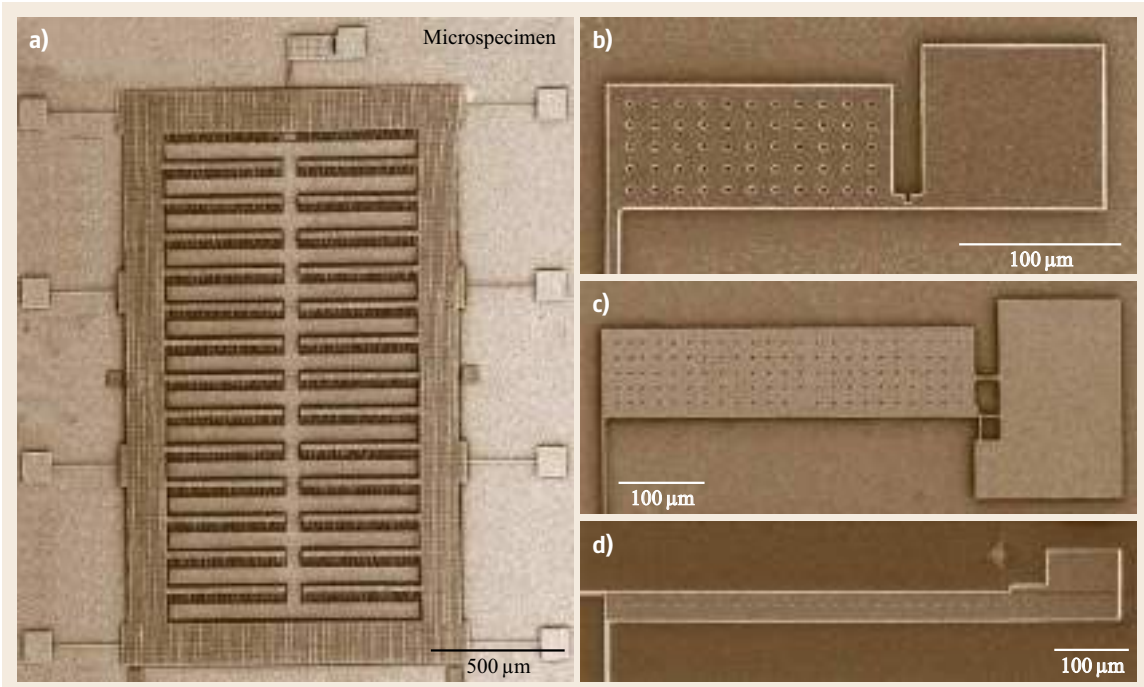


Fig. 41.15 **(a)** SEM of a micromachined device for conducting strength tests; the device consists of a large comb drive electrostatic actuator integrated with a microspecimen; **(b–d)** SEMs of various microspecimens for testing bend strength, tensile strength, and fracture toughness, respectively (after [41.13])

actuator it moves downward, as oriented in Fig. 41.15. This pulls down on the left end of each of the three microspecimens, which are anchored on the right. The actuator contains 1486 pairs of comb fingers. The maximum voltage that can be applied is limited by the breakdown voltage of the medium in which the test takes place. In air, that limits the voltage to less than 200 V. As a result, given a finger height of 4 µm and a gap of 2 µm, and using (41.8), the maximum force generated by this actuator is limited to about 1 mN. Standard optical photolithography has a minimum feature size of about 2 µm. As a result, the electrostatic actuator cannot generate sufficient force to perform a standard tensile test on MEMS structural materials such as polysilicon. Therefore, the microspecimens shown in Fig. 41.15 are designed such that the stress is amplified.

The specimen shown in Fig. 41.15b is designed to measure bend strength. It contains a micromachined notch with a root radius of 1 µm. When the actuator pulls downward on the left end of this specimen, the notch serve as a stress concentration, and when the stress at the notch root exceeds σ_{crit}, the specimen fractures. The specimen in Fig. 41.15c is designed to test tensile strength. When the left end of this specimen is pulled downward, a tensile stress is generated in the upper thin horizontal beam near the right end of the specimen. As the actuator continues to move downward, the tensile stress in this beam will exceed the tensile strength, causing fracture. Finally, the specimen in Fig. 41.15d is similar to that in Fig. 41.15b, except that the notch is replaced by a sharp precrack that was produced by the Vickers indent placed on the substrate near the specimen. When this specimen is loaded, a stress intensity K is generated at the crack tip. When the stress intensity exceeds a critical value K_{Ic} the crack propagates. K_{Ic} is also referred to as the fracture toughness.

The force generated by the electrostatic actuator can be calculated using (41.8). However, (41.8) assumes a perfectly planar, two-dimensional device. In fact, when actuated, the electric fields extend out of the plane of the device, and so (41.8) is just an approximation. Instead, similar to many of the techniques discussed in this section, the test is continuously monitored and the actuator displacement at the time of fracture is recorded. Then, FEA is used to determine the magnitude of the stress or stress intensity seen by the specimen at the point of fracture. To generate sufficient force to conduct tensile tests, a similar device to that shown in Fig. 41.15 has been designed which uses an array of parallel plate capacitors to provide the force,

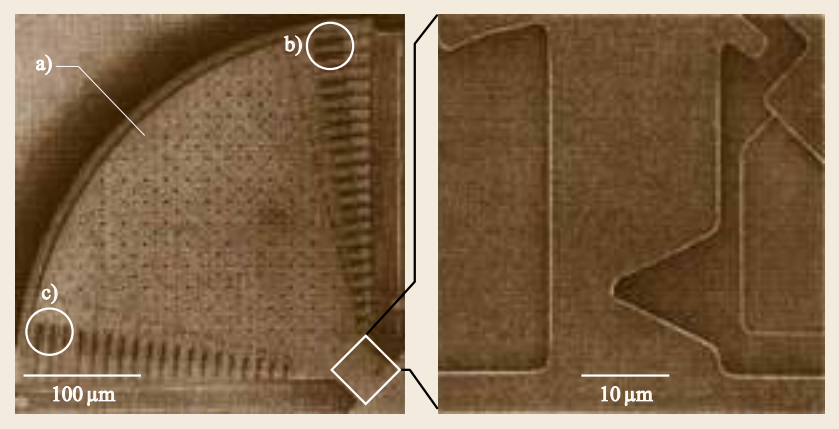


Fig. 41.16 SEM of a device for investigating fatigue; the image *on the right* is a higher magnification view of the notch near the base of the moving part of the structure. (a) Mass, (b) comb-drive actuator, (c) capacitive displacement sensor (after [41.36])

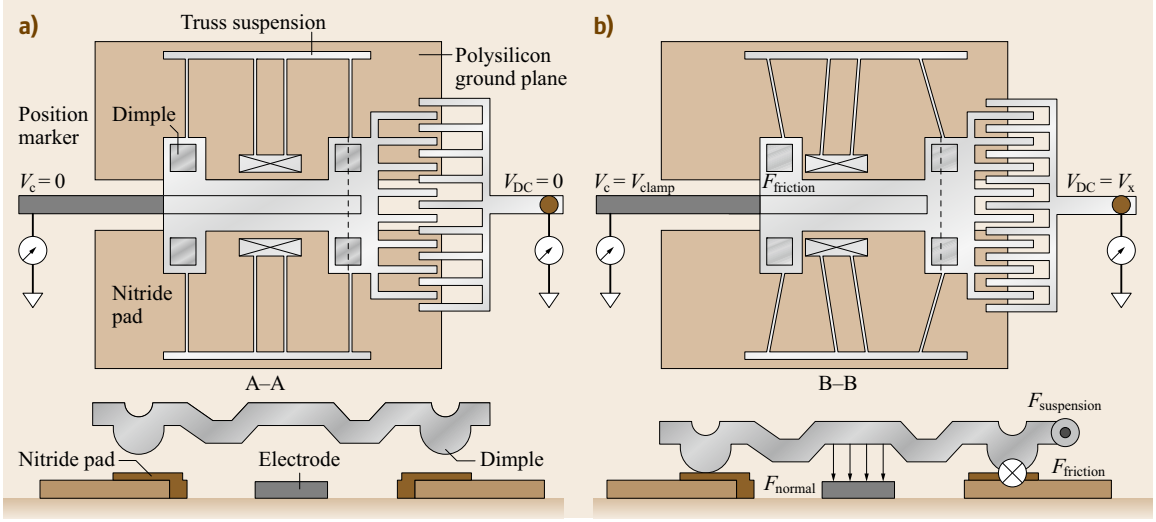


Fig. 41.17a,b Schematic drawings of a device used to study friction. **(a)** Shows top and side views of the device in its original position, and **(b)** shows views of the device after it has been displaced using the comb drive, and clamped using the substrate electrode (after [41.37])

instead of comb-drives [41.38]. In this way the available force is increased, but the maximum stroke is severely limited. Similar devices have also been designed for in situ testing in the SEM or TEM (transmission electron microscope) [41.39, 40].

Fatigue Measurements

A benefit of the electrostatic actuator shown in Fig. 41.15 is that, besides monotonic loading, it can generate cyclic loading. This allows the fatigue resistance of materials to be studied. Simply by using an AC signal instead of a DC voltage, the device can be driven at its resonant frequency. The amplitude of the resonance depends on the magnitude of the AC signal. This amplitude can be increased until the specimen breaks; this will investigate the low-cycle fatigue resistance. Otherwise, the amplitude of resonance can be left constant at a level below that required for fast fracture, and the device will resonate indefinitely until the specimen breaks; this will investigate high-cycle fatigue. It should be noted that the resonance frequency of such a device is about 10 kHz. Therefore, it is possible to stress a specimen over 10^9 cycles in less than a day. In addition to simple cyclic loading, if a DC bias is added to the AC signal, a mean stress can be superimposed on the cyclic load. In this way, nonsymmetric cyclic loading (either with a large tensile stress alternating with a small compressive stress, or vice versa) can be studied.

Another device that can be used to investigate fatigue resistance in MEMS materials is shown in Fig. 41.16 [41.36]. In this case, a large mass is attached

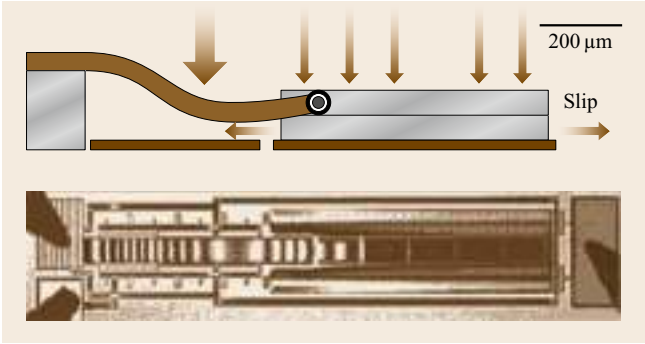


Fig. 41.18 Schematic cross-section and top-view optical micrograph of a hinged-cantilever test structure for measuring friction in micromachined devices (after [41.41])

to the end of a notched cantilever beam. The mass contains two comb drives on opposite ends. When an AC signal is applied to one comb drive, the device will resonate, cyclically loading the notch. The comb drive on the opposite side is used as a capacitive displacement sensor. This device contains many fewer comb fingers than the device shown in Fig. 41.15. As a result, it can apply cyclic loads by exploiting the resonance frequency of the device, but cannot supply sufficient force to achieve monotonic loading.

Fatigue loading has also been studied using the same external loading techniques shown in Fig. 41.7. In this case, the frequency of the cyclic load is considerably lower, since the resonance frequency of the device is not being utilized. This leads to longer high-cycle testing times. However, since the force is essentially unlimited, this technique allows a variety of frequencies to be studied, to determine their effect on the fatigue behavior.

Friction and Wear Measurements

Friction is another property that has been studied in micromachined structures. To study friction, of course, two surfaces must be brought into contact with each other. This is usually avoided at all costs for these devices because of the risk of stiction (stiction is the term used when two surfaces that come into contact adhere so strongly that the surfaces cannot be separated.) Even so, there have been a few devices designed to investigate friction. One of these is shown schematically in Fig. 41.17 [41.37]. It consists of a movable structure with a comb drive on one end and a cantilever beam on the other. Beneath the cantilever, on the substrate, is a planar electrode. The device is moved to one side using the comb drive. Then a voltage is applied between the cantilever beam and the substrate electrode. The voltage on the comb drive is then released. The device would normally return to its original position, to relax the

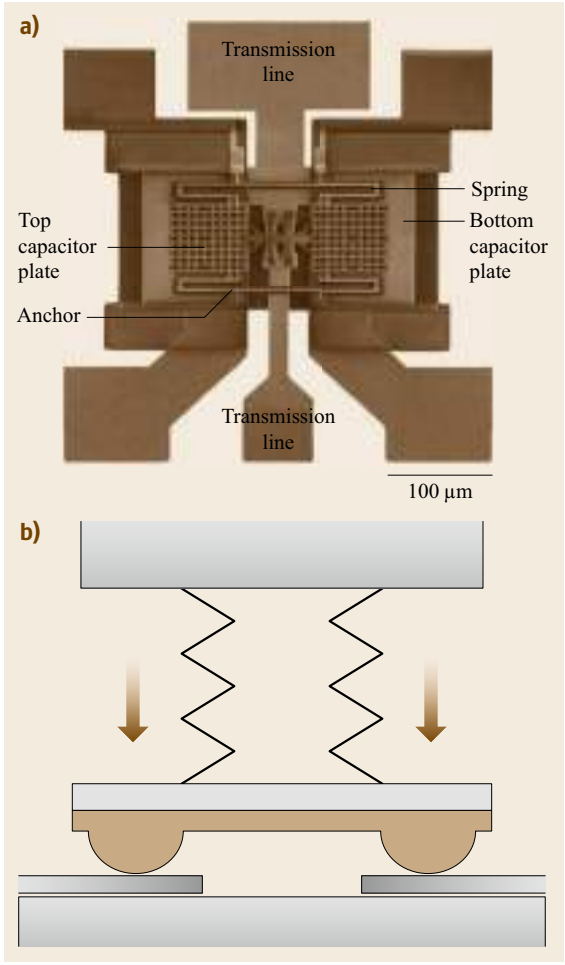


Fig. 41.19 **(a)** Micrograph of an RF MEMS switch designed by Rockwell Scientific, and **(b)** schematic drawing showing the movement of the metal bridge as it comes into contact with the underlying metal plate during switch operation (after [41.42])

deflection in the truss suspensions, but the friction between the cantilever and the substrate electrode holds it in place. The voltage to the substrate electrode is slowly decreased until the device starts to slide. Knowing the electrostatic force generated by the substrate electrode and the stiffness of the truss suspensions, the static friction can be determined. For this device, bumps were fabricated on the bottoms of the cantilever beams. This limited the surface area that came into contact with the substrate, and so lowered the risk of stiction.

Another device that has been designed to study friction is shown in Fig. 41.18 [41.41]. This technique uses a hinged cantilever. The portion near the free end acts as the friction test structure, and the portion near the anchored end acts as the driver. The friction test structure

is attracted to the substrate by means of electrostatic actuation, and when a second electrostatic actuator pulls down the driver, the friction test structure slips forward by a length which is proportional to the forces involved, including the frictional force. This distance, however, has a maximum of 30 nm, so all measurements must be exceedingly accurate in order to investigate a range of forces. This test structure can be used to determine the friction coefficients for surfaces with and without lubricating coatings.

Wear can occur whenever friction between two contacting components is nonzero. Wear mechanisms can entail plastic deformation or brittle fracture. Adhesive wear involves fusion between asperities of two different components, followed by fracture. Wear in polysilicon devices has been investigated using transmission electron microscopy [41.43].

Metal-to-Metal Contact

Recently, radiofrequency (RF) MEMS switches have been developed that are vast improvements over conventional technology. During operation, these devices endure unique mechanical metal-to-metal contacts that can be serious reliability concerns. An example of an RF MEMS switch is shown in Fig. 41.19 [41.42]. As illustrated in Fig. 41.19b, to close the switch the metal bumps are brought into contact with a metal plate, creating a short circuit. When this occurs, significant pressures are developed, depending on the roughness and asperity heights on the metal surfaces. Also, if currents are flowing, Joule heating and arcing can also take place. If the contact area becomes large, through surface roughness modification during operation, the adhesive forces between the two metal components can exceed the restoring force of the springs, and the switch can become stuck in the closed position. Alternatively, high currents can result in melting and short circuits [41.44]. To investigate this mechanical deformation during repeated impact, the RF switches themselves are usually examined before and after operation. A separate testing device has not been designed that could provide any accelerated testing conditions.

41.3 Measurements of Mechanical Properties

All of the techniques discussed above have been used to measure the mechanical properties of MEMS and NEMS materials. As a general rule, the results from the various techniques have agreed well with each other, and the argument becomes which of the measurement techniques is easiest and most reliable to perform. It is crucial to keep in mind, however, that certain properties, such as strength, are process dependent, and so the results taken at one laboratory will not necessarily match those taken from another.

41.3.1 Mechanical Properties of Silicon and Polysilicon

In current MEMS technology, the most widely used structural material is silicon: either single-crystal silicon or polysilicon deposited by low-pressure chemical vapor deposition (LPCVD). One reason for the prevalence of silicon and polysilicon is the large body of processing knowledge for this material that has been developed by the integrated circuit community. Another reason is that silicon possesses a number of qualities beneficial for MEMS devices, particularly its high strength and Young's modulus. As a natural result, most of the mechanical properties investigations on MEMS materials have focused on silicon and polysilicon.

Residual Stresses in Polysilicon

The residual stresses of LPCVD polysilicon have been thoroughly characterized using both the wafer curvature technique and the microstrain gauges. The results from both techniques give consistent values. Figure 41.20 summarizes the residual stress measurements as a function of deposition temperature taken from five different investigations at five different laboratories [41.45]. All five sets of data show the same trend. The stresses change from compressive at the lowest deposition temperatures to tensile at intermediate temperatures and back to compressive at the highest temperatures. The exact transition temperatures vary somewhat among the different investigations, probably due to differences in the deposition conditions:

- The silane or dichlorosilane pressure
- The gas flow rate
- The geometry of the deposition system
- The temperature uniformity.

However, in each data set the transitions are easily discernible. The origin of these residual stress changes lies with the microstructure of the LPCVD polysilicon films.

As with all deposited films, the microstructure of LPCVD polysilicon films is dependent on the deposition conditions. In general, the films are amor-

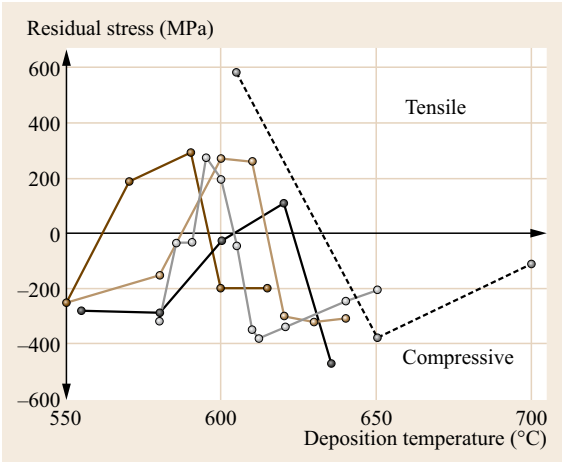


Fig. 41.20 Results for residual stress of LPCVD polysilicon films taken from five different investigations (after [41.45]). The data from each investigation are connected by a *line*

phous at the lowest growth temperatures (lower than $\approx 570\,^{\circ}\mathrm{C}$), display fine ($\approx 0.1\,\mu\mathrm{m}$ diameter) grains at intermediate temperatures (≈ 570–$610\,^{\circ}\mathrm{C}$), and contain columnar (110)-textured grains with a thin fine-grained nucleation layer at the substrate interface at higher temperatures (≈ 610–$700\,^{\circ}\mathrm{C}$) [41.45]. The fine-grained microstructure results from the homogeneous nucleation and growth of silicon grains within an as-deposited amorphous silicon film. In this regime, the deposition rate is just slightly faster than the crystallization rate. The as-deposited films will be crystalline near the substrate interface and amorphous at the free surface (the amorphous fraction can be quickly crystallized by annealing above $610\,^{\circ}\mathrm{C}$). The columnar microstructure seen at the higher growth temperatures results from the formation of crystalline silicon films as-deposited, with growth being fastest in the ⟨110⟩ directions.

The origin of the tensile stress in the fine-grained polysilicon arises from the volume decrease that accompanies the crystallization of the as-deposited amorphous material. The origins of the compressive stresses in the amorphous and columnar films are less well understood. One proposed explanation for compressive stress generation during thin film growth postulates that an increase in the surface chemical potential is caused by the deposition of atoms from the vapor; the increase in surface chemical potential induces atoms to flow into newly formed grain boundaries, creating a compressive stress in the film [41.46]. Investigations of PECVD (plasma-enhanced chemical vapor deposition) silicon films associates an increase in compressive stress with the presence of hydrogen [41.47], which may also be the origin of the compressive stresses in amorphous LPCVD silicon films.

Stress gradients are also typical of LPCVD polysilicon films. The partially amorphous films contain large stress gradients since they are essentially bilayers of compressive amorphous silicon on top of tensile fine-grained polysilicon. The fully crystalline films also exhibit stress gradients. The columnar compressive films are most highly stressed at the film–substrate interface, with the compressive stresses decreasing as the film thickness increases; the fine-grained films are less tensile at the film–substrate interface, with the tensile stresses increasing as the film thickness increases [41.45]. Both stress gradients are associated with microstructural variations. For the columnar films, the initial nucleation layer corresponds to a very high compressive stress, which decreases as the columnar morphology develops. For the fine-grained films, the region near the film-substrate interface has a slightly smaller average grain size, due to heterogeneous nucleation at the interface. This region displays a slightly lower tensile stress than the rest of the film, since the increased grain boundary area reduces the local density.

Young's Modulus of Silicon and Polysilicon

Young's moduli of polysilicon films have been measured using all the techniques discussed above. A good review of the experimental results taken from bulge testing, tensile testing, beam bending, and lateral resonators are contained in [41.48]. All of the reported results are in reasonable agreement, varying from 130–175 GPa, though many values are reported with a relatively high experimental scatter. The main origin of the error in these results is the uncertainties involving the geometries of the small specimens used to make the measurements. For example, from (41.13), the Young's modulus determined by the lateral resonators depends on the cube of the tether beam width, which is typically about $2\,\mu\mathrm{m}$. In general, the beam width and other dimensions can be measured with scanning electron microscopy to within about $0.1\,\mu\mathrm{m}$. However, the width of the beams is not perfectly constant along the entire length, or even through the thickness. These uncertainties in geometry lead to uncertainties in modulus.

The various experimental measurements lie close to the Voigt and Reuss bounds for Young's modulus calculated using the elastic stiffnesses and compliances for single-crystal silicon [41.48]. This strongly implies that Young's modulus of micro- and nanomachined polysilicon structures will be the same as for bulk samples made from polysilicon. This is not unexpected, since Young's modulus is a material property. It is related to the interatomic interactions and should have no dependence on the geometry of the sample. It should be

noted that polysilicon can display a preferred crystallographic orientation depending on the deposition conditions, and that this could affect the Young's modulus of the material, since the Young's modulus of silicon is not isotropic. Young's modulus of single-crystal silicon depends on the orientation of the crystal. Bulk measurements have been carried out and reviewed [41.49]. However, the anisotropy is fairly small for the cubic silicon.

An investigation that utilized electrostatically actuated cantilevers and interferometric deflection detection yielded Young's modulus of 164 GPa [41.24]. They found the grains in their polysilicon films to be randomly oriented, and calculated the Voigt and Reuss bounds to be 163.4–164.4 GPa. This appears to be a very reliable value for randomly oriented polysilicon.

Fracture Toughness and Strength of Silicon and Polysilicon

Using the device shown in Fig. 41.15a and the specimen shown in Fig. 41.15d, the fracture toughness, K_{Ic}, of polysilicon has been shown to be $1.0 \pm 0.1\,\mathrm{MPa\,m^{1/2}}$ [41.50]. Several different polysilicon microstructures were tested, including fine-grained, columnar, and multilayered. Amorphous silicon was also investigated. All of the microstructures displayed the same K_{Ic}. This indicates that, like Young's modulus, fracture toughness is a material property, independent of the material microstructure or the geometry of the sample.

A tensile test, such as shown in Fig. 41.7, but using a sample with indentation-induced sharp precracks reveals a K_{Ic} of $0.86\,\mathrm{MPa\,m^{1/2}}$ [41.51]. The passive, residual stress-loaded beams with sharp precracks, shown in Fig. 41.6, gave a K_{Ic} of $0.81\,\mathrm{MPa\,m^{1/2}}$ [41.14].

Given that K_{Ic} is a material property for polysilicon, measured fracture strength, σ_{crit}, is related to K_{Ic} by

$$K_{Ic} = c\sigma_{crit}(\pi a)^{1/2}\,, \tag{41.14}$$

where a is the crack-initiating flaw size, and c is a constant of order unity. The value for c will depend on the exact size, shape, and orientation of the flaw; for a semicircular flaw, c is equal to 0.71 [41.52]. Therefore, any differences in the reported fracture strength of polysilicon will be due to changes in a.

A good review of the experimental results available in the literature for polysilicon strength is contained in [41.53]. The tensile strength data vary from about 0.5 to 5 GPa. Like many brittle materials, the measured strength of polysilicon is found to obey Weibull statistics [41.54, 55]. This implies that the polysilicon samples contain a random distribution of flaws of various sizes, and that the failure of any particular specimen will occur at the largest flaw that experiences the highest stress. One consequence of this behavior is that, since larger specimens have a greater probability of containing larger flaws, they will exhibit decreased strengths. More specifically, it was found that the most important geometrical parameter is the surface area of the sidewalls of polysilicon [41.53] and single-crystal silicon [41.56] specimens. The sidewalls, as opposed to the top and bottom surfaces, are those surfaces that are created by etching. This is not surprising, since LPCVD polysilicon films contain essentially no flaws within the bulk, and the top and bottom surfaces are typically very smooth.

As a result, the etching techniques used to create the structures will have a strong impact on the fracture strength of the material. For single-crystal silicon specimens it was found that the choice of etchant could change the observed tensile strength by a factor of two [41.57, 58], and by about 20% for polysilicon [41.59]. In addition, the bend strength of amorphous silicon was measured to be twice that of polysilicon for specimens processed identically [41.50]. It was found that the reactive ion etching used to fabricate the specimens produced much rougher sidewalls on the polysilicon than on the amorphous silicon.

An investigation using high-throughput devices achieved 1500 strength measurements in identical polysilicon specimens [41.60]. A strength threshold was detected, implying a maximum flaw size.

The tensile strength of single-crystal silicon has also been measured using a technique similar to that shown in Fig. 41.8. Sharp notches were introduced into the beams using a focused ion beam, and the apparent fracture toughness was measured for varying planes parallel to the notch front, along which the crack propagated. For the {110} notch plane, the fracture toughness was about $1\,\mathrm{MPa\,m^{1/2}}$ [41.61, 62], and for the {100} notch plane it was about $1\,\mathrm{MPa\,m^{1/2}}$ [41.62] or about $2\,\mathrm{MPa\,m^{1/2}}$ [41.61]. The suspended membrane technique has also been used to study creep in silicon at elevated temperatures [41.63].

Fatigue of Silicon and Polysilicon

Fatigue failure involves fracture after a number of load cycles, when each individual load is not sufficient by itself to generate catastrophic cracking in the material. For ductile materials, such as metals, fatigue occurs due to accumulated damage at the site of maximum stress, and involves local plasticity. As a brittle material, polysilicon would not be expected to be susceptible to cyclic fatigue. However, fatigue has been observed for polysilicon tensile samples [41.51], polysilicon bend specimens with notches [41.36, 64], polysilicon bend specimens with sharp cracks [41.65], and polysilicon

torsion specimens [41.66]. The exact origins of the fatigue behavior are still subject to debate. However, some aspects of the experimental data are that the fatigue lifetime does not depend on the loading frequency [41.51], the fatigue behavior is affected by the ambient [41.14, 65], fatigue occurs faster in higher humidities [41.67], and the fatigue depends on the ratio of compressive to tensile stresses seen in the load cycle [41.14].

Fatigue in polysilicon has been found to either cause thickening of the native oxide in the highly stressed regions [41.68], or to cause no thickening of the native oxide in the highly stressed regions [41.69]. Even more strangely, fatigue cycling under certain conditions can lead to apparent strengthening of the polysilicon [41.70, 71]. While the exact mechanism of fatigue in polysilicon is not known, dislocations have been observed in tensile fracture of polysilicon at room temperature [41.72], which implies a role for dislocations in fatigue behavior.

Fatigue in single-crystal silicon has also been found to be affected by humidity, and also by loading frequency [41.73, 74]. In an oxygen-free ambient, silicon devices with smooth sidewalls did not exhibit fatigue, while silicon devices with rough sidewalls covered with thin SiO_2 layers did [41.75]; the authors concluded that neither dislocations nor environmental reactions contributed to fatigue.

Friction of Polysilicon

The friction of polysilicon structures has been measured using the techniques described above. The measured coefficient of friction was found to vary from 4.9 [41.37] to 7.8 [41.41]. The static and dynamic friction of polysilicon coated with monolayer lubricants has been measured with a device similar to that shown in Fig. 41.17, but using a scratch-drive actuator instead of a comb-drive [41.76]. The dynamic friction at 0.2 m/s was approximately 80% of the static friction value; the static friction at zero applied load was due to an adhesive force of 0.95 nN μm^{-2}.

41.3.2 Mechanical Properties of Other Materials

As discussed above, of all the materials used for MEMS and NEMS, polysilicon has generated the most interest as well as the most research in mechanical properties characterization. However, there have been measurements done on other materials, and these are summarized in this section.

As discussed above, one advantage of the externally loaded tension test, as shown in Fig. 41.7, is that essentially any material can be tested using this technique. As such, tensile strengths have been measured to be 0.6–1.9 GPa for SiO_2 [41.77] and 0.7–1.1 GPa for titanium [41.78]. The yield strength for electrodeposited nickel was found to vary from 370 to 900 MPa, depending on annealing temperature [41.19]. In addition, the yield strength was strongly affected by the current density during the electrodeposition process. Both the annealing and current density effects were correlated to changes in the microstructure of the material. Tensile strength and grain size were also found to be affected by electrolyte bath composition [41.79], and strength is increased by the addition of 5–15% tungsten [41.80]. The fatigue behavior of electrodeposited Ni was also investigated [41.81]. Microstructual changes due to thermal treatments also caused the yield strength of AlCuMgMn films to vary from 400 to 1000 MPa [41.82]. Young's moduli were determined to be 100 GPa for titanium [41.78] and 215 GPa for electrodeposited nickel [41.19], and 96 GPa for permalloy [41.27]. Young's modulus for electrodeposited nickel was found to be 159 GPa, using a resonating beam technique [41.30]. The tensile strength, Young's modulus, and Poisson's ratio of silicon nitride were measured to be 5.9 GPa, 257 GPa, and 0.23, respectively [41.83], and the same properties of amorphous diamond-like carbon were found to be 7.3 GPa, 759 GPa, and 0.17, respectively [41.84], with a fracture toughness of 4.4 MPa $m^{1/2}$ [41.85]. Young's modulus and creep behavior of SiC was studied up to 1000 °C [41.86].

The technique of bending cantilever beams, shown in Fig. 41.8, can also be performed on a variety of materials. The yield strength and Young's modulus of gold were found to vary from 260 MPa and 57 GPa, respectively [41.18], to 300 MPa and 120 GPa, respectively [41.87], using this method. The same properties in Al were measured to be 150 MPa and 80 GPa, respectively [41.87], and in silicon nitride to be 6.9 GPa and 260 GPa, respectively [41.88]. Strength, Young's modulus, and maximum strain were determined in carbon-infiltrated carbon nanotube structures to vary from 102 MPa, 4.1 GPa, and 2.5% when stressed perpendicular to the nanotube growth direction to 903 MPa, 11.0 GPa, and 9.2% when stressed parallel to the nanotube growth direction [41.89]. Using a technique similar to that shown in Fig. 41.8, except that a doubly clamped beam was used instead of a cantilever beam, the fracture toughness of ultrananocrystalline diamond was measured as 4.5 MPa $m^{1/2}$ [41.25].

Resonating cantilever beams revealed a Young's modulus of gold of 47 GPa [41.90], and Young's moduli of 1.8 and 14.4 GPa of silica and alumina aerogel thin films, respectively [41.91]. Resonating

beams were also used to determine Young's moduli of GaN and AlN to be 261 and 339 GPa, respectively [41.92].

Another technique that can be used with a number of materials is the membrane deflection method, shown in Fig. 41.12. Silicon nitride measured with this technique revealed a Young's modulus of 258 [41.83] to 325 GPa and a burst strength of 7.1 GPa [41.93]. A polyimide membrane gave a residual stress of 32 MPa, a Young's modulus of 3.0 GPa, and an ultimate strain of about 4% [41.28]. Membranes were also fabricated from polycrystalline SiC films with two different grain structures [41.94]. The film with (110)-texture columnar grains had a residual stress of 434 MPa and a Young's modulus of 349 GPa. The film with equiaxed (110)- and (111)-textured grains had a residual stress of 446 MPa and a Young's modulus of 456 GPa. A separate investigation reported Young's moduli of 417 GPa for polycrystalline and 388 GPa for single-crystal SiC [41.95]. Al and Ni membranes revealed Young's moduli of 70 and 200 GPa, respectively [41.96]. Polymeric SU-8 displayed a fracture strength of 73 MPa and a very low creep rate [41.97]. Al_2O_3 membranes revealed a fracture strength of about 2.6 GPa [41.98].

Other devices that are used to measure mechanical properties require more complicated micromachining, namely patterning, etching, and release, in order to operate. These devices are more difficult to fabricate with materials that are not commonly used as MEMS structural materials. However, the following examples demonstrate work in this area. The structure shown in Fig. 41.5b was fabricated from Si_xN_y, and revealed an apparent fracture toughness of 1.8 MPa $m^{1/2}$ [41.12]. Devices shown in Fig. 41.6 revealed a fracture toughness of atmospheric-pressure CVD SiC of 3.1 MPa $m^{1/2}$ [41.99] and low-pressure CVD SiC of 2.9 MPa $m^{1/2}$ [41.100]. Similar composite beams gave a yield strength of 625 MPa and a strain at crack initiation of 0.3 for Al [41.15]. Devices as shown in Fig. 41.10 showed a fracture toughness of diamond of 4.6 MPa $m^{1/2}$ [41.25], Young's modulus of Cu of 115 GPa [41.101], and Young's modulus and yield strength of Mo of 231 and 1.76 GPa, respectively [41.102].

Lateral resonators of the type shown in Fig. 41.13 were processed using single-crystal and polycrystalline SiC, and Young's modulus was determined to be between 360 GPa [41.103] and 426 GPa [41.104]. Resonating beams revealed the temperature coefficient of Young's modulus to be $-53\,\text{ppm}\,\text{K}^{-1}$ [41.105]. The device shown in Fig. 41.14 was fabricated from polycrystalline germanium and used to measure a bend strength of 1.5 GPa for unannealed Ge and 2.2 GPa for annealed Ge [41.106]. The same device was also fabricated from SiC and revealed a bend strength of 23 GPa [41.107]. Devices similar to that shown in Fig. 41.15 revealed the tensile strength of silicon nitride to be 6.4 GPa [41.93], and the Young's modulus and yield strength of aluminum to be 74.6 GPa and 330 MPa, respectively [41.108]. Young's modulus of an Al-0.5%Cu thin film was measured to be 87 GPa, with a yield strength that depended on the specimen length and thickness [41.109]. The device shown in Fig. 41.16 was used to study the fatigue behavior of electrodeposited nickel [41.110].

The metal-to-metal contact RF switches, such as shown in Fig. 41.19 typically use Au as the contact metal, due to its low electrical resistance and high oxidation resistance. In these devices, adhesive forces were found to increase logarithmically with *cold* contact cycles (cycling with no current passing through the switch), consistent with creep as the underlying physical mechanism [41.42]. At high currents, no adhesion was found between the metal components, but melting led to extrusions and short circuits [41.44]. Both of these results are consistent with separate modeling investigations [41.111, 112].

41.4 Summary

The mechanical properties of the materials comprising MEMS and NEMS are critical to their performance and reliability. Some of these properties, such as stiffness and toughness, are inherent to the materials, while others, such as strength and residual stress, are strongly influenced by the processing techniques used to make and pattern the structures. Various methods have been developed to determine the mechanical properties of micromachined structures – some of them relatively simple and others quite complex. The resulting measurements can be used to design devices with optimal performance.

References

41.1 W. Nix: Mechanical properties of thin films, Metall. Trans. A **20**, 2217–2245 (1989)

41.2 G.G. Stoney: The tension of metallic films deposited by electrolysis, Proc. R. Soc. Lond. A **82**, 172–175 (1909)

41.3 W. Brantley: Calculated elastic constants for stress problems associated with semiconductor devices, J. Appl. Phys. **44**, 534–535 (1973)

41.4 A. Ni, D. Sherman, R. Ballarini, H. Kahn, B. Mi, S.M. Phillips, A.H. Heuer: Optimum design of multilayered polysilicon films for prescribed curvature, J. Mater. Sci. **38**, 4169–4173 (2002)

41.5 J.A. Floro, E. Chason, S.R. Lee, R.D. Twesten, R.Q. Hwang, L.B. Freund: Real-time stress evolution during $Si_{1-x}Ge_x$ heteroepitaxy: Dislocations, islanding, and segregation, J. Electron. Mater. **26**, 969–979 (1997)

41.6 X. Li, B. Bhusan: Micro-nanomechanical characterization of ceramic films for microdevices, Thin Solid Films **340**, 210–217 (1999)

41.7 M.P. de Boer, H. Huang, J.C. Nelson, Z.P. Jiang, W.W. Gerberich: Fracture toughness of silicon and thin film micro-structures by wedge indentation, Mat. Res. Soc. Symp. Proc. **308**, 647–652 (1993)

41.8 H. Guckel, D. Burns, C. Rutigliano, E. Lovell, B. Choi: Diagnostic microstructures for the measurement of intrinsic strain in thin films, J. Micromech. Microeng. **2**, 86–95 (1992)

41.9 F. Ericson, S. Greek, J. Soderkvist, J.-A. Schweitz: High sensitivity surface micromachined structures for internal stress and stress gradient evaluation, J. Micromech. Microeng. **7**, 30–36 (1997)

41.10 L.S. Fan, R.S. Muller, W. Yun, R.T. Howe, J. Huang: Spiral microstructures for the measurement of average strain gradients in thin films. In: *Proc. IEEE MEMSYS* (1990) pp. 177–182 doi:10.1109/MEMSYS.1990.110272

41.11 M. Biebl, H. von Philipsborn: Fracture strength of doped and undoped polysilicon. In: *Proc. Int. Conf. Solid-State Sens. Act* (1995) pp. 72–75 doi:10.1109/SENSOR.1995.721747

41.12 L.S. Fan, R.T. Howe, R.S. Muller: Fracture toughness characterization of brittle films, Sens. Act. A **21–23**, 872–874 (1990)

41.13 H. Kahn, N. Tayebi, R. Ballarini, R.L. Mullen, A.H. Heuer: Wafer-level strength and fracture toughness testing of surface-micromachined MEMS devices, Mat. Res. Soc. Symp. Proc. **605**, 25–30 (2000)

41.14 H. Kahn, R. Ballarini, J.J. Bellante, A.H. Heuer: Fatigue failure in polysilicon not due to simple stress corrosion cracking, Science **298**, 1215–1218 (2002)

41.15 N. André, M. Coulombier, V. De Longueville, D. Fabrègue, T. Gets, S. Gravier, T. Pardoen, J.-P. Raskin: Microfabrication-based nanomechanical laboratory for testing the ductility of submicron aluminum films, Microelectron. Eng. **84**, 2714–2718 (2007)

41.16 A. Boe, A. Safi, M. Coulombier, D. Fabregue, T. Pardoen, J.-P. Raskin: MEMS-based microstructures for nanomechanical characterization of thin films, Smart Mater. Struct. **18**, 115018 (2009)

41.17 W.N. Sharpe Jr., B. Yuan, R.L. Edwards: A new technique for measuring the mechanical properties of thin films, J. Microelectromech. Syst. **6**, 193–199 (1997)

41.18 T.P. Weihs, S. Hong, J.C. Bravman, W.D. Nix: Mechanical deflection of cantilever microbeams: A new technique for testing the mechanical properties of thin films, J. Mater. Res. **3**, 931–942 (1988)

41.19 H.S. Cho, W.G. Babcock, H. Last, K.J. Hemker: Annealing effects on the microstructure and mechanical properties of LIGA nickel for MEMS, Mat. Res. Soc. Symp. Proc. **657**, EE5.23.1–EE5.23.6 (2001)

41.20 W. Suwito, M.L. Dunn, S.J. Cunningham, D.T. Read: Elastic moduli, strength, and fracture initiation at sharp notches in etched single crystal silicon microstructures, J. Appl. Phys. **85**, 3519–3534 (1999)

41.21 C.-S. Oh, W.N. Sharpe: Techniques for measuring thermal expansion and creep of polysilicon, Sens. Act. A **112**, 66–73 (2004)

41.22 L. Li, J. Gomes, G. Brown, D. Uttamchandani, W. Pan, D. Weiland, M. Begbie, C. Lowrie, M.P.Y. Desmulliez: Simultaneous determination of the Young's modulus and Poisson's ratio in micro–nano materials, J. Micromech. Microeng. **19**, 125027 (2009)

41.23 G. Tosolini, L.G. Villanueva, F. Perez-Murano, J. Bausells: Fast on-wafer electrical, mechanical, and electromechanical characterization of piezoresistive cantilever force sensors, Rev. Sci. Inst. **83**, 015002 (2012)

41.24 B.D. Jensen, M.P. de Boer, N.D. Masters, F. Bitsie, D.A. Van La: Interferometry of actuated microcantilevers to determine material properties and test structure nonidealities in MEMS, J. Microelectromech. Syst. **10**, 336–346 (2001)

41.25 H.D. Espinosa, B. Peng: A new methodology to investigate fracture toughness of freestanding MEMS and advanced materials in thin film form, J. Microelectromech. Syst. **14**, 153–159 (2005)

41.26 V. Mulloni, S. Colpo, A. Faes, B. Margesin: A simple analytical method for residual stress measurement on suspended MEM structures using surface profilometry, J. Micromech. Microeng. **23**, 025025 (2013)

41.27 X. Li, G. Ding, T. Ando, M. Shikida, K. Sato: Micromechanical characterization of electroplated permalloy films for MEMS, Microsyst. Technol. **14**, 131–134 (2007)

41.28 M.G. Allen, M. Mehregany, R.T. Howe, S.D. Senturia: Microfabricated structures for the in situ measurement of residual stress, Young's modulus, and ultimate strain of thin films, Appl. Phys. Lett. **51**, 241–243 (1987)

41.29 L. Kiesewetter, J.-M. Zhang, D. Houdeau, A. Steckenborn: Determination of Young's moduli of mi-

cromechanical thin films using the resonance method, Sens. Act. A **35**, 153–159 (1992)
41.30 S. He, J.S. Chang, L. Li, H. Ho: Characterization of Young's modulus and residual stress gradient of MetalMUMPs electroplated nickel film, Sens. Act. A **154**, 149–156 (2009)
41.31 E. Boyd, V. Nock, D. Weiland, X. Li, D. Uttamchandani: Direct comparison of stylus and resonant methods for determining Young's modulus of single and multilayer MEMS cantilevers, Sens. Act. A **172**, 440–446 (2011)
41.32 H. Kahn, S. Stemmer, K. Nandakumar, A.H. Heuer, R.L. Mullen, R. Ballarini, M.A. Huff: Mechanical properties of thick, surface micromachined polysilicon films. In: *Proc. IEEE MEMSYS* (1996) pp. 343–348
41.33 W.C. Tang, T.-C.H. Nguyen, R.T. Howe: Laterally driven polysilicon resonant microstructures, Sens. Act. A **20**, 25–32 (1989)
41.34 P.T. Jones, G.C. Johnson, R.T. Howe: Fracture strength of polycrystalline silicon. In: *Mat. Res. Soc. Symp. Proc., San Francisco*, Vol. 518 (1998) pp. 197–202
41.35 P. Minotti, R. Le Moal, E. Joseph, G. Bourbon: Toward standard method for microelectromechanical systems material measurement through on-chip electrostatic probing of micrometer size polysilicon tensile specimens, Jpn. J. Appl. Phys. **40**, L120–L122 (2001)
41.36 C.L. Muhlstein, E.A. Stach, R.O. Ritchie: A reaction-layer mechanism for the delayed failure of micron-scale polycrystalline silicon structural films subjected to high-cycle fatigue loading, Acta Mater **50**, 3579–3595 (2002)
41.37 M.G. Lim, J.C. Chang, D.P. Schultz, R.T. Howe, R.M. White: Polysilicon microstructures to characterize static friction. In: *Proc. IEEE MEMSYS, Napa Valley* (1990) pp. 82–88
41.38 A. Corigliano, B. De Masi, A. Frangi, C. Comi, A. Villa, M. Marchi: Mechanical characterization of polysilicon through on-chip tensile tests, J. Microelectromech. Syst. **13**, 200–219 (2004)
41.39 W. Kang, J.H. Han, M.T.A. Saif: A novel method for in situ uniaxial tests at the micro/nanoscale – Part II: Experiment, J. Microelectromech. Syst. **19**, 1322–1330 (2010)
41.40 D. Zhang, J.-M. Breguet, R. Clavel, V. Sivakov, S. Christiansen, J. Michler: In situ electron microscopy mechanical testing of silicon nanowires using electrostatically actuated tensile stages, J. Microelectromech. Syst. **19**, 663–674 (2010)
41.41 B.T. Crozier, M.P. de Boer, J.M. Redmond, D.F. Bahr, T.A. Michalske: Friction measurement in MEMS using a new test structure, Mat. Res. Soc. Symp. Proc., Boston **605**, 129–134 (2000)
41.42 G. Gregori, D.R. Clarke: The interrelation between adhesion, contact creep, and roughness on the life of gold contacts in radio-frequency microswitches, J. Appl. Phys. **100**, 094904 (2006)
41.43 D.H. Alsem, E.A. Stach, M.T. Dugger, M. Enachescu, R.O. Ritchie: An electron microscopy study of wear in polysilicon microelectromechanical systems in ambient air, Thin Solid Films **515**, 3259–3266 (2007)
41.44 S.T. Patton, J.S. Zabinski: Fundamental studies of Au contacts in MEMS RF switches, Tribol. Lett. **18**, 215–230 (2005)
41.45 J. Yang, H. Kahn, A.Q. He, S.M. Phillips, A.H. Heuer: A new technique for producing large-area as-deposited zero-stress LPCVD polysilicon films: The multipoly process, J. Microelectromech. Syst. **9**, 485–494 (2000)
41.46 E. Chason, B.W. Sheldon, L.B. Freund, J.A. Floro, S.J. Hearne: Origin of compressive residual stress in polycrystalline thin films, Phys. Rev. Lett. **88**, 156103 (2002)
41.47 J. Gaspar, O. Paul, V. Chu, J.P. Conde: Mechanical properties of thin silicon films deposited at low temperatures by PECVD, J. Micromech. Microeng. **20**, 035022 (2010)
41.48 S. Jayaraman, R.L. Edwards, K.J. Hemker: Relating mechanical testing and microstructural features of polysilicon thin films, J. Mater. Res. **14**, 688–697 (1999)
41.49 M.A. Hopcroft, W.D. Nix, T.W. Kenny: What is the Young's modulus of silicon?, J. Microelectromech. Syst. **19**, 229–238 (2010)
41.50 R. Ballarini, H. Kahn, N. Tayebi, A.H. Heuer: Effects of microstructure on the strength and fracture toughness of polysilicon: A wafer level testing approach, ASTM STP (2001) doi:10.1520/STP109795
41.51 J. Bagdahn, J. Schischka, M. Petzold, W.N. Sharpe Jr.: Fracture toughness and fatigue investigations of polycrystalline silicon. In: *Proc. SPIE Conf. Reliab. Testing Charact. MEMS–MOEMS*, Vol. 4558 (2001) pp. 159–168
41.52 I.S. Raju, J.C. Newman Jr.: Stress intensity factors for a wide range of semi-elliptical surface cracks in finite-thickness plates, Eng. Frac. Mech. **11**, 817–829 (1979)
41.53 J. Bagdahn, W.N. Sharpe Jr., O. Jadaan: Fracture strength of polysilicon at stress concentrations, J. Microelectromech. Syst. **12**, 302 (2003)
41.54 R. Boroch, J. Wiaranowski, R. Mueller-Fiedler, M. Ebert, J. Bagdahn: Characterization of strength properties of thin polycrystalline silicon films for MEMS applications, Fatigue Fract. Eng. Mater. Struct. **30**, 2–12 (2007)
41.55 B.L. Boyce, J.M. Grazier, T.E. Buchheit, M.J. Shaw: Strength distributions in polycrystalline silicon MEMS, J. Microelectromech. Syst. **16**, 179–190 (2007)
41.56 D.C. Miller, B.L. Boyce, M.T. Dugger, T.E. Buchheit, K. Gall: Characteristics of a commercially available silicon-on-insulator MEMS material, Sens. Act. A **138**, 130–144 (2007)
41.57 T. Yi, L. Li, C.-J. Kim: Microscale material testing of single crystalline silicon: Process effects on surface morphology and tensile strength, Sens. Act. A **83**, 172–178 (2000)
41.58 M.S. Gaither, R.S. Gates, R. Kirkpatrick, R.F. Cook, F.W. DelRio: Etching process effects on surface structure, fracture strength, and reliability of single-crystal silicon theta-like specimens, J. Micro-

electromech. Syst. **22**, 589–602 (2013)
41.59 T. Alan, M.A. Hines, A.T. Zehnder: Effect of surface morphology on the fracture strength of silicon nanobeams, Appl. Phys. Lett. **89**, 091901 (2006)
41.60 E.D. Reedy Jr., B.L. Boyce, J.W. Foulk III, R.V. Field Jr., M.P. de Boer, S.S. Hazra: Predicting fracture in micrometer-scale polycrystalline silicon MEMS structures, J. Microelectromech. Syst. **20**, 922–932 (2011)
41.61 X. Li, T. Kasai, S. Nakao, T. Ando, M. Shikida, K. Sato, H. Tanaka: Anisotropy in fracture of single crystal silicon film characterized under uniaxial tensile condition, Sens. Act. A **117**, 143–150 (2005)
41.62 S. Koyama, K. Takashima, Y. Higo: Fracture toughness measurement of a micro-sized single crystal silicon, Key Eng. Mater. **297–300**, 292–298 (2005)
41.63 J. Ren, M. Ward, P. Kinnell, R. Craddock, X. Wei: Plastic deformation of micromachined silicon diaphragms with a sealed cavity at high temperatures, Sensors **16**, s16020204 (2016)
41.64 H. Kahn, R. Ballarini, R.L. Mullen, A.H. Heuer: Electrostatically actuated failure of microfabricated polysilicon fracture mechanics specimens, Proc. R. Soc. Lond. A **455**, 3807–3923 (1999)
41.65 W.W. Van Arsdell, S.B. Brown: Subcritical crack growth in silicon MEMS, J. Microelectromech. Syst. **8**, 319–327 (1999)
41.66 J.-N. Hung, H. Hocheng, K. Sato: Torsion fatigue testing of polycrystalline silicon cross-microbridge structures, Jpn. J. Appl. Phys. **50**, 06GM07 (2011)
41.67 Y. Yamaji, K. Sugano, O. Tabata, T. Tsuchiya: Tensile-mode fatigue tests and fatigue life predictions of single crystal silicon in humidity controlled environments. In: *Proc. IEEE MEMS* (2007) pp. 267–270
41.68 D.H. Alsem, R. Timmerman, B.L. Boyce, E.A. Stach, J.T.M. De Hosson, R.O. Ritchie: Very high-cycle fatigue failure in micron-scale polycrystalline silicon films: Effects of environment and surface oxide thickness, J. Appl. Phys. **101**, 013515 (2007)
41.69 H. Kahn, A. Avishai, R. Ballarini, A.H. Heuer: Surface oxide effects on failure of polysilicon MEMS after cyclic and monotonic loading, Scr. Mater. **59**, 912–915 (2008)
41.70 H. Kahn, L. Chen, R. Ballarini, A.H. Heuer: Mechanical fatigue of polysilicon: Effects of mean stress and stress amplitude, Acta Mater **54**, 667–678 (2006)
41.71 R.E. Boroch, R. Müller-Fiedler, J. Bagdahn, P. Gumbsch: High cycle fatigue and strengthening in polycrystalline silicon, Scr. Mater. **59**, 936–940 (2008)
41.72 S. Nakao, T. Ando, M. Shikida, K. Sato: Effects of environmental condition on the strength of submicron-thick single crystal silicon film. In: *Proc. 14th Int. Conf. Transducers* (2007) pp. 375–378
41.73 T. Ikehara, T. Tsuchiya: Low-cycle to ultrahigh-cycle fatigue lifetime measurement of single-crystal-silicon specimens using a microresonator test device, J. Microelectromech. Syst. **21**, 830–839 (2012)
41.74 E.K. Baumert, P.-O. Theillet, O.N. Pierron: Investigation of the low-cycle fatigue mechanism for micron-scale monocrystalline silicon films, Acta Mater **58**, 2854–2863 (2010)
41.75 V.A. Hong, S. Yoneoka, M.W. Messana, A.B. Graham, J.C. Salvia, T.T. Branchflower, E.J. Ng, T.W. Kenny: Fatigue experiments on single crystal silicon in an oxygen-free environment, J. Microelectromech. Syst. **24**, 351–359 (2015)
41.76 A. Corwin, M.P. de Boer: Effect of adhesion on dynamic and static friction in surface micromachining, Appl. Phys. Lett. **84**, 2451–2453 (2004)
41.77 T. Tsuchiya, A. Inoue, J. Sakata: Tensile testing of insulating thin films; humidity effect on tensile strength of SiO_2 films, Sens. Act. A **82**, 286–290 (2000)
41.78 H. Ogawa, K. Suzuki, S. Kaneko, Y. Nakano, Y. Ishikawa, T. Kitahara: Measurements of mechanical properties of microfabricated thin films. In: *Proc. IEEE MEMSYS* (1997) pp. 430–435
41.79 J. Tang, H. Wang, X.Q. Guo, R. Liu, X.H. Dai, G. Ding, C.S. Yang: An investigation of microstructure and mechanical properties of UV-LIGA nickel thin films electroplated in different electrolytes, J. Micromech. Microeng. **20**, 025033 (2010)
41.80 S.J. Suresha, M. Haj-Taieb, K. Bade, J. Aktaa, K.J. Hemker: The influence of tungsten on the thermal stability and mechanical behavior of electrodeposited nickel MEMS structures, Scr. Mater. **63**, 1141–1144 (2010)
41.81 Y. Yang, B.I. Imasogie, S.M. Allameh, B. Boyce, K. Lian, J. Lou, W.O. Soboyejo: Mechanisms of fatigue in LIGA Ni MEMS thin films, Mater. Sci. Eng. A **444**, 39–50 (2007)
41.82 R. Modlinski, R. Puers, I. De Wolf: AlCuMgMn micro-tensile samples mechanical characterization of MEMS materials at micro-scale, Sens. Act. A **143**, 120–128 (2008)
41.83 R.L. Edwards, G. Coles, W.N. Sharpe: Comparison of tensile and bulge tests for thin-film silicon nitride, Exp. Mech. **44**, 49–54 (2004)
41.84 S. Cho, I. Chasiotis, T.A. Friedmann, J.P. Sullivan: Young's modulus, Poisson's ratio and failure properties of tetrahedral amorphous diamond-like carbon for MEMS devices, J. Micromech. Microeng. **15**, 728–735 (2005)
41.85 K.N. Jonnalagadda, I. Chasiotis: Strength and fracture resistance of amorphous diamond-like carbon films for MEMS, J. Nanomater. **2009**, 204281 (2009)
41.86 D. Liesen, R. Rusanov, F. Rohlfing, T. Fuchs, C. Eberl, H. Riesch-Opperman, O. Kraft: Mechanical characterization between room temperature and 1000 °C of SiC free-standing thin films by a novel high-temperature micro-tensile setup, Rev. Sci. Inst. **86**, 055104 (2015)
41.87 D. Son, J.-H. Jeong, D. Kwon: Film-thickness considerations in microcantilever-beam test in measuring mechanical properties of metal thin film, Thin Solid Films **437**, 182–187 (2003)
41.88 W.-H. Chuang, T. Luger, R.K. Fettig, R. Ghodssi: Mechanical property characterization of LPCVD sil-

icon nitride thin films at cryogenic temperatures, J. Microelectomech. Syst. **13**, 870–879 (2004)
41.89 B.H. Hanna, W.C. Fazio, J.D. Tanner, J.M. Lund, T.S. Wood, R.C. Davis, R.R. Vanfleet, B.D. Jensen: Mechanical property measurement of carbon infiltrated carbon nanotube structures for compliant micromechanisms, J. Microelectromech. Syst. **23**, 1330–1339 (2014)
41.90 C.-W. Baek, Y.-K. Kim, Y. Ahn, Y.-H. Kim: Measurement of the mechanical properties of electroplated gold thin films using micromachined beam structures, Sens. Act. A **117**, 17–27 (2005)
41.91 R. Yokokawa, J.-A. Paik, B. Dunn, N. Kitazawa, H. Kotera, C.-J. Kim: Mechanical properties of aerogel-like thin films used for MEMS, J. Micromech. Microeng. **14**, 681–686 (2004)
41.92 A.B. Amar, M. Faucher, V. Brandli, Y. Cordier, D. Theron: Young's modulus extraction of epitaxial heterostructure AlGaN/GaN for MEMS applications, Phys. Status Solidi A **211**, 1655–1659 (2014)
41.93 A. Kaushik, H. Kahn, A.H. Heuer: Wafer-level mechanical characterization of silicon nitride MEMS, J. Microelectromech. Syst. **14**, 359–367 (2005)
41.94 S. Roy, C.A. Zorman, M. Mehregany: The mechanical properties of polycrystalline silicon carbide films determined using bulk micromachined diaphragms. In: *Mat. Res. Soc. Symp. Proc., Boston*, Vol. 657 (2001) pp. EE9.5.1–EE9.5.6
41.95 O. Pabst, M. Schiffer, E. Obermeier, T. Tekin, K.D. Lang, H.-D. Ngo: Measurement of young's modulus and residual stress of thin SiC layers for MEMS high temperature applications, Microsyst. Technol. **18**, 945–953 (2012)
41.96 A. Reddy, H. Kahn, A.H. Heuer: A MEMS-based evaluation of the mechanical properties of metallic thin films, J. Microelectromech. Syst. **16**, 650–658 (2007)
41.97 B. Schoeberle, M. Wendlandt, C. Hierold: Long-term creep behavior of SU-8 membranes: Application of the time-stress superposition principle to determine the master creep compliance curve, Sens. Act. A **142**, 242–249 (2008)
41.98 M. Berdova, O.M.E. Ylivaara, V. Rontu, P.T. Torma, R.L. Puurunen, S. Franssila: Fracture properties of atomic layer deposited aluminum oxide freestanding membranes, J. Vac. Sci. Technol. A **33**, 01A106 (2015)
41.99 J.J. Bellante, H. Kahn, R. Ballarini, C.A. Zorman, M. Mehregany, A.H. Heuer: Fracture toughness of polycrystalline silicon carbide thin films, Appl. Phys. Lett. **86**, 071920 (2005)
41.100 V. Hatty, H. Kahn, J. Trevino, C.A. Zorman, M. Mehregany, R. Ballarini, A.H. Heuer: Fracture toughness of low-pressure chemical-vapor-deposited polycrystalline silicon carbide thin films, J. Appl. Phys. **99**, 013517 (2006)
41.101 Z.M. Zhou, Y. Zhou, C.S. Yang, J.A. Chen, W. Ding, G.F. Ding: The evaluation of Young's modulus and residual stress of copper films by microbridge testing, Sens. Act. A **127**, 392–397 (2006)
41.102 J.-H. Kim, H.-J. Lee, S.-W. Han, J.-Y. Kim, J.-S. Kim, J.-Y. Kang, S.-H. Choa, C.-S. Lee: Mechanical behavior of freestanding Mo thin films for RF-MEMS devices, Int. J. Mod. Phys. B **20**, 3757–3762 (2006)
41.103 W. Chang, C. Zorman: Determination of Young's moduli of 3C (110) single-crystal and (111) polycrystalline silicon carbide from operating frequencies, J. Mater. Sci. **43**, 4512–4517 (2008)
41.104 A.J. Fleischman, X. Wei, C.A. Zorman, M. Mehregany: Surface micromachining of polycrystalline SiC deposited on SiO_2 by APCVD, Mater. Sci. Forum **264–268**, 885–888 (1998)
41.105 M. Pozzi, M. Hassan, A.J. Harris, J.S. Burdess, L. Jiang, K.K. Lee, R. Cheung, G.J. Phelps, N.G. Wright, C.A. Zorman, M. Mehregany: Mechanical properties of a 3C-SiC film between room temperature and 600 °C, J. Phys. D **40**, 3335–3342 (2007)
41.106 A.E. Franke, E. Bilic, D.T. Chang, P.T. Jones, T.-J. King, R.T. Howe, G.C. Johnson: Post-CMOS integration of germanium microstructures. In: *Proc. IEEE MEMSYS* (1999) doi:10.1109/MEMSYS.1999.746901
41.107 D. Gao, C. Carraro, V. Radmilovic, R.T. Howe, R. Maboudian: Fracture of polycrystalline 3C-SiC films in microelectromechanical systems, J. Microelectromech. Syst. **13**, 972–976 (2004)
41.108 M.A. Haque, M.T.A. Saif: A review of MEMS-based microscale and nanoscale tensile and bending testing, Exp. Mech. **43**, 248–255 (2003)
41.109 M.E. Saleh, J.L. Beuth, Y.N. Picard, M.P. de Boer: In situ platform for isothermal testing of thin-film mechanical properties using thermal actuators, J. Microelectromech. Syst. **24**, 2008–2018 (2015)
41.110 E.K. Baumert, O.N. Pierron: Fatigue degradation properties of LIGA Ni films using kilohertz microresonators, J. Microelectromech. Syst. **22**, 16–25 (2013)
41.111 J. Song, D.J. Srolovitz: Adhesion effects in material transfer in mechanical contacts, Acta Mater **54**, 5305–5312 (2006)
41.112 O. Rezvanian, M.A. Zikry, C. Brown, J. Krim: Surface roughness, asperity contact and gold RF MEMS switch behavior, J. Micromech. Microeng. **17**, 2006–2015 (2007)

42. High Volume Manufacturing and Field Stability of MEMS Products

T. Kieran Nunan, Mark G. da Silva

Low-volume microelectromechanical systems (MEMS)/nanoelectromechanical systems (NEMS) production is practical when an attractive concept is implemented with business, manufacturing, packaging, and test support. Moving beyond this to high-volume production adds requirements on design, process control, quality, product stability, market size, market maturity, internal capacity, and capital investment or transfer to foundry and business systems. In a broad sense, this chapter uses a case study approach: It describes and compares the silicon-based MEMS accelerometers and gyroscopes that are in high-volume production. What is described here also applies to other MEMS products such as pressure sensors, image projection systems, microphones, etc. Although they serve several markets, these businesses have common characteristics. For example, the manufacturing lines use automated semiconductor equipment and standard material sets to make consistent products in large quantities. Standard well-controlled processes are sometimes modified for a MEMS product. However, novel processes that cannot run with standard equipment and material sets are avoided when possible. When transferring to an external foundry, existing processes are modified to utilize the foundry equipment and processes where possible. This reliance on semiconductor tools, as well as the organizational practices required to manufacture clean, particle-free products, partially explains why the MEMS market leaders are integrated circuit manufacturers. There are other factors. MEMS and NEMS are enabling technologies, so it can take several years for high-volume applications to develop. Indeed, market size is usually a strong function of price. This becomes a vicious circle, because low price requires low cost – a result that is normally achieved only after a product is in high-volume production. During the early years, IC companies reduce cost and financial risk by using existing facilities for low-volume MEMS production. As a result, product architectures are partially determined by capabilities developed for previous products. This chapter includes a discussion of MEMS product architecture with particular attention to the impact of electronic integration, packaging, and surfaces. Packaging and testing are critical, because they are significant factors in MEMS product cost. MEMS devices have extremely high surface-to-volume ratio, so performance and stability may depend on the control of surface characteristics *after packaging*. Looking into the future, the competitive advantage of IC suppliers is decreasing because MEMS foundries are growing and small companies are learning to integrate MEMS/NEMS devices with die from CMOS foundries in one package. Packaging challenges still remain, because most MEMS/NEMS products must interact with the environment without degrading stability or reliability.

42.1 **High-Volume Manufacturing Strategy** 1482
42.1.1 Foundry versus Internal Fab 1482
42.1.2 The ADI Experience 1482
42.1.3 Foundry Manufacturing..... 1486
42.1.4 High-Volume Inertial MEMS Manufacturing Worldwide 1487

42.2 **Robust Design for Volume Manufacturing**..... 1491
42.2.1 MEMS Design for Manufacturability (DfM) 1491

42.3 **Stable Field Performance**..... 1496
42.3.1 Electric Field Stability..... 1496
42.3.2 Mechanical Field Stability..... 1497

42.4 **Internet of Things (IoT)** 1498
42.4.1 Heterogeneous Integration..... 1499

42.5 **Conclusions and Outlook** 1503

References..... 1503

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_42

This chapter builds on the summary presented in [42.1], and covers high-volume inertial MEMS manufacturing over the last 15 years. MEMS are not limited to inertial products, and there are many other products such as inkjet printers and optical MEMS devices such as optical switches for the telecommunications field and optical light projectors such as TI's Digital Light Processing (DLP) system that use MEMS technologies. Except for optical switches, the inkjet and DLP MEMS technologies are mature, with most of the business being captured by a few dominant players. Other high-volume MEMS products such as pressure sensors, microphones, and MEMS switches will not be mentioned, and the main focus will be on inertial sensors. Newer MEMS products such as biomedical microelectromechanical systems (bioMEMS), environmental and chemical sensors, energy harvesting, and energy storage devices will not be discussed, as such technologies are still in the early stages of adoption, although seen to be a growing business with the onset of the Internet of Things (IoT) [42.2]. Maybe some insight into the evolution of these newer sensors and devices can be gained from this review of the evolution of the inertial MEMS sensor business.

42.1 High-Volume Manufacturing Strategy

In our discussion of MEMS manufacturing strategy, we will use the Analog Devices Inc. (ADI) experience and how the manufacturing strategy at ADI evolved. We will also discuss the manufacturing strategy for other high-volume inertial MEMS manufacturing companies. We hope that, through this discussion, the reader will see how MEMS high-volume manufacturing is closely linked to the semiconductor IC high-volume manufacturing strategy and the importance of external foundries in high-volume manufacturing. We will also discuss the challenges in using external foundries for MEMS manufacturing. The biggest challenge is that MEMS products typically require their own technology platform or different process modules tailored to MEMS product performance needs.

42.1.1 Foundry versus Internal Fab

In the early 2000s, inertial MEMS were a commodity product for many applications, including inertial sensing and pressure sensing. MEMS microphones were also beginning to gain some traction. Currently, there are two major approaches for high-volume manufacturing, either using an internal fab, e.g., STM and Bosch, or using external MEMS foundries, e.g., companies such as Knowles, InvenSense, and mCube. The main foundries used were IC foundries in Asia (e.g., Knowles uses a MEMS foundry in China to make their MEMS microphone).

Large foundries offer MEMS foundry services in their older 6 inch and 8 inch fabs to help with fab utilization, as standard IC production in these fabs transitioned to processing more advanced submicron complementary metal-oxide-semiconductor (CMOS) technology. The preferred MEMS technologies that could utilize this model were inertial, microphone, and pressure sensor designs that could take advantage of the standard equipment set in these older IC fabs, thereby avoiding large investment in new capital, and that were compatible with the other IC processes running in the fabs.

42.1.2 The ADI Experience

Evolution of Inertial MEMS

ADI started out with integrated MEMS devices based on the iMEMS® process [42.3–5], see Fig. 42.1.

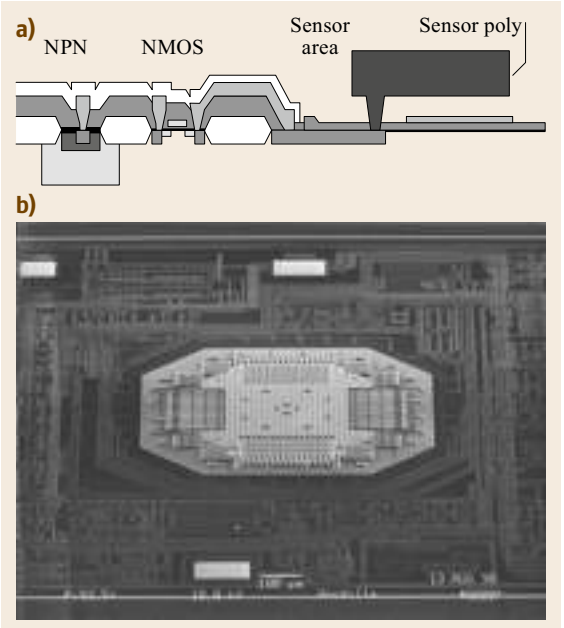


Fig. 42.1a,b iMEMS® sensor and signal processing devices on the same chip. **(a)** Cross section showing BIMOS circuit elements and sensor beam on the same chip. **(b)** Top down view of the integrated MEMS die with the sensor surrounded by the BIMOS circuits (courtesy of Analog Devices Inc.)

The evolution of this core technology was to increase the thickness of the MEMS element and reduce the parasitic capacitance to improve the signal-to-noise performance. The major changes were to replace the N^+ runners for the electrical connections between the sensor and electronics [42.6] with polysilicon runners that had lower series resistance and lower parasitic capacitance. At the same time, the sensor thickness was increased from 2 to 3 μm and eventually to 4 μm. Other changes centered on the packaging to drive down the cost and utilize original equipment manufacturer (OEM)-supplied packaging. This evolution went from metal can for the ADXL50 to ceramic and ceramic dual in-line package (CERDIP) cavity packages to eventually low-cost plastic overmolded packages. To enable overmold packaging, the sensor had to be capped [42.7, 8], see Fig. 42.2, and coated with antistiction coating technology [42.9] at wafer level, requiring development of a wafer-level antistiction coating process and a wafer-level capping technology.

The development of silicon-on-insulator microelectromechanical systems (SOIMEMS) technology [42.10–12], see Fig. 42.3, enabled MEMS integration with submicron CMOS technology and a 10 μm-thick sensor element.

So, the trend over the years has been to improve signal-to-noise performance and overload performance by increasing the sensor element thickness and reducing noise sources such as parasitic capacitance and Brownian noise while additionally reducing die size to reduce cost further.

ADI MEMS-Only Process

Before 2000, ADI MEMS products were manufactured using their integrated MEMS process flows such as *iMEMS*® and *SOIMEMS*. These integrated processes were highly customized to ADI internal IC processes as well as the unique MEMS process steps to integrate the MEMS element with the electronics directly on the same chip. As customers demanded greater functionality, including high-density digital signal processing, it became cost prohibitive to add the high-density circuity onto a single integrated chip. This resulted in a paradigm shift from discrete chip integration to newer products having two discrete chips in a single package: a MEMS chip with the front-end electronics for MEMS signal processing, and a second application-specific integrated circuit (ASIC) chip for additional high-density digital signal processing. This was still a very high-cost solution, so to drive down the cost, in particular for products targeted at high-volume commercial applications, ADI developed their first MEMS-only process, see Fig. 42.4a and 42.3b [42.13, 14].

ADI took some of the front-end poly interconnect and sensor anchor steps from the iMEMS® process and the process know-how for these steps and the back-end high-aspect-ratio deep silicon etch and sensor release steps from the SOIMEMS process to create the process flow for the MEMS-only process. The first MEMS-only process had two levels of poly interconnect and an 8 μm-thick polysilicon MEMS element. By using a thick substrate oxide, and two levels of poly interconnect with a low-pressure chemical vapor deposition (LPCVD) nitride etch-stop layer, the parasitic capacitance to the sensitive node in the MEMS device could be kept to a minimum, enabling manufacture of small high-performance inertial MEMS devices at relatively low cost.

Foundry MEMS at ADI

The next step in reducing the manufacturing cost was to transfer some if not all of the manufacturing steps to an external foundry capable of processing higher wafer volume rates. Other than the thick sensor polysilicon process that required careful control of the deposition conditions, all the other steps up to sensor element re-

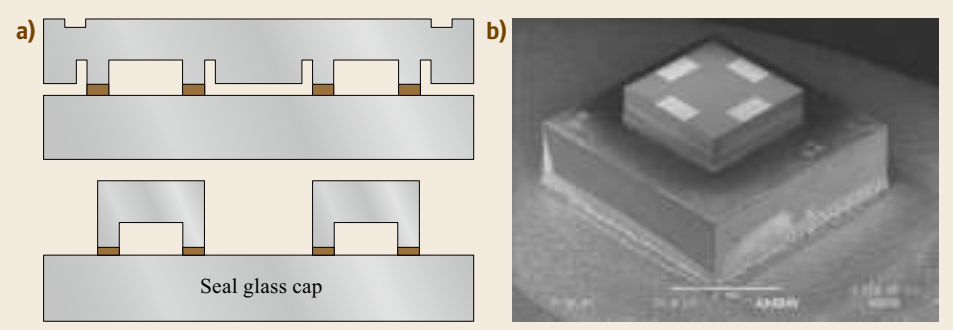


Fig. 42.2a,b Wafer-level seal glass frit capping process for MEMS devices using dicing for cap singulation (courtesy of Analog Devices Inc.); **(a)** cross section showing wafer level capping process using glass frit, **(b)** SEM image of finished capped MEMS devices after dicing

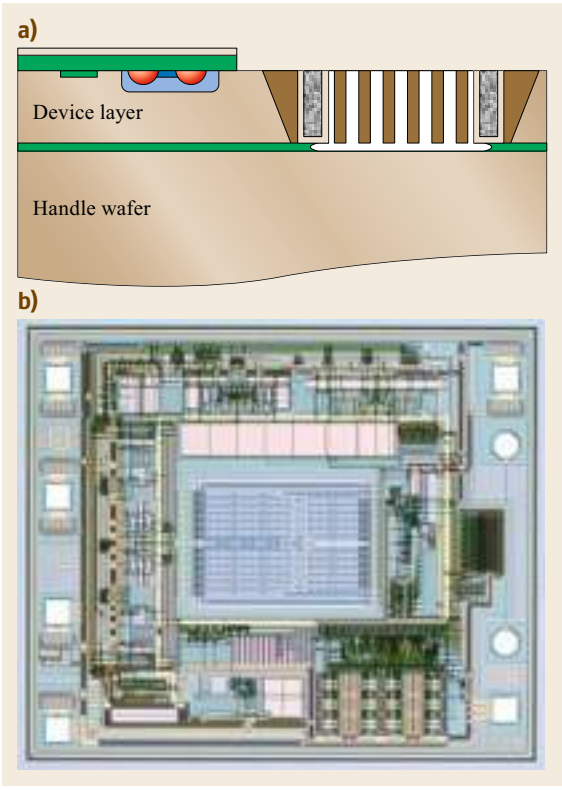


Fig. 42.3a,b SOIMEMS sensor and signal processing devices on the same chip (courtesy of Analog Devices Inc.); **(a)** cross section showing sub-micron circuit elements and 10 μm sensor element on the same chip, **(b)** top down view of the integrated MEMS die with the sensor surrounded by the sub-micron CMOS circuit elements on a Bonded silicon on insulator substrate

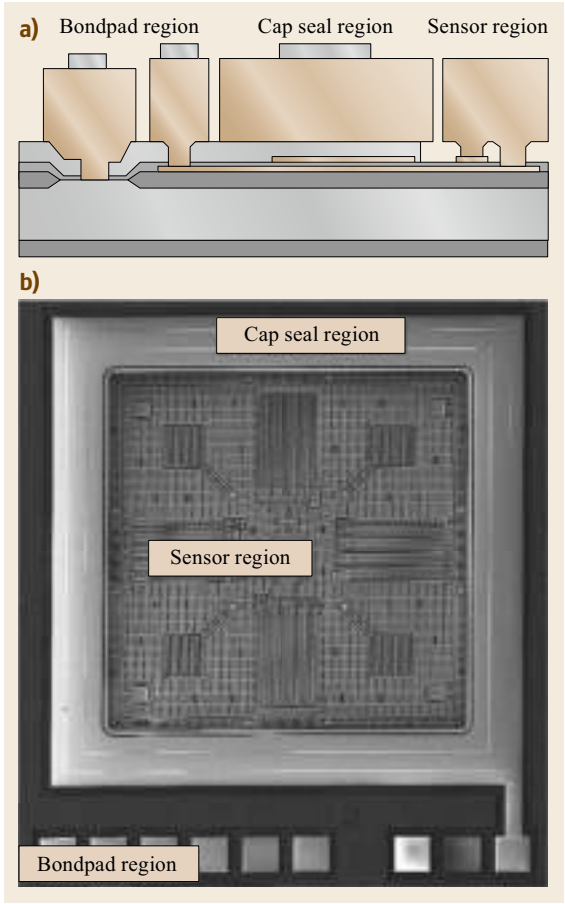


Fig. 42.4a,b MEMS-only process with 8 μm- or 16 μm-thick sensor polysilicon (courtesy of Analog Devices Inc.); **(a)** cross section showing the layers for the MEMS-only process, **(b)** top down view of the MEMS-only die before capping

lease were relatively standard processes available in a conventional IC foundry. ADI partnered with an external foundry to set up their first MEMS-only foundry process. This external foundry sent a team of engineers to work with the ADI MEMS engineers to understand all the process steps that were going to be run at the foundry, including the thick sensor polysilicon process. Every effort was taken to transfer the process as exactly as possible, but some process steps were changed to a different deposition tool but with similar film characteristics. The decisions to transfer exactly (as for the sensor poly process) or use similar but different films for other layers were made based on the expected impact on device properties and manufacturability. So, for a relatively small investment in capital, the ADI MEMS process up to the sensor element release step was transferred to the foundry with minimal impact on the other processes running in the same fab. This could be achieved as all the MEMS process steps up to the release step were compatible with the foundry fab and could use most of the existing equipment in the IC foundry fab.

For the first MEMS-only process to run at the foundry, this external foundry processed the MEMS wafers up to the sensor element release step and the cap wafers up to the seal glass frit application step to be ready to bond to the MEMS device wafers. ADI then completed the MEMS fabrication steps that included the proprietary sensor element release using a buffered oxide etch (BOE), and the application of the antistiction coating [42.9] to prevent in-use *stiction* on the MEMS device wafer. ADI also completed the processing steps on the cap wafer including the application of the seal glass frit bonding material, the wafer bonding, and final cap singulation to separate the individual die. These steps done at ADI are unique

MEMS steps requiring customized equipment and processes.

With the success of this first MEMS-only process foundry transfer, ADI and this external foundry partner continued to codevelop a full MEMS-only high-volume foundry process. This was done on 8 inch wafers on one of the 8 inch lines at the foundry. To do all the processing at the foundry, a number of process steps that would be compatible with the ADI MEMS-only process and design architectures had to be developed. This included an OEM-supplied vapor HF (vHF) release process and an OEM-supplied antistiction coating process for the MEMS element release. The cap process and the cap wafer for the MEMS wafer bonding were redesigned to take advantage of the fab equipment and process available at this foundry, such as the bonders for the bond step and the grind and polish tools for the cap singulation process, see Fig. 42.5.

ADI developed unit process steps for the vHF release process by working with an OEM supplier for the vHF step that would be compatible with the ADI MEMS-only process using a *low-stress nitride* (LSN) etch-stop layer. ADI also developed an antistiction coating process on an OEM-supplied antistiction coating tool as well as a new capping process that was compatible with a metal eutectic bond [42.15] and also compatible with the standard equipment and process steps in the CMOS fab. Further development was required for the cap singulation step to make it compatible with the new cap bonding process using equipment that was available at the foundry for the cap grinding steps. The bulk of the process development centered around integration of these new unit process steps to maintain sensor device performance and yield. Numerous adjustments or tweaks to the layout and process steps had to be made to address issues that surfaced with the integration of each of these new process steps. Process tweaks were made regarding the integrity of the etch-stop layer to the vHF and residues such as ammonium fluoride from the vHF release. Significant development effort went into achieving a predictable high-yielding metal bond process and a defect-free cap singulation process. There was heavy reliance on the foundry process development and manufacturing team to understand and optimize the process steps to fix the various issues that surfaced during the integration and manufacturing ramp-up. Through this close relationship between the foundry process teams and the ADI development team, a high-yielding and reliable MEMS-only process was qualified and released for manufacturing at the foundry. The strategy adopted and the key to success was redesigning the MEMS-only process to minimize the number of new processes and tools required for the foundry process and designing the process to be compatible with the existing cleanroom. Additional benefits were also realized; e.g., with the vHF release process, it was relatively easy to increase the sensor thickness from 8 to 16 µm, to increase the S/N performance and overload performance for the MEMS sensor element.

Strategy for Using MEMS Foundry Services

For a fabless company or organization looking to use external MEMS foundry services, there are two basic options or strategies. One option is to use a MEMS foundry process that the foundry owns and offers. The MEMS company may provide guidelines to the foundry to develop a MEMS process that is compatible with their MEMS designs. This works well for fabless MEMS companies (e.g., mCube). The approach is to obtain the design rules and design guide or process design kit (PDK) from the MEMS foundry and design the MEMS device using these. This is a well-established

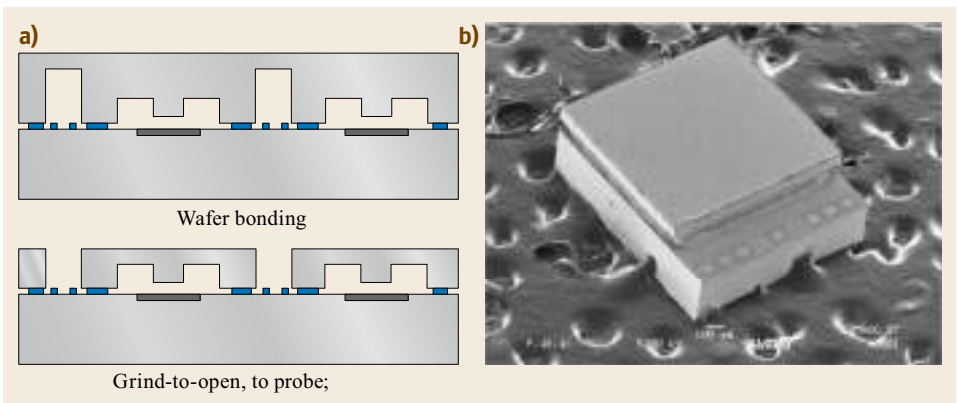


Fig. 42.5a,b MEMS cap process using grind-to-open for cap singulation (courtesy of Analog Devices Inc.); **(a)** cross section showing capping and cap grinding using Grind-to-open process, **(b)** SEM image of the 8"MEMS-only sensor die after dicing

methodology for semiconductor IC development, as the processes are well established and characterization data for individual active and passive devices manufactured using this foundry process are very well characterized and accurately simulatable. Also, the interaction of the IC devices when making a fully functional product is very well understood, so if the design does not violate any of the design rules, the product will with high likelihood operate as expected and pass qualification. The same approach for an off-the-shelf MEMS foundry process has been found to be not as predictable (Sect. 42.2). There are many variants in the sensor design between different MEMS sensor suppliers, and it is very challenging for a MEMS foundry to predict all the variants in the sensor design and how they will interact with the fabrication process flow. Typically, MEMS suppliers have developed models to predict the sensor performance and the effects of package stress and temperature on the sensor performance. These models are validated during characterization and qualification and are then adjusted (within statistical limits) to match the actual product performance. Package stress effects, anchor design, spring design, and surface parameters such as roughness and surface charge can all influence device performance. This takes a number of iterations to optimize (Robust Design for Volume Manufacturing methodology in Sect. 42.2), and often the process is modified to address issues found, in order to meet the target performance, as it may not always be possible to modify the sensor element design. In short, it has been the general experience that, to take advantage of the external foundry process, the product supplier will need a few cycles of learning to understand these interactions and implement a design for manufacturability (DfM) strategy to be successful. These process design iterations could result in a less optimal product in terms of cost, size, or performance in order to meet time-to-market requirements.

The other model for the MEMS product supplier is to develop a fabrication process internally (or at local R&D facilities) and then transfer the process to an external high-volume foundry, similar to the approach adopted by ADI and InvenSense. The difference between ADI and InvenSense is that ADI developed their MEMS processes in house using captive equipment sets and later transferred these individual process steps to the external foundry. In the case of InvenSense, the approach was to use many external MEMS foundry services to develop the process (short loop experiments and prototypes) and products and then transfer all of the processing know-how to a foundry for high-volume manufacturing. If the process is transferred exactly (including equipment utilized), then there can be high confidence that the designs will work as predicted using this process. Also, since the transferred process is proprietary to the MEMS product supplier, they have the flexibility to make modifications to the process for new products or features.

The foundry option has various advantages and disadvantages compared with using an internal fab dedicated to making MEMS devices in-house. The main advantage of internal manufacturing is the close working relationship between the MEMS designers, assembly and testing, and the fab and the flexibility to make process tweaks to address product issues as well as process changes to increase device performance and reduce cost for each product generation. Continuous constant feedback between the process and the MEMS design is critical for designing new MEMS products with quick design of experiments (DOE) cycle times. This freedom to adjust the process and add new process modules has significant advantages for new product generations. The disadvantage is the cost of maintaining a dedicated in-house wafer fab and maintain a positive variance by operating the fab close to capacity. In contrast, a dedicated MEMS fab relies on high-volume wafer starts, which is typically consumer products with lower margins. Furthermore, the consumer market is unpredictable due to its highly competitive nature and extremely short product cycles, so using a foundry service can help mitigate these risks.

42.1.3 Foundry Manufacturing

MEMS Industry Group (MIG) [42.16] provides a general guideline for anyone considering using a foundry service on their website [42.17].

In general, a statement of work (SOW) is agreed upon, which defines the scope of the work, the expected process flow and specifications, and either a timeline for product development, process transfer, and qualification, or a codevelopment agreement. The most straightforward development is a pure product development on an established *standard* MEMS fabrication process offered by the foundry, which is very rare. This will generally involve the foundry taking most of the responsibility for manufacturing the product and shipping the wafers when completed. Key items that need to be agreed upon include wafer cost, schedule, and intellectual property (IP) ownership. Both parties need to be aligned on general expectations regarding expected yield, reliability, and quality, since these can depend on unanticipated product–process interactions that will need to be resolved.

For a MEMS process transfer, additional items need to be included in the SOW and the contract. These will include setting up engineering teams at both the process owner(s) and foundry, to work out the details for

each of the fabrication steps and how each of these will be transferred or replicated at the foundry. This may involve adding some capital equipment and training and also includes sharing of equipment recipes, process specifications and control limits, and test structures. In practice, a sensor from an already released product is usually selected to qualify the process transfer and validate reliability and quality.

42.1.4 High-Volume Inertial MEMS Manufacturing Worldwide

Similar trends can be observed for other high-volume MEMS manufacturing companies. Referring to market research reports [42.18, 19], it can be observed that Bosch has also been making steady progress in increasing the sensor thickness as well as the dielectric thickness to increase the signal-to-noise performance. A review of tear down reports show that over time they have reduced the die size and increased the wafer size to reduce cost and add improved performance and functionality. Other trends include going from seal glass frit to metal bond for the capping process and combo devices in the same package. ST Microelectronics, another high-volume supplier of commodity MEMS inertial products, has made similar evolutions on a similar MEMS process. InvenSense and mCube are examples of two fabless MEMS companies that use foundries for their MEMS process for high-volume production. Their MEMS process architecture is quite different from those of ADI, Bosch, and ST, as they use a significantly different approach to make the MEMS sensor and the electrical interconnect between the ASIC and the MEMS. One of the main differences is that they use oxide chemical–mechanical polishing (CMP) and fusion bonding as some of the core process steps for building their MEMS devices. These process steps are similar to unit process steps in high-volume IC foundry fabs for making IC products, so it did not require significant process development or capital outlay to adapt them for MEMS products.

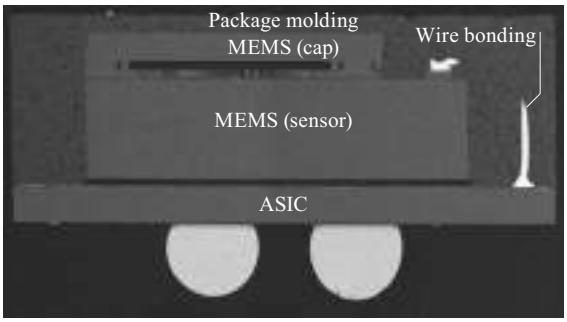


Fig. 42.6 Bosch BMA355 three-axis accelerometer cross section (courtesy of System Plus Consulting)

Bosch Sensor Fabrication Process

The Bosch version of the MEMS-only process [42.19] is an elegant process where a thick oxide is formed, using the first polysilicon layer (*poly1*) to route the electrical connections to the sensor and using a thick epi-polysilicon layer for the sensor. Figure 42.6 presents a typical example, showing a cross section of the construction of the Bosch three-axis accelerometer device.

Fig. 42.7 shows some of the sensor structure for the Bosch IMU product. Based on a previous review [42.18], the evolution of the process at Bosch is quite interesting, showing that early discrete MEMS sensor devices used 11 µm-thick sensor polysilicon with no etch-stop layer for sacrificial oxide removal. They also used conventional seal glass frit bonding for the wafer-level capping process to protect the MEMS device and enable low-cost plastic packaging. Later generations point to 20 µm epi poly and a second poly layer to improve device performance and decrease sensor size, while further driving down manufacturing

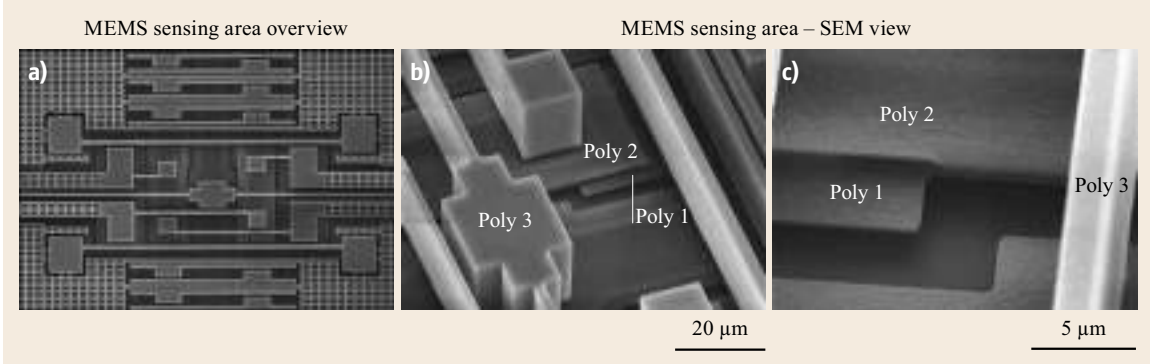


Fig. 42.7a–c Bosch SMI130 six-axis MEMS inertial measurement unit (IMU) (courtesy of System Plus Consulting); views **(a)**, **(b)** and **(c)** showing different details for the sensor poly (poly3) and the interconnect poly (poly1 and poly2)

cost. Additional cost and size reduction was achieved in subsequent iterations, with a bumped ASIC with through-silicon via (TSV) through the ASIC die. This part was still plastic encapsulated, but at the wafer level, providing significant reduction in assembly and test costs if final testing was done on the bumped ASIC before dicing. Bosch was also one of the leaders in implementing Al-Ge eutectic bonding instead of seal glass frit to further decrease the die size and therefore die cost. In contrast, the discrete sensor gyro devices show that the sensor poly thickness was 10 µm with an Al-Ge eutectic bonding capping process implemented.

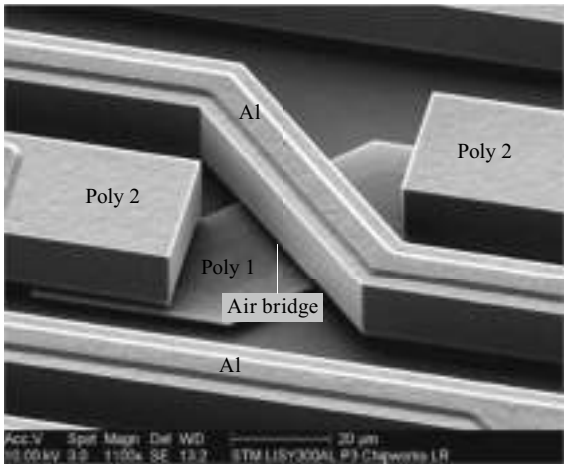


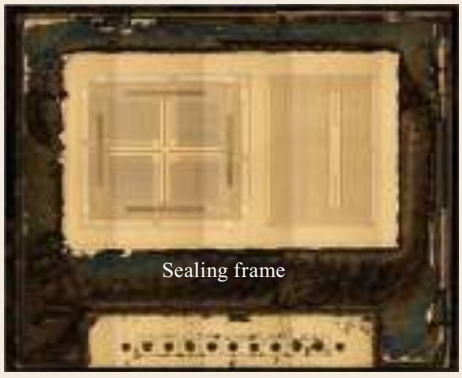
Fig. 42.8 ST LIS331DL three-axis accelerometer – metal on poly routing (courtesy of TechInsights Inc., formerly Chipworks)

There is no *getter* layer evident in the construction, suggesting that the switch to Al-Ge may be to maintain low cavity pressure, which is needed for improved gyro performance. While increasing the sensor layer thickness and mass improves the S/N performance of accelerometer devices, provided that the Brownian noise is kept to a minimum, one does not get the same improvement in device performance for gyro designs, as they are influenced much more by the noise floor of the device. High-performance accelerometer devices can benefit directly in terms of both S/N and overload performance by increasing the sensor thickness and bonding at lower pressure to achieve cavity pressures in the range of 30–500 mbar. For gyro devices the Brownian noise has a greater effect, increasing in proportional to the sensor thickness, so there is not the same advantage with increasing the sensor thickness for gyro designs. Improvement in performance is mostly gained by decreasing the cavity pressure to below 10 mbar, which is very difficult to achieve with seal glass frit bonding, although such lower pressures can be achieved by using a metal eutectic bond (such as Al-Ge). This may explain why, for the multi-DOF (combined accelerometer and gyro die in the same package) designs, the manufacturer chose to use 10 µm sensor poly with Al-Ge metal eutectic bond for the gyro die while the accelerometer die in the same package has a 20 µm sensor layer with seal glass frit bonding [42.19].

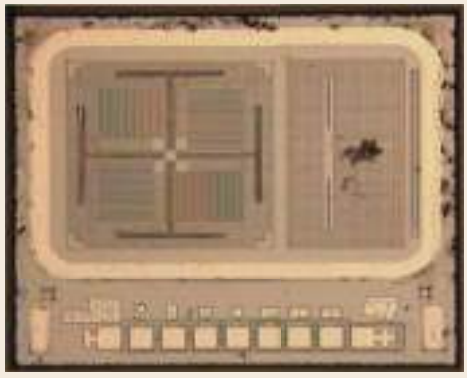
ST Microsensor Fabrication Process

STM also have an elegant process for manufacturing discrete sensors, very similar to the Bosch process with

a) LIS331DLH (2008)



b) LIS3DH (2010)



c)

	Die area	Sensing area	Sealing	Sealing frame width
LIS331DLH	4.71 mm^2 2.30 mm × 2.05 mm	1.44 mm^2 (≈ 30% of die area) 1.6 mm × 0.9 mm	Pb glass frit	180–220 µm
LIS3DH	2.01 mm^2 1.52 mm × 1.32 mm	0.94 mm^2 (≈ 47% of die area) 1.25 mm × 0.75 mm	Gold	70 µm

Fig. 42.9a–c LIS331DLH and LIS3DH, showing the process evolution for ST's low-power three-axis accelerometer process (courtesy of **(a)** TechInsights Inc. (formerly Chipworks) and **(b)** System Plus Consulting [42.18])

a thick substrate layer, polysilicon runners (signal routing), and a 15 μm-thick epi-polysilicon sensor layer in their early products. They also used conventional seal glass frit bonding for wafer-level encapsulation of the sensor to enable low-cost plastic packaging. Fig. 42.8 and Fig. 42.9 is a typical example showing some of the sensor structure of STM inertial sensors.

STM also use polysilicon for the sensor element, as well as additional polysilicon layers for signal routing and with an added metal layer (similar to the bond pad metal) on top of the thick polysilicon routing to further reduce the series resistance. For their later products, STM increased the sensor polysilicon to 24 μm and replaced the seal glass frit capping bond with a gold eutectic (Au-Au) thermal compression metal bond to reduce the die size (and cost) and also improve the accelerometer device performance. Other changes appear to be the use of materials such as tungsten silicide for the polysilicon routing layer to reduce the series resistance of the electrical connections. For their multi-DOF devices, the manufacturer chose to use the same polysilicon thickness of 24 μm for both the gyro and accelerometer sensor elements but added a *getter* layer to the capping process for the gyro sensor (deposited within the cap cavity) to maintain the cap cavity at very low pressure.

Interestingly, in other multi-DOF devices from the manufacturer, the decision was made to combine the accelerometer and gyro sensor elements on the same chip but with separate capped environments for each element; i. e., the cavity for the accelerometer and gyro is separated using a conventional seal glass frit bond. With a getter in the gyro sensor element cavity and no getter in the accelerometer cavity, both cavities can be fabricated at substantially different pressures. This highlights the simple fact that optimum performance for both the accelerometer and gyroscope cannot be achieved if the sensor thickness and cavity pressure are identical for both devices. Other process changes implemented in this fabrication flow include TSV through the sensor substrate to further reduce overall die size, and also added polysilicon layer routing paths in more recent products [42.19].

InvenSense Sensor Fabrication Process

InvenSense was one of the first fabless MEMS companies to make MEMS gyroscopic sensors using an external MEMS volume manufacturing foundry. Their first-generation MEMS gyro product was a two-axis device with a 35 μm sensor element layer using Al-Ge metal eutectic (likely related to manufacturability decisions at the external foundry) to bond the MEMS device to the ASIC and also to form the electrical connections between the sensor nodes and the ASIC. Fig. 42.10 and Fig. 42.11 shows some of the sensor details for the InvenSence products.

The process flow begins with a cap wafer with cavities etched into it, followed by silicon-to-silicon fusion bonding of the MEMS layer wafer to the cap wafer. The MEMS wafer is then ground back to the desired sensor thickness (35 μm), and germanium (Ge) is patterned on the previously bonded MEMS–cap stack. The sensor design is then formed within the 35 μm MEMS layer using a standard deep reactive-ion etch (DRIE). At this point, the capped MEMS wafer is combined with a conventionally fabricated standard CMOS ASIC wafer, but utilizing the top metal ASIC layer as the bonding nodes and the bond seal ring. Early device iterations had deep cavities etched into the ASIC die, which was likely not favored due to the necessary increase in the ASIC die size. However, more recent process iterations show that designs evolved to work without a cavity but used the top metal layer as a shield or sense node under the MEMS die.

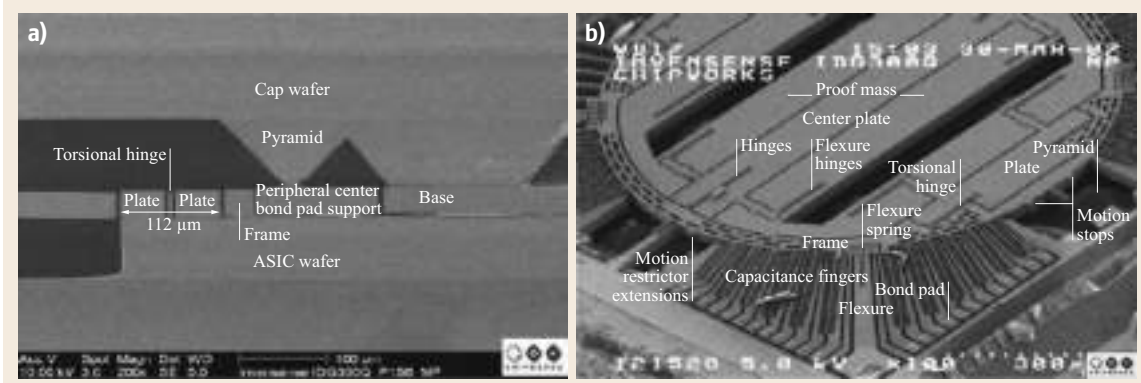


Fig. 42.10a,b InvenSense IDG-300 dual-axis MEMS gyroscope. **(a)** Cross sectional view and **(b)** top down view (courtesy of TechInsights Inc., formerly Chipworks)

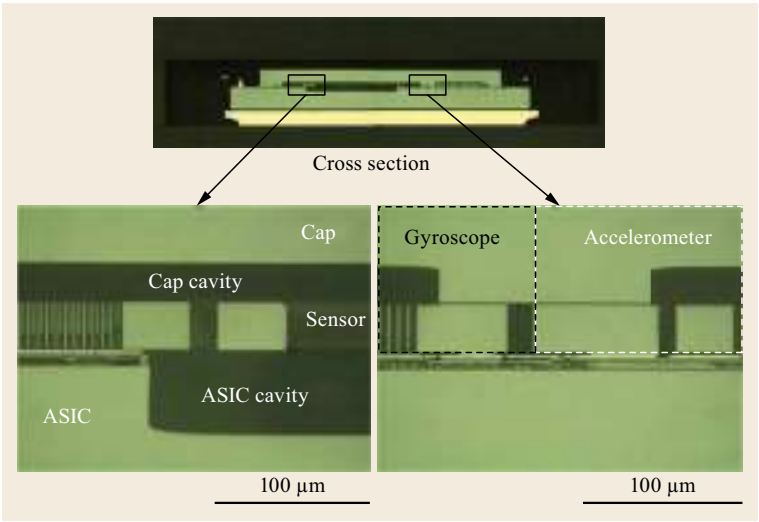


Fig. 42.11 InvenSense MPU-6000 three-axis accelerometer and three-axis gyro (courtesy of System Plus Consulting)

While there was some evolution in the process, like going from KOH etching to pattern the cap wafer to DRIE and using the top metal as an electrode for the MEMS sensing, most of the fabrication process evolution appears to have been directed towards the MEMS sensor element design and the ASIC technology nodes to shrink the die size and enable multi-DOF devices (combined three-axis gyroscope and three-axis accelerometer) on a single die.

mCube Sensor Fabrication Process

Another recent example of a company directly deploying their production to an external high-volume MEMS foundry is mCube, a new fabless inertial MEMS company. Reports [42.19] indicate that mCube appears to be using a fairly well-characterized MEMS process flow available to implement their own unique sensor design. Fig. 42.12 shows some of the sensor details for mCube process. The process has some similarities to the InvenSense sensor fabrication process and appears to use similar unit processes such as oxide-to-silicon fusion bonding and Al-Ge eutectic bonding. Both InvenSense and mCube processes do not require sacrificial oxide removal. Similar to earlier descriptions, the MEMS process uses fusion bonding to bond the MEMS wafer to the ASIC wafer, followed by a grind-back step of the MEMS wafer to the target sensor element thickness. Electrical connections through the sensor layers are realized using tungsten (W)-filled vias to make contact directly to the ASIC sensor nodes and interestingly to form the sensor anchors. The sensor element is patterned using a conventional DRIE, which also serves to isolate the sensor electrical connections, and finally the MEMS die is capped using a cavity cap wafer with metal Al-Ge eutectic bond.

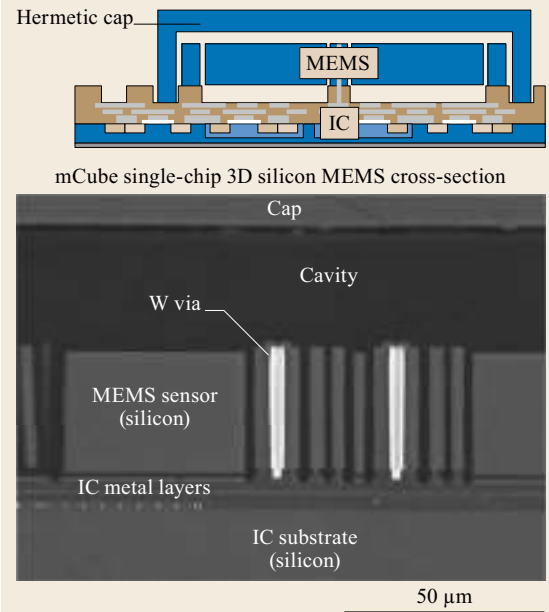


Fig. 42.12 mCube MC3635 three-axis accelerometer cross section (courtesy of System Plus Consulting in "mCube MC3635 Accelerometer Reverse Costing, 2006"). The mechanical layer of the MEMS accelerometer is constituted of silicon layer fusion bonded on top of the metal layers of the CMOS IC. The silicon layer is patterned by DRIE to form the mechanical structure. The electrical connection between the MEMS silicon layer ($\approx 30\,\mu$m thick) and the IC metal layer is realized with tungsten

One significant difference emerging between the MEMS processes used by ADI, Bosch, and STM for inertial sensors compared with InvenSense and mCube is

that the former have the flexibility to make the MEMS die and ASIC die independently as small as possible without any constraints on the MEMS or ASIC, while for the latter the MEMS and the ASIC die size are constrained to be the same, which does imply some limitations in terms of potential for manufacturing cost reduction. The other limitation is that it is difficult to evolve the latter for multi-DOF combined products that take advantage of either different cavity pressures or different device layer thicknesses. InvenSense in particular has been able to work around these limitations by using smart sensor designs and have introduced both 6DOF and 9DOF products to the market.

42.2 Robust Design for Volume Manufacturing

Robust design for manufacturing of MEMS implies deploying a MEMS element design strategy that results in high-yielding, controllable fabrication processes that provide functional and reliable sensor chips, as directly measured as the percentage of chips in a finished lot that pass all tests and function as specified. As seen in the previous section, the main requirement to build reliable parts is well-qualified manufacturing processes (Sect. 42.1); however, in this section, we take a look at the design strategy and in particular design for manufacturability (DfM) and reliability (DfR) [42.20], which are equally important for high volume MEMS manufacturing. In the MEMS industry, the challenges encountered in volume manufacturing have generally been underestimated in terms of the technical challenges they present due to the level of custom process development needed (*one product – one process*), and the effort required to resolve failure modes due to the relative immaturity of the process integration required. Section 42.1 covered some ground in describing available MEMS high-volume processes and their advantages/disadvantages and limitations, so the focus for the next few pages will be on DfM techniques to ensure successful, robust high-volume production.

42.2.1 MEMS Design for Manufacturability (DfM)

Development of MEMS-enabled products has generally followed a *one product – one process* [42.21] paradigm. Although a transition to use or reuse of standard scalable platforms has begun [42.22], as described in the previous section, there still exists a significant challenge for developing manufacturable and reliable [42.23] MEMS-enabled sensor products. DfM as a product development philosophy is well known [42.20, 21, 23], but the application of some common DfM principles (such as use of modular components, design for reuse, etc.) faces adoption challenges for MEMS development in particular (DfM incorporates the manufacturing process as part of the design and development process, including concurrent engineering design principles to produce manufacturable designs for cost-effective volume production [42.20]). The benefits of incorporating DfM for MEMS-enabled products are numerous, but at the top of the list would be reduced product development costs and time, being essential to meet the increased pace of system solution (IoT) introduction and competition. These improvements in manufacturability are typically characterized by:

- Minimization of the number of components in the manufacturing flow – specifically use of fewer masks and process modules. As seen in the cases in Sect. 42.1.4, even necessary modifications to the sensor process flow can be time consuming.
- Use of standard design (e.g., CAD models, material properties) or process elements (e.g., tools, fabrication steps, and instrumentation). In Sect. 42.1.4, the manufacturers (e.g., Bosch, STM etc.) chose to use existing, available process elements (flow) and implemented their unique sensor design within in it to ensure manufacturability.
- Available process flows or platforms can be used for a portfolio of MEMS-enabled products, which can translate to enormous potential to cut costs and reduce time to market. As noted in Sect. 42.1.4, all the cited manufacturers adopted this approach, as it ensures expandability of the product portfolio with minimal manufacturing development risk.

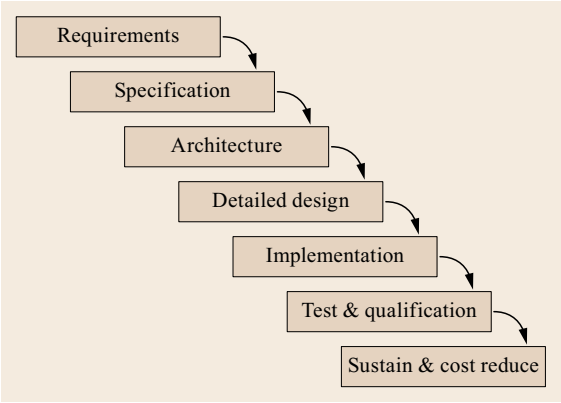


Fig. 42.13 *Waterfall* model for product development

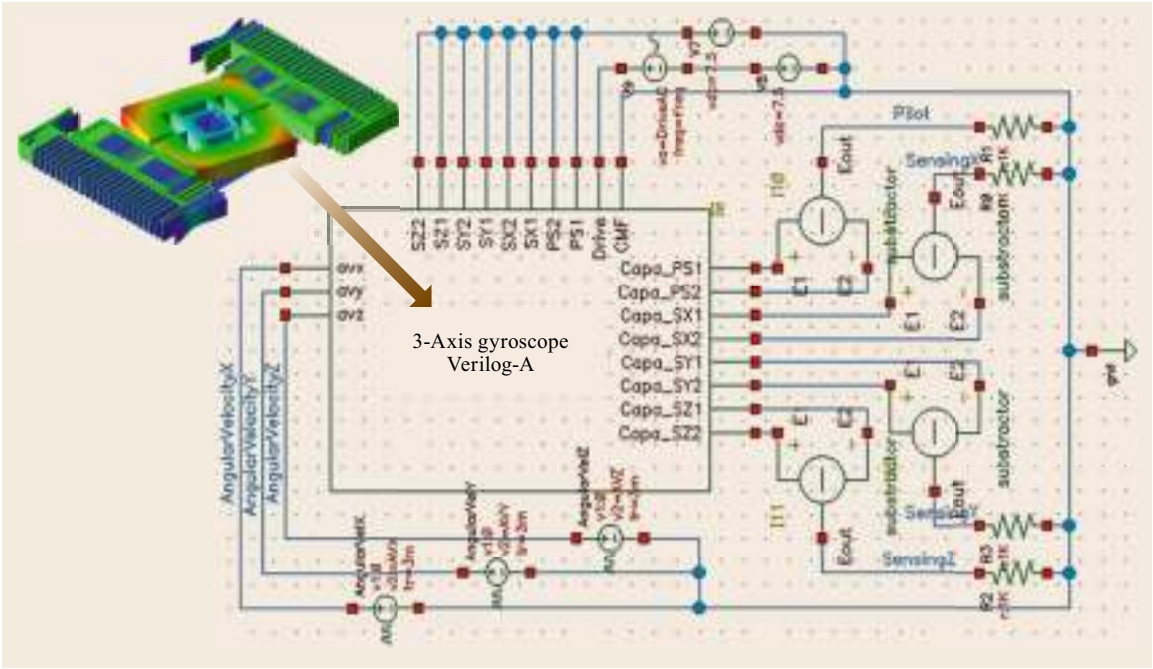


Fig. 42.14 Schematic model of a MEMS inertial sensor realized in MEMS+® (reprinted with permission from Coventor Inc. 2017)

In a typical MEMS component development flow, the product design phase has several overlapping phases that flow together, in what is widely referred to as the *waterfall* model (Fig. 42.13). Although this has been covered in significant detail elsewhere [42.23], it is worth highlighting that in the primary development stages, these phases have a significant impact on the course of product development and the robustness of the final product.

Technical Requirements and Architectural Design Phase, where the customer needs of performance, cost target, and component description are clearly articulated through the definition of a clear, detailed technical specification for the final product. This is a crucial step in the process, and one that can take many iterations. It typically involves data analysis, simulation, and even in some cases test chips to translate product requirements into precise engineering requirements and subsystem specifications that can serve as necessary inputs for the subsequent, more complex design phase.

Before any detailed design work can commence, the overall product architecture must be established. This is not a trivial task, as it has deep implications not only at the manufacturing level (which technology to employ, e.g., wafer-level chip-scale packaging (WLCSP), 0.18 μm CMOS, etc.) but also for the fundamental strategy at the design level (e.g., open/closed loop, power saving schemes, etc.), with the most critical issue being harmony between the design and process flow selected. As mentioned above, the down-selection process (especially for the sensor element design) often includes early evaluation of multiple sensor element designs [42.24], and this can typically result in evaluation of early test chips to identify critical limits of the design or process. Through iteration, once the system architecture has been established (based on general considerations of manufacturing availability; e.g., 0.18 μm CMOS ultralow leakage for the ASIC, or 30 μm-thick polysilicon for the sensor element are typical choices), evaluation of multiple design concepts needs to be completed in order to select candidate designs that are closest to meeting the required specifications. On the ASIC side, this translates to evaluating approaches for various subblocks (analog to digital converter (ADC), oscillator etc.), whereas on the sensor element side, this means understanding the design envelope available (spring constants, position of stoppers, etc.).

For the design of the ASIC signal chain, well-characterized PDKs are available from high-volume foundries (e.g., TSMC, Global Foundries, etc.) with well-characterized design rules, which makes it less burdensome as long as the design operates within established process limits. However, for MEMS element design, PDKs are not always available [42.25]. A MEMS foundry may make a limited set of design rules available, which sets design constraints for certain

but not all primary design parameters, and although this allows designers to proceed with the task of creating and evaluating multiple design concepts (using specialized design automation tools, e.g., CoventorWare™ or Intellisuite™), the interaction of the design and process (e.g., etch hole size, stopper shape, etc.) can have a significant impact on the robustness of the design once fabricated, and long-term reliability is virtually unknown without further characterization and testing. However, even with these limitations, this approach enables evaluation of sensor performance *and* to some extent manufacturability within a *set of constraints*, meaning that only designs that could potentially be manufactured within a particular process are evaluated.

Robust Design Phase: the primary goal in this phase is to thoroughly analyze and optimize the performance of the lead conceptual design(s) selected in the previous phase, using a variety of standard semi-analytical, analytical, and numerical techniques [42.26] to perform *optimization* studies, *tolerance-based design*, and *statistical design* to thoroughly evaluate the ability of the candidate designs to meet the required specifications. With modern design automation tools (e.g., MEMS+, Cadence; Fig. 42.14), *virtual manufacturing* analysis has become more widely available, and while only as accurate as the models used, it is gaining in importance as a necessary task, enabling optimization of device performance considering fabrication tolerances and process corners. Additionally, functional yield can be verified using a variety of techniques including *Monte Carlo* methods, and *sensitivity* analyses as shown in Fig. 42.15. At this stage, the focus of the designer is on quantifying the effects of *random variations* and *errors* on typical performance parameters (*sensitivity*, *cross-axis*, *dynamic range*, etc.) of the part [42.26] within the available process corners and operating ranges.

To the extent possible, determined of course by the manufacturing strategy decisions outlined above, it is beneficial to analyze the impact of manufacturing performance variations on the design *before* taping out any mask sets or fabrication of actual silicon. As discussed in Sect. 42.1, when an external foundry makes sufficient details regarding the established fabrication process(es) available, and the process is well controlled and characterized by available design rules, it becomes straightforward to predict functional yield and design sensitivities [42.26]. However, in cases where processes are not that well established, it is necessary to quantify achievable process limits before attempting such detailed designs [42.27]. The effort towards quantifying achievable limits overlaps with standardization efforts (either industry wide as has occurred in the IC industry and to a limited extent in MEMS, but also within a foundry) for measurement and characterization of material properties and process control parameters. The primary benefits are first-pass silicon success (design accuracy) and also more rapid development of new process technologies.

Package Design – The *co-design* of the package with the sensor element is critical, and preliminary assessment of the packaging options available must be initiated simultaneously with the *detailed design phase*. Ideally, the design team performs the package design as an *extension* of the sensor design to minimize the potential for redesign at a future stage. The performance of the part and potential sensor–package interaction failure modes that could arise (due to earlier sensor design decisions), as well as the package design and material selection and their impact on reliability and manufac-

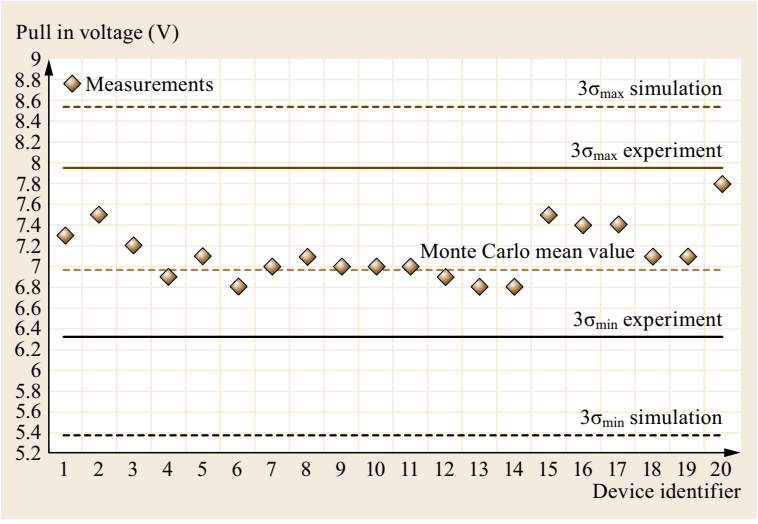


Fig. 42.15 Measured pull-in voltages for 20 suspended-plate variable capacitors in comparison with simulation results (after [42.24])

turability, has been described in detail elsewhere [42.23] but remains a critical factor, especially for the development of inertial sensors (Sect. 42.1). Today, low-cost plastic overmold packaging [42.28] is quite common (Fig. 42.16) for most MEMS inertial sensor products [42.18, 19]; however, in some cases, specialty packaging is needed (optical systems, Vital-Sign Monitoring (VSM) applications, or MEMS microphone packages), requiring additional development, and this necessary task must be factored into the overall development. Lastly, detailed decisions about the package design, especially pad configuration and placement, overall package dimensions and tolerances, maximum ratings, and even specific material choices, need to be made. At some level, similar choices face the design team, i. e., custom versus commercial off-the-shelf (COTS) package, highlighting that, in the development of MEMS products, using off-the-shelf packaging technologies will save development time and reduce overall risk.

Test (ATE) Development – The third important component of part design and one area that often gets overlooked during the design phase is the testability of the part, often referred to as design-for-test (DfT). The design team must also include robustness during final testing as part of the sensor–package design. In production, the part is handled by automated test equipment (ATE) and will experience unique thermomechanical forces due to handling and other back-end operations that may not be accounted for in the specification; For example, the packaged sensor could see higher than normal forces due to differences in the tolerances of the test socket compared with the package dimensional tolerances. Tolerance-based design practices must ensure survival of the component as well as no accumulated damage during this back-end stage.

Once the preliminary detailed design iteration has been completed and the design has successfully passed all available design rule checks (DRCs), the manufacturing stage can begin. In situations where process variability is still in question, the sensitivity of the design to these variations has to be quantified by collaborative effort between the design and process teams, resulting in the first tapeout of the design.

Manufacturing Phase: The manufacturing phase begins development of the MEMS sensor and ASIC fabrication process, as well as the assembly process for the product. Essential steps highlighted earlier (characterization of new process steps, etc.) involved in developing the sensor fabrication process must begin in parallel with the *detailed design* phase, and should include detailed process requirements and evaluation of available process platforms. Evaluation of available fabrication processes is necessary because there are basically three options with varying levels (as described in Sect. 42.1) of risk and cost structure: (a) available standard processes, generally found at captive foundries, (b) available standard processes with minor modifications, or even a well-established process transferred to another facility, or (c) a completely new process flow. One obvious consideration is ownership or control of the available process – as seen in Sect. 42.1 a majority of high-volume MEMS manufacturing process flows only exist in captive foundries, although this continues to evolve, and today there are multiple MEMS foundries that provide contract manufacturing services [42.29]. Each manufacturing platform process flow is composed of several unit process steps (that either add or remove material in a precise order [42.30]), and so in defining process requirements, it is necessary to factor in the tolerance limits and variations of these process characteristics such that specific performance parameters are constrained appropriately. In the situation where available platforms are either insufficient or require modification, validation of process development and integration experimental results must be performed to ensure manufacturability. Standardization efforts within the foundry should focus on identification and selection of relevant material properties and the conditions and metrology that influence these properties. Various approaches involving test structures (both in-line and monitor wafers) and material property measurements (off-line and small sample testing) have been employed with some success, but these approaches are by no means uniformly common in high-volume foundries. The goal of in-line measurements for MEMS process flows (similar to wafer acceptance test (WAT)

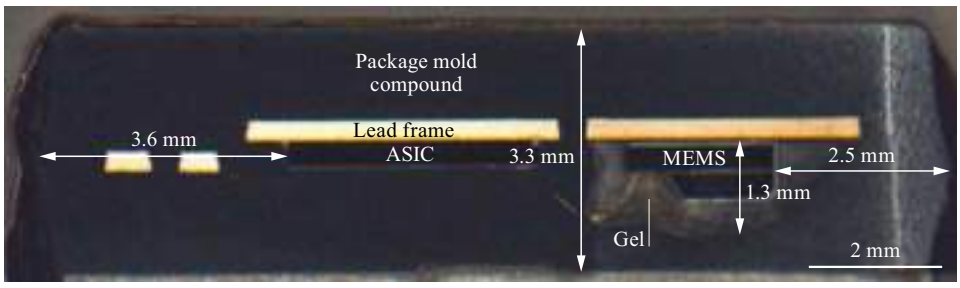


Fig. 42.16 Cross-section of a typical plastic overmolded MEMS sensor (courtesy of TechInsights Inc., formerly Chipworks)

structures used in IC fabrication) remains, and several industry groups including NIST continue to work on standardization. A table of material properties and process control parameters that are needed for MEMS is described in detail in [42.27] and replicated in Table 42.1 for convenience.

A manufacturing organization that is tasked with MEMS fabrication needs to ensure stable and reproducible fabrication processes to manufacture reliable products. As described above, fabrication of MEMS-enabled products requires in-depth analysis and review of product architecture and decisions regarding integration, size, and the process steps to use; the overall number of process steps and where they occur are among a large set of variables that affect the risk and cost of the end product, but more importantly determine success. A MEMS process is said to be stable when all of the parametric measures of the process have means, variances and distributions over time [42.31] leading to consistent chute yields over time. The path to achieving a high level of confidence in the manufacturing process is time consuming, requiring a step-by-step confidence building effort that begins with design of experiments (DOE) to identify significant process variables. The results of the DOE are then implemented on batches of validation material to confirm the outcome of the experiments and establish process baselines. Further DOEs may be needed to skew the primary variables to establish precise design rules that ultimately lead to a well-qualified, stable and reproducible process flow for future designs. It is generally true that, in the development or deployment of a completely new process flow, one is unlikely to achieve the low-ppm level of reliability within that process until a significant number of wafers have been processed. This sometime implies that the first designs released to the market using such a process may be vulnerable to field returns, but subsequent products using the same manufacturing flow will no doubt benefit. Over time, the process will mature to the point where it is possible to have extremely low field failure rates (a few ppm), and so the obvious challenge for the engineering team is to reach that level of field failures as soon as possible.

As described in earlier sections, the evolution of MEMS products has gone through several stages of integration. From the early days of integration of MEMS and signal processing side by side on the same die (e.g., iMEMS®) to discrete MEMS and ASIC die (e.g., integration through assembly) to the present-day wafer-level chip-scale packaging (WLCSP) integration of die by wafer bonding, the impetus has been to reduce size and cost by increasing the die per wafer (DPW) with fewer overall masks and process steps. Of course, this is dependent on the particular design and fabrication goals in each case, but design for manufacturability (*DfM*) is a powerful methodology that both the design and manufacturing teams need to adopt to improve manufacturing quality and yield of MEMS-enabled products.

Table 42.1 Material properties and process control parameters for MEMS fabrication (after [42.27])

Geometry	Primary material properties	Process control parameters	Reliability properties
Thickness Undercut Line width Line spacing Sidewall profile Step coverage Radii of curvature Voids Minimum feature sizes Misalignments Surface roughness Others	I. *Mechanical properties* Residual stress Residual stress gradients Anisotropic moduli Temperature dependence of moduli Nonlinear elasticity and plasticity Piezoelectric coefficients Others II. *Electrical properties* Dielectric permittivity Electrical conductivity Frequency dependence of properties Others III. *Thermal properties* Anisotropic coefficients of thermal expansion Specific heats, etc. Thermal conductivity Others	This list can include all parameters from the three neighboring lists, assessing statistical distributions, as well as variation of the mean value itself, variations within a die, die-to-die variation, wafer-to-wafer variation, batch-to-batch variation, machine-to-machine variation, and line-to-line variation	Loss mechanisms Material damping Contact between surfaces Cavity pressure hermeticity Failure mechanisms Fracture strength Fatigue strength Delamination/bond strength Creep UV compatibility Others

42.3 Stable Field Performance

Field stability refers to the ability of a MEMS product to maintain stable performance parametrics, i. e., meet datasheet specifications, over the life of the component. Robust design of MEMS devices, and precisely controlled manufacturing processes, are needed to ensure stable field performance. However, even well-designed and well-fabricated parts can fail in-use due to a variety of factors that are usually considered out of bounds in terms of the design or process capability as well as use case. Typically, detailed analysis of the potential failure modes is performed using failure modes and effects analysis (FMEA) during the development cycle to try to anticipate and mitigate such failures. Clearly, there exists a strong linkage between design, manufacturing, and stable field performance [42.23], but a large proportion of MEMS product failures generally fall into two categories – electrical (e.g., dielectric charging and electrostatic discharge (ESD)) and mechanical (e.g., stiction, shock, and creep) failures. In this section, some of the primary failure modes within these categories are described, along with typical approaches to mitigate or prevent such failures.

42.3.1 Electric Field Stability

Due to the increased complexity of the ASIC signal chains in modern MEMS sensor products, field returns due to anomalous performance of specific ASIC functionality is not uncommon. Usually, these types of field returns are related to the sensitivity of the design to a process variable that may shift over time or from one lot to the next. By and large, the root cause of these failures can be reproduced by simulation and appropriate steps (and mitigated through either test-based screens or redesign), but for MEMS products, with the additional degree of freedom due to the sensor and package, the most common in-field failures experienced include dielectric charging and electrical breakdown (electrostatic discharge (ESD) or electrical overstress (EOS)) [42.23]. These electrical failure modes can play a significant role in a component's lifetime, and can cause unstable performance through either catastrophic failure or performance drift. The task of preventing these potential failure modes begins in the detailed design phase, but proper control of fabrication materials and surface chemistry control is also absolutely necessary for their mitigation.

Dielectric Charging: High electric fields (10^8 V/m) experienced across bulk and surfaces within some MEMS sensors can lead to charge injection resulting in trapped charges, and leakage currents, which are typically strong functions of temperature and humidity. Some dielectrics used in electrostatic MEMS are tailored to have low stress (as they are part of moving elements and have to have predictable behavior), and achieving this often requires a tradeoff with film electrical properties (reduced breakdown voltage, etc.). Charging of dielectrics is a complex phenomenon at the atomic scale and can be highly dependent on film composition and stoichiometry, deposition rates, times at temperature (as well as heating/cooling rates), and the specific techniques used for making the film. Mitigation of such charging effects involves a combination of design approaches such as geometry changes (removing dielectrics from certain areas, shielding, etc.) and material changes (i. e., composition of dielectrics) [42.32]. An example is deposition or growth of a thin conductive layer on top of a dielectric to bleed off surface charge, but this requires careful consideration of the design configuration and placement of various elements. Charging results in unpredictable spurious electrical fields that can disturb the stability of the moving MEMS element, leading to drift over time or even stiction. In accelerometers and gyros, this mechanism causes offset drift over time, leading to gradual loss of performance and eventually failure.

Electrostatic Discharge (ESD) and Electrical Overstress (EOS) – This refers to the sudden or instantaneous transfer of charge due to large potential differences between sensitive nodes in the component and can occur through a solid dielectric or through the gas (in between layers). In conventional ICs, protection against such failures is well established, but much less so for MEMS where ESD can cause both electrical (Fig. 42.17) and mechanical damage (pull-in or breakage of sensor element). Often, the mechanism can lead to catastrophic failure of the MEMS component, although less damaging erratic performance has also been observed in inertial sensors. Currently, techniques are available to quantitatively measure (human body model (HBM), machine model (MM), and charge discharge model (CDM)) and quantify device susceptibility, but they may not capture specific in-use situations encountered in the field.

Although there are clearly many design practices [42.23] that can be implemented to mitigate the harmful effects of ESD/EOS, from a manufacturing standpoint, the main choices are again geometrical (well-characterized design rules for spacings) and/or material (choice of appropriate materials). It is important to also mention the role of proper handling in the back-end of the production line so that opportunities for triboelectric charge generation are completely eliminated.

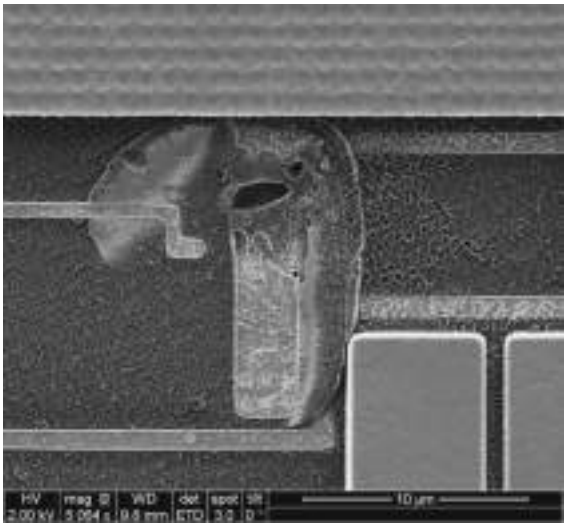


Fig. 42.17 EOS damage at ESD protection pad (courtesy of Analog Devices Inc.)



Fig. 42.18 Pinned and free-standing cantilever beams (reprinted with permission from [42.36, 37]. © 1993 IEEE)

The electric field stability of MEMS devices is also known to exhibit failure modes related to galvanic corrosion, anodic oxidation, pad damage, electrical discharge, etc., at either sensor die or package level, and these can be mitigated by proper control of fabrication steps (e.g., metal pad corrosion).

42.3.2 Mechanical Field Stability

Besides electrical field stability concerns, mechanical field stability concerns such as shock response, bias stability, noise stability etc. are often encountered in MEMS devices but the most common field stability issue in inertial MEMS is stiction. It is one of the primary tribological failure mechanisms in MEMS [42.33, 34], occurring when a suspended structure exhibits unexpected adhesion to adjacent contacting surfaces [42.34]. Although stiction usually manifests during fabrication, under certain circumstances it can also occur during use, i. e., *in-use* stiction.

Adhesion of two adjacent MEMS surfaces in contact occurs due to a combination of a variety of physical mechanisms such as van der Waals forces, electrostatic forces (trapped charge), and liquid capillary forces or a combination of these forces [42.35]. The use of wet chemical processes in MEMS fabrication is known to leave residual ions, dangling bonds, and even monolayer films of water and/or polymers. In such cases, adhesion predominantly occurs due to surface tension forces [42.38], although other triboelectric mechanisms can also influence this mechanism.

During component usage, the physical and chemical state of the contacting surfaces is a critical factor for the development of adhesive forces. A high-energy surface, which can evolve due to contamination, moisture, and film out-gassing during field use because of either known or unplanned failure modes (the surface energy can change over time due to repeated contact or due to a change in moisture levels, ambient gas, or surface roughness), can cause stiction to occur, leading to catastrophic failure of the component and potentially the system it resides within. Within the manufacturing process, it therefore becomes necessary to consider the balance and control of the process variables more precisely to prevent stiction [42.35]. In most inertial and optical MEMS devices where such failure modes are to be prevented, a combination of design and process modifications can be deployed to ensure mitigation of this failure mode. On the design side, it is relatively common to include bumps, stoppers, and other geometrical features to minimize contact surface areas. Additionally, the sensor element mechanics can be designed to ensure sufficient tolerance to applied voltages or displacements. On the process side, besides proper sequences of process steps to ensure low-energy surfaces, antistiction coatings are commonly used to further improve performance and eliminate stiction.

42.4 Internet of Things (IoT)

The IoT refers to the third wave of technology development, as depicted in Fig. 42.19. The mid 1950s saw the first wave with the development of solid-state devices and mainframe computers for businesses. The 1960–1970s saw a major growth period, which then began to level off as businesses and technology matured. The second wave was the development of personal computers, laptops, smartphones, and tablets. As the chart shows, there was a very rapid development of these devices starting in the 1980s and maturing in the 2000s. The connectivity offered by these devices, especially smartphones, has opened up a new area of technology development, which we call the Internet of Things or IoT. With the development of the internet and personal electronic devices, most of the modern world is now connected to the web and one another, providing ready access to information and the sharing of data. This provided a new opportunity to automate much of our daily activities, the methods we use to control our daily lives, and how we manage our personal tasks such as our health, finances, the efficiency of our homes, our communications, and work. To support this new paradigm of living, we will need ubiquitous sensing and signal processing to monitor our health and our environment, which can interface with our personal electronic devices, and the *intelligent* network to interpret and act on all this data as well as to provide a structure for the efficient use of this data. This will require multiple sensors, signal processing, wireless communication, cyber security, data analysis, storage, and applications as well as algorithms to manage and interpret all this data so that this information can help guide meaningful decisions.

Among the major areas where sensors will be used, smart agriculture, smart homes, and the smart environment are expected to be the industries seeing the greatest impact. The focus will be on developing systems to monitor our food for freshness, the cleanliness of the air we breathe, and the water we drink, as well as to detect hazardous chemicals or biological agents. In agriculture, soil quality, fertilizer concentrations, and insecticide levels will be monitored continuously to grow crops better and more efficiently. In the case of such systems deployed in homes, the goal would be to increase our energy efficiency by automatically turning on lights and other appliances in the most efficient manner and also to monitor our personal health in real time. It is predicted [42.2] that there will be 50 trillion connected devices with multiple sensors in the next 25 years. This includes wearables leading to e-health enabled by the mobile market as well as remote, fully contained systems for monitoring soil, clean food, air, water, roads, and building infrastructure. In our homes and workplaces, sensors will be used to monitor all our appliances, energy usage, security, etc. that are remotely accessed and monitored to reduce waste and improve their efficiency. In addition, a significant majority of these sensors will need energy-harvesting solutions for their (rechargeable) power source due to the size and cost of replacing disposable batteries.

All of this monitoring will demand a huge number of sensors of various types, including chemical, magnetic, inertial, and pressure sensors to name but a few, and MEMS manufacturing technology will need to scale significantly to meet the demand from such applications. To make these sensors, we need high-volume and low-cost sensor manufacturing technologies and likely novel three-dimensional (3-D) printer-based technologies to meet this huge demand at low cost.

While identification of this new massive business opportunity has generated lots of excitement and exploratory research and development work, there are significant challenges, specifically in evolving technologies and material science advances to provide solutions that meet the cost, manufacturability, reliability, and performance requirements. These challenges will require huge resources or individual companies to accurately predict narrow areas to focus on to develop new, profitable business directions. Over time, the materials and manufacturing technology approaches and strategies will determine winners and losers; at this point of time, the risks are still high, and it is not clear which direction IoT deployment (industry) will take to ensure success.

The IoT will require complete standalone system solutions; autonomous is another term that best describes such systems. So, in addition to the sensor, the

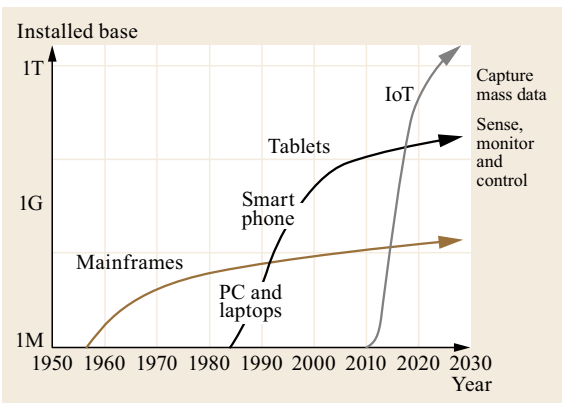


Fig. 42.19 Three waves of technology development. The IoT is predicted to have 1000× more devices compared with the second wave

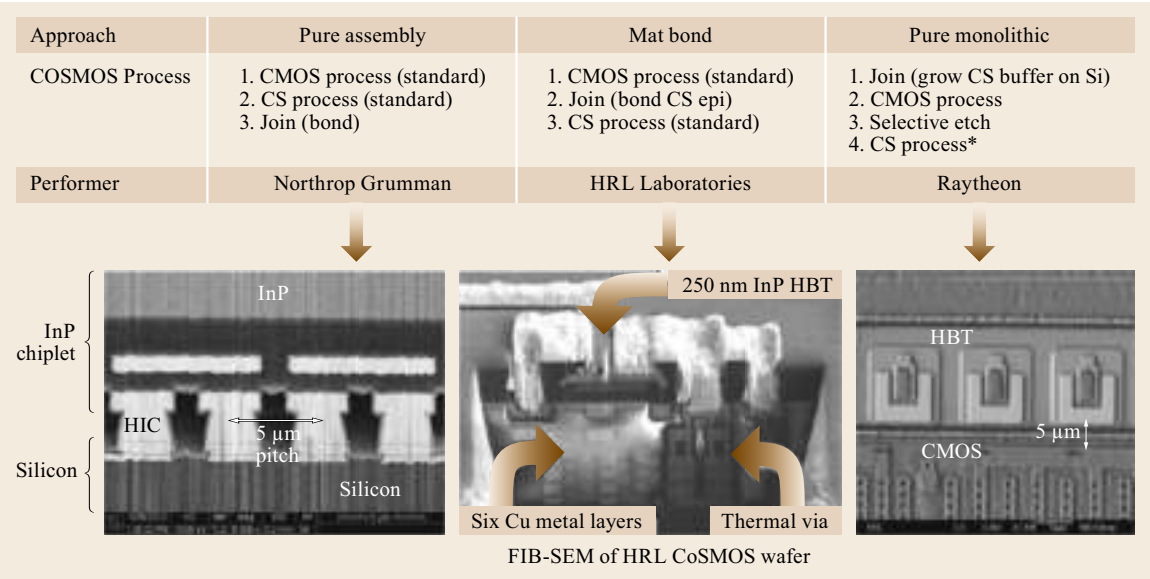


Fig. 42.20 Heterogeneous integration techniques funded by the DARPA COSMOS program (*CS*: chip scale)

system will need signal processing, wireless communication to your mobile device, and a renewable power source for implantable or remote location. MEMS devices and manufacturing technologies can be retooled and scaled to support this new sensor revolution. Many MEMS devices can be repurposed, and the fundamental science used for MEMS such as capacitive, piezoresistive, and piezoelectric sensors can be used to make a vast number of new IoT sensors. Also, MEMS unit processes such as bonding, deep silicon etching, depositing thick and thin films, and interfacing the device with the environment can be used as unit process steps in the development of the next generation of sensors, which suggests that existing MEMS manufacturers should have a head-start if not a clear advantage. In short, the requirements for the new IoT era that will be the main technology drivers are low-cost high-volume manufacturing, ultralow-power signal processing and networking, energy harvesting, and energy storage for the power source and algorithms and software for data management, providing secure error-free standalone sensor systems.

42.4.1 Heterogeneous Integration

Initially, new sensors and systems for the IoT will be silicon based, building on the experience to date in MEMS technology, wireless communication, and recent development efforts for energy harvesters.

The IoT will require rethinking of conventional semiconductor packaging and manufacturing approaches. The emerging concept of *heterogeneous integration* will be an enabling technology for the IoT. Heterogeneous integration refers to technologies which closely combine different types of semiconductor devices in ways that push far beyond what is now done using traditional semiconductor packaging techniques; For example, future heterogeneous integration technologies will enable combination of III–V semiconductor devices with silicon semiconductor devices with micron-scale placement accuracy and interconnect parasitics approaching that of monolithic integration. This will enable solutions that use the *right junction for the function*; that is, the optimum semiconductor material will be used for each particular function in a system. ICs will be manufactured which consist of, for example, GaAs or InP low-noise and radiofrequency (RF) power amplifiers, highly scaled CMOS for data converters, logic, and memory, and piezoelectric materials for sensors, all integrated together into a single piece of silicon in a single package. Cost-effective realization of this vision will enable IoT nodes with size, weight, and power that are impossible with traditional monolithic manufacturing and die packaging technologies.

Heterointegration will be needed to integrate all these different technologies into one system solution. Heterointegration is the process of transferring devices from one wafer with devices built using one technology to another wafer with devices built using a completely different technology so that the individual devices that make up the complete system can be integrated together. Heterointegration takes the system in package (SIP) approach for individual integration of different chips to the wafer level.

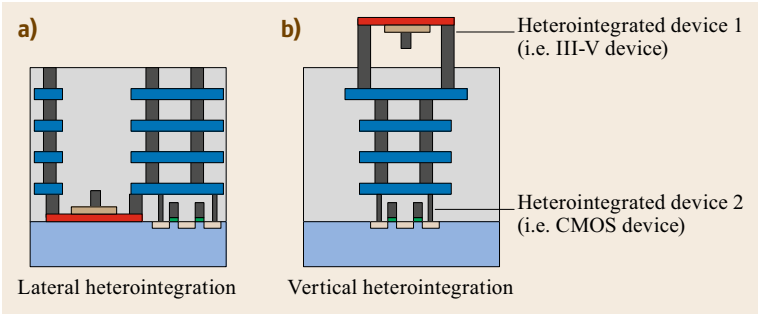


Fig. 42.21a,b Pictorial depiction of **(a)** lateral and **(b)** vertical heterointegration

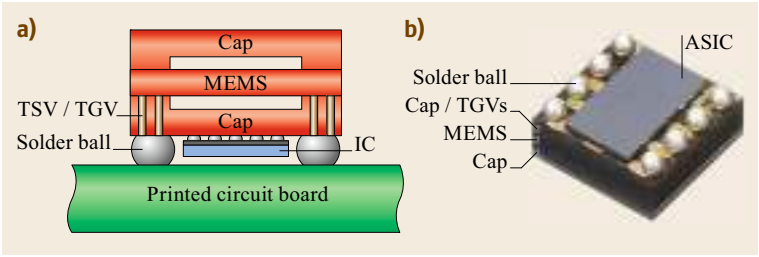


Fig. 42.22a,b Vertically integrated wafer bonding to integrate a MEMS device with a CMOS signal processing chip at wafer level (after [42.39]); **(a)** cross section, **(b)** image of diced device (*TGV*: through glass via)

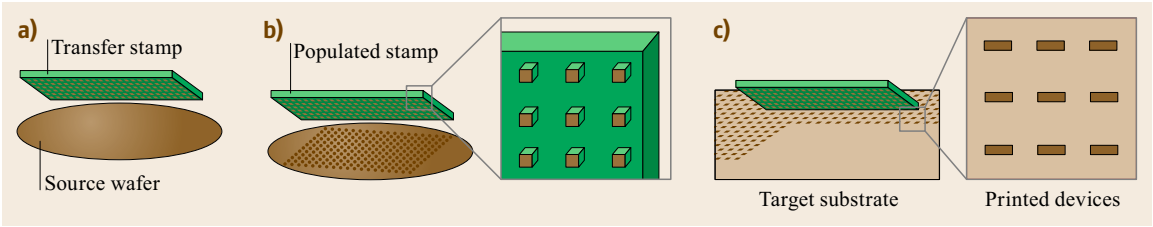


Fig. 42.23a–c Depiction of transfer printing process. A polymer stamp **(a)** is used to pick up large arrays of thin devices **(b)** and place them on a second wafer **(c)**

Heterointegration can take devices from a processed III–V wafer and integrate them with a foundry CMOS wafer. Another example could be a MEMS device on a MEMS wafer with either a processed III–V wafer, or both a III–V and a MEMS wafer to a foundry CMOS wafer. Heterointegration is highly flexible in enabling integration of different devices to provide a finished, self-contained product solution; For example, it can combine RF and high-frequency analog devices with foundry CMOS or any other unique process such as high-speed SiGe or InP devices, GaN high-speed high-power devices, photonics, or the new emerging technologies for energy storage or energy harvesters. So, developing heterointegration processes to mix many different types of chip at wafer level is one approach to develop a low-cost high-performance self-contained solution for emerging IoT products. Another advantage of heterointegration over SIP is that it allows microscale integration of different devices to reduce size and parasitic capacitance, resistance, and inductance between devices. With SIP integration, the minimum device sizes are relatively large and they have high interconnect parasitics. Heterointegration will enable integration of smaller devices and through parallel processing provide higher accuracy and matching of device placement and significantly lower manufacturing costs.

Many groups are pursuing different types of *heterogeneous integration*. The US Defense Advanced Research Program Agency (DARPA) funded a variety of these approaches in the COSMOS program. The different approaches taken in that work are shown in Fig. 42.20. Three different approaches were pursued. Raytheon pursued a heteroepitaxy approach to enable monolithic integration of InP devices with CMOS. Northrop Grumman pursued an approach which combines various *chiplets* on an interposer wafer. Hughes Research Labs used a wafer bonding approach. In all cases, III–V devices were joined with silicon devices with micron-scale spacing between the III–V device,

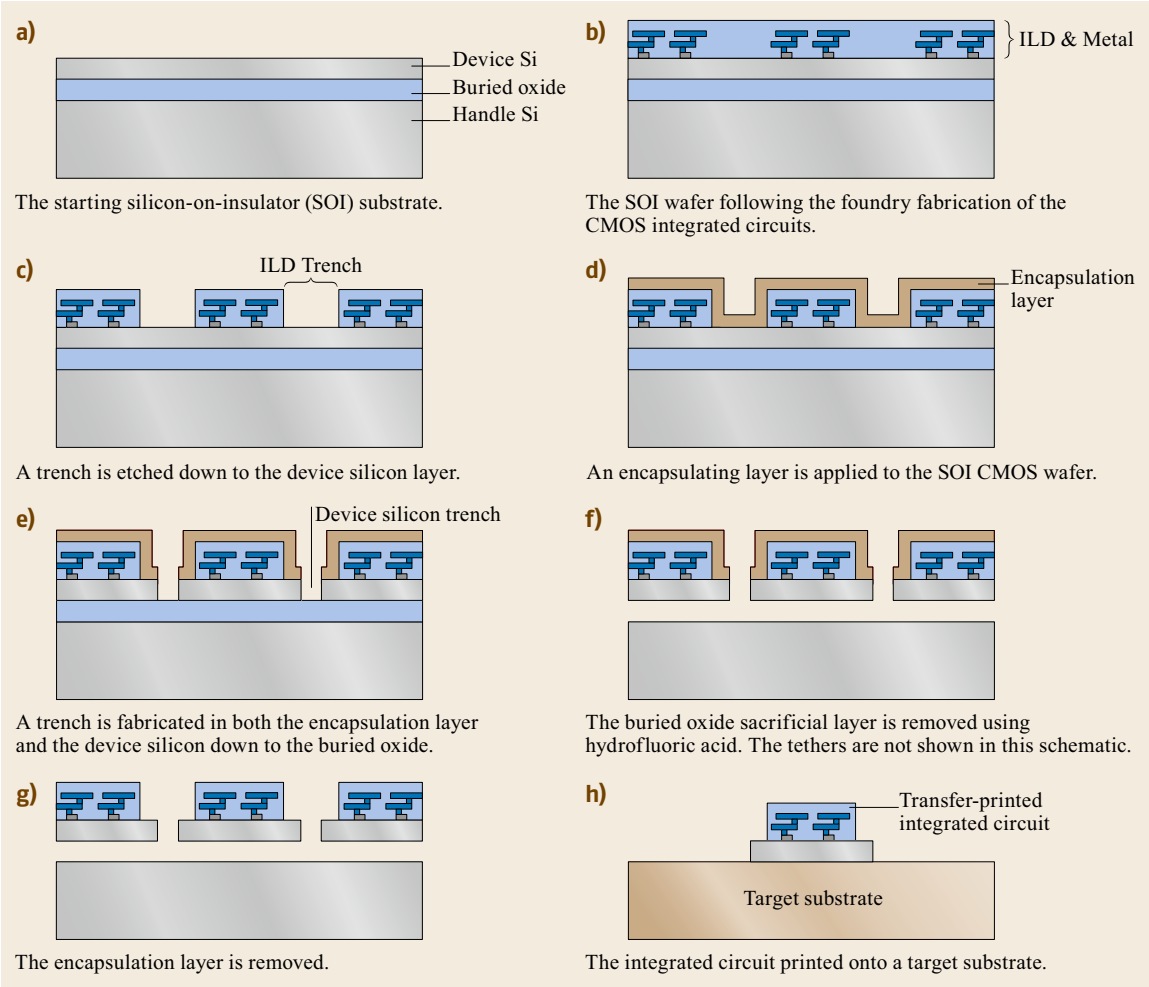


Fig. 42.24a–h Depiction of use of the transfer stamping process to transfer devices from an SOI wafer to a quartz wafer for display back-plane applications (*ILD*: inter level dielectric)

and with interconnect parasitics which approach what can be achieved in a monolithic process.

There are two basic modes of heterointegration, i. e., lateral (side by side) and vertical (stacked), as shown in Fig. 42.21.

Wafer-to-wafer bonding with TSVs is one way to integrate a MEMS wafer with submicron CMOS (Fig. 42.22). Here, Al-Ge eutectic bonding is used to form the electrical interconnect between the two different wafers, and a second major advantage is that it serves to form the cap for the MEMS device as well. TSVs through the CMOS or MEMS wafer act as input and output connections for the vertically heterointegrated device. The main advantages of this style of wafer bonding are close lateral alignment and the ability to form a large

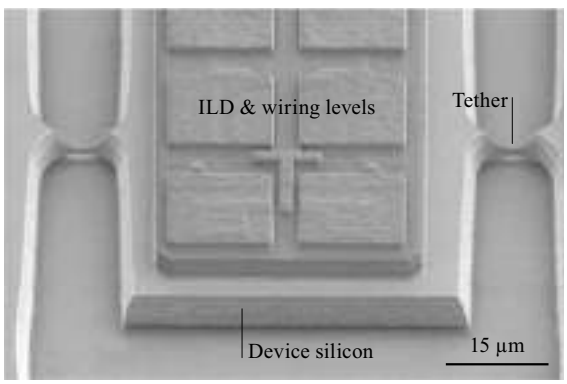


Fig. 42.25 SEM of a tethered, fully processed CMOS device (courtesy of Semprius Inc.)

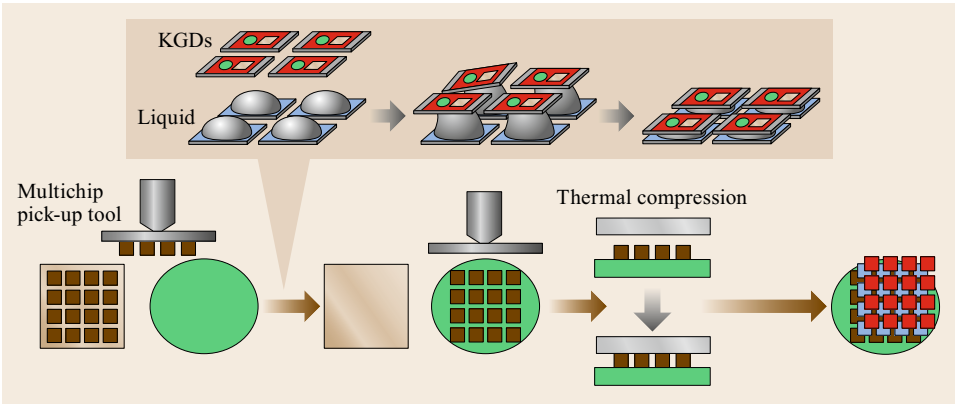


Fig. 42.26 Self-assembly-based multi-chip-to-wafer bonding (SA-MCtW)

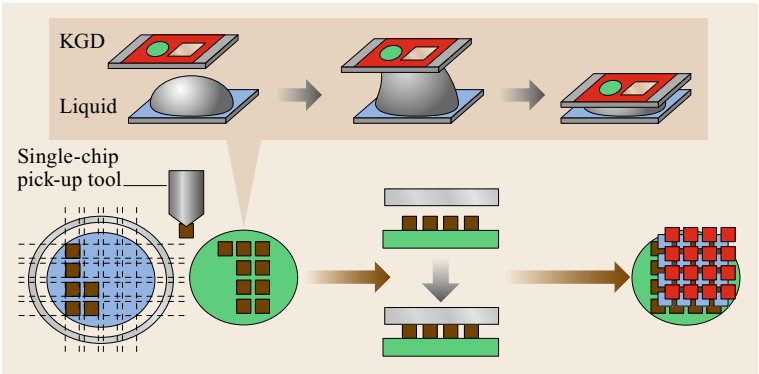


Fig. 42.27 Hybrid self-assembly-based chip-to-wafer bonding (HAS-CtW)

number of interconnects, but the main disadvantage is that one cannot easily integrate wafers of different sizes, and that the overall size is now determined by the size of the largest active device (which could leave valuable space unused).

This is a great example of vertical heterointegration. Another technique that is worth considering is microscale transfer printing [42.40], where a polymer array is used to pick up chiplets from one wafer and transfer them to a second wafer, as shown in Fig. 42.23.

Transfer stamping is another variant of this technology, as depicted in Fig. 42.24 and Fig. 42.25. Both approaches take chiplets from one wafer and transfer them to receptor regions on a second wafer.

Other approaches involve die to wafer transfer, where a pick and place tool can take individual die and flip them to align and bond with interconnects on the acceptor wafer.

This could be scaled up to take an array of singulated die and flip and bond to the acceptor wafer. Here once again, additional processing is performed to form electrical interconnects between the die that has been transferred and the die on the receiving wafer.

A variant on die to wafer transfer is a process that uses self-assembly techniques [42.41]. This technique requires special treatment of the acceptor wafer to obtain regions with specific surface tension to attract the chiplets and then uses thermocompression for bonding the die to the wafer. Two examples are shown in Figs. 42.26 and 42.27. One technique developed by Tohoku University [42.42] is to place a droplet of water on the receiving wafer where the chiplets are to be placed and use a pick and place tool to take singulated die from the donor wafers and place them on the water droplets. The chiplets self-align on the water droplets due to the surface tension of the water droplet and are then fixed in place using a thermocompression bond.

There are other ways to achieve heterointegration, but they lie outside the scope of this chapter. The aim of this discussion of heterointegration is to show that wafer-level bonding is currently used for heterointegration of MEMS with foundry CMOS, and that this and other heterointegration techniques are expected to gain momentum with the development of new sensors for the IoT.

42.5 Conclusions and Outlook

In this chapter we have discussed the evolution of MEMS high-volume manufacturing at ADI and some other high-volume MEMS manufacturing companies. There are two basic approaches. One is to make the MEMS device using a separate process and wafer and then connect the MEMS die with a signal processing die in a two- or multichip package using wire bonding. Future trends will be to use heterointegration to connect the MEMS die with the signal processing die through conductive bonding. Both the MEMS and CMOS processing can be done internally or using an external MEMS foundry. The second approach is monolithic integration of the MEMS device with the CMOS using fusion bonding followed by postprocessing. This approach generally requires that all the processing be done either internally or using an external foundry. The InvenSense process is a hybrid between these two approaches. The MEMS industry has been around for over 30 years now, and we are beginning to see commoditization of many MEMS products such as inertial sensors, microphones, and pressure sensors. Just as the semiconductor industry was the core technology to build MEMS devices, MEMS technology will become the core technology for new IoT sensors and devices.

References

42.1 J. Martin: High volume manufacturing and field stability of MEMS products. In: *Springer Handbook of Nanotechnology* (Springer, Berlin, Heidelberg 2010)

42.2 TSensors Summit – MEMS Industry Group Enterprise, http://www.tsensorssummit.org/Resources/TSensors%20Roadmap%20v1.pdf

42.3 T.A. Core, W.K. Tsang, S.J. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol **36**(10), 39–47 (1993)

42.4 K. Nunan, G. Ready, J. Sledziewski: LPCVD & PECVD operations designed for iMEMS sensor devices, Vac. Technol. Coat. **2**(1), 26–37 (2001)

42.5 K.H. Chau, R.E. Suloff: Technology for the high-volume manufacturing of integrated surface-micromachined accelerometer products, Microelectron. J. **29**, 579–586 (1998)

42.6 W. Kuenel, S. Sherman: A surface micromachined silicon accelerometer with on-chip detection circuitry, Sens. Actuators Phys. **45**(1), 7–16 (1994)

42.7 M. Schirmer, R. Goggin, P. Fitzgerald, D. Rohan, J.-E. Wong: MEMS Switch Capping and Passivation Method. US Patent 8124436 (2012)

42.8 L.E. Felton, P.W. Farrell, J. Luo, D.J. Collins, J.R. Martin, W.A. Webster: MEMS Capping Method and Apparatus. US Patent 6893574 (2005)

42.9 J.R. Martin: Process for Wafer Level Treatment to Reduce Stiction and Passivate Micromachined Surfaces and Compounds Used Therefor. US Patent 6674140 (2004)

42.10 A. Solanki, K. Prasad, K. Nunan, R. Oreilly: Comparing process flow of monolithic CMOS-MEMS integration on SOI wafers with monolithic BiMOS-MEMS integration on Silicon wafers, iMEMS fabrication incorporating MEMS and electronics on a single chip. In: *Proc. 53rd IEEE Int. Midwest Symp. Circuits Syst.* (2010) doi:10.1109/MWSCAS.2010.5548876

42.11 S. Lewis, S. Alie, T. Brosnihan, C. Core, T. Core, R. Howe, J. Geen, D. Hollocher, M. Judy, J. Memishian, K. Nunan, R. Paine, S. Sherman, B. Tsang, B. Wachtman: Integrated sensor and electronics processing for >108 iMEMS inertial measurement unit components. In: *Proc. IEEE Int. Electron Devices Meet* (2003) doi:10.1109/IEDM.2003.1269435

42.12 G.K. Fedder, J. Chae, K. Najafi, T. Denison, J. Kuang, S. Lewis: Monolithically integrated inertial sensors. In: *CMOS-MEMS*, ed. by O. Brand, G.K. Fedder (Wiley-VCH, Weinheim 2005)

42.13 D. Hollocher, X. Zhang, A. Sparks, S. Bart, W. Sawyer, P. Narayanasamy, C. Pipitone, J. Memishian, H. Samuels, S.-L. Ng, R. Mhatre, D. Whitley, F. Sammoura, M. Bhagavat, C. Tsau, K. Nunan, M. Judy, M. Farrington, K. Yang: A very low cost, 3-axis, MEMS accelerometer for consumer applications. In: *Proc. IEEE Sens* (2009) pp. 953–957 doi:10.1109/ICSENS.2009.5398189

42.14 T.K. Nunan: Polysilicon Deposition and Anneal Process Enabling Thick Polysilcon Films for MEMS Applications. US Patent 7754617 (2008)

42.15 S. Sood: *CMOS Compatible Hermetic Wafer Level Packaging for Inertial MEMS* (SUSS MicroTech Inc, Sunnyvale 2013)

42.16 MEMS Industry Group: http://www.semi.org/en/msig-information-hub

42.17 MEMS Industrial Group's Foundry Engagement Guide Steering Committee: http://memsfoundry.wikia.com/wiki/MEMS_Foundry_Engagement_Guide_Wiki

42.18 Yole: Yole MEMS & Sensors Industry Reports, http://www.i-micronews.com/reports/mems-sensors-report.html

42.19 SystemPlus Consulting: MEMS Accelerometer Reports, (2016) http://www.systemplus.fr/device-type/accelerometer

42.20 D.M. Anderson: *Design for Manufacturability* (CRC Press, Boca Raton 2014)

42.21 M.G. da Silva, R. Giasolli, S. Cunningham, D. DeRoo: MEMS design for manufacturability. In: *Sensors Expo, Boston* (2002)

42.22 P. Doe: New approaches needed to sustain MEMS growth, (2015) http://www.semi.org/en/node/56321

(originally published in EE Times)
42.23 A.L. Hartzell, M.G. da Silva, H. Shea: *MEMS Reliability* (Springer, New York 2011)
42.24 G. Schropfer, M. McNie, M.G. da Silva, R. Davies, A. Rickard, F.X. Musalem: Designing manufacturable MEMS in CMOS compatible processes – Methodology and case studies, Proc. SPIE (2004) doi:10.1117/12.544971
42.25 B. Romanowicz, M.H. Zaman, S.F. Bart, V.L. Rabinovich, I. Tchertkov, C. Hsu, J.R. Gilbert: A methodology and associated CAD tools for Support of concurrent design of MEMS. In: *VLSI: Systems on a Chip*, ed. by L.M. Silveira, S. Devadas, R.A. Reis (Springer, New York 2000)
42.26 S. Maity, S. Liu, S. Rouvillois, G. Lorenz, M. Kamon: Rapidly analyzing parametric resonance and manufacturing yield of MEMS 2D scanning mirrors using hybrid finite-element/behavioral modeling, Proc. SPIE (2014) doi:10.1117/12.2041067
42.27 M.G. da Silva, S. Bouwstra: Critical comparison of metrology techniques for MEMS, Proc. SPIE (2007) doi:10.1117/12.714852
42.28 L.E. Felton, N. Hablutzel, W.A. Webster, K.P. Harney: Chip scale packaging of a MEMS accelerometer. In: *Proc. 54th Electron. Compon. Technol. Conf.* (2004) doi:10.1109/ECTC.2004.1319439
42.29 MEMS Industry Group: Foundry Engagement Guide, MEMS Industry Group (2016) http://www.memsindustrygroup.org/?page=FoundryEngagement
42.30 M.J. Madou: *Fundamentals of Microfabrication* (CRC Press, Boca Raton 2000)
42.31 T. Rogers, N. Aitken, K. Stribley, J. Boyd: Improvements in MEMS gyroscope production as a result of using in situ, aligned, current-limited anodic bonding, Sens. Actuators A **123/124**, 106–110 (2005)
42.32 M.R. Douglass: DMD reliability: A MEMS success story, Proc. SPIE (2003) doi:10.1117/12.478212
42.33 R. Maboudian, R.T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B **15**, 1 (1997)
42.34 C.H. Mastrangelo: Surface force induced failures in microelectromechanical systems. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer Academic, Dordrecht 1998) pp. 367–395
42.35 R. Maboudian, R.T. Howe: Stiction reduction processes for surface micromachines, Tribology Lett **3**(3), 215–221 (1997)
42.36 C.H. Mastrangelo, C.H. Hsu: Mechanical stability and adhesion of microstructures under capillary forces: Part I. Basic theory, J. Microelectromech. Syst. **2**(1), 33–43 (1993)
42.37 C.H. Mastrangelo, C.H. Hsu: Mechanical stability and adhesion of microstructures under capillary forces: Part II. Experiments, J. Microelectromech. Syst. **2**(1), 44–55 (1993)
42.38 W.M. van Spengen, R. Puers, I. De Wolf: A physical model to predict stiction in MEMS, J. Micromech. Microeng. **12**, 702–713 (2002)
42.39 A.C. Fischer, F. Forsberg, M. Lapisa, S.J. Bleiker, G. Stemme, N. Roxshed, F. Niklaus: Integrating MEMS and ICs, Microsyst. Nanoneng. (2015) doi:10.1038/micronano.2015.5
42.40 C.A. Bower, E. Menard, S. Bonafede, S. Burroughs: Transfer-printed microscale integrated circuits. In: *Proc. Electron. Compon. Technol. Conf.* (2009) doi:10.1109/ECTC.2009.5074077
42.41 A. Technology Corporation: *Fluidic Self-Assemby White Paper* (Allen Technology Corporation, Addison 1999)
42.42 T. Fukushima, H. Hashiguchi, J. Bea, Y. Ohara, M. Murugesan, K.-W. Lee, T. Tanaka, M. Koyanagi: New chip-to-wafer 3D integration technology using hybrid self-assembly and electrostatic temporary bonding. In: *Proc. IEEE Int. Electron Dev. Meet. (IEDM)* (2012) doi:10.1109/IEDM.2012.6479157

43. Packaging and Reliability Issues in Micro/Nano Systems

Yu-Chuan Su, Jongbaeg Kim, Yu-Ting Cheng, Mu Chiao, Liwei Lin

The potential of microelectromechanical systems (MEMS)/nanoelectromechanical systems (NEMS) technologies has been viewed as a revolution comparable to or even greater than that of microelectronics. The scientific and engineering advancements in MEMS/NEMS could enable applications that were previously unthinkable, from space systems, environmental instruments, to appliances for use in daily life. As presented in previous chapters, development of core MEMS/NEMS processes has already demonstrated many commercial applications as well as potential for advanced functionality in the future. However, low-cost and reliable packaging for protection of these MEMS/NEMS products remains a very difficult challenge. Without addressing these packaging and reliability issues, no commercial products can be sold on the market. Packaging design and modeling, packaging material selection, packaging process integration, and packaging cost are the main issues to be considered when developing a new MEMS packaging process. In this chapter, we present the fundamentals of MEMS/NEMS packaging technology, including packaging processes, hermetic and vacuum encapsulation, wafer-level packaging, three-dimensional (3-D) packaging, polymer-MEMS assembly and encapsulation, thermal issues, packaging reliability, and future packaging trends. Specifically, development of MEMS packaging will rely on successful implementation of several unique techniques, including packaging design kits for system and circuit designers, low-cost and high-yield wafer-level, chip-scale packaging techniques, effective testing techniques at wafer level to reduce overall testing costs, and reliable fabrication of an interposer [43.1] with vertical through interconnects for device integration.

43.1 **Introduction to MEMS Packaging** 1505
43.1.1 MEMS Packaging Fundamentals 1506
43.1.2 Contemporary MEMS Packaging Approaches 1508
43.1.3 Bonding Processes for MEMS Packaging Applications 1508
43.2 **Hermetic and Vacuum Packaging** 1511
43.2.1 Integrated Micromachining Processes 1511
43.2.2 Postpackaging Processes 1513
43.2.3 Localized Heating and Bonding Processes 1515
43.3 **Emerging Packaging Approaches** 1518
43.3.1 Wafer-Level Packaging 1518
43.3.2 3-D Packaging 1521
43.3.3 Polymer-MEMS Packaging 1524
43.4 **Thermal Issues and Packaging Reliability** 1526
43.4.1 Thermal Issues in Packaging 1526
43.4.2 Packaging Reliability 1528
43.4.3 Long-Term and Accelerated MEMS Packaging Tests 1530
43.5 **Future Trends and Summary** 1533
References 1534

43.1 Introduction to MEMS Packaging

MEMS are miniaturized systems of micrometer to millimeter size, integrating mechanical, chemical, or biomedical features with IC circuitry for sensor or actuator applications [43.2]; For example, pressure [43.3], temperature, flow [43.4], acceleration [43.5], gyroscopic [43.6], and chemical sensors [43.7] can be fabricated using MEMS technologies for sensing applications, while fluidic valves [43.8], pumps [43.9], and inkjet printer heads are examples of actuation devices for medical, environmental, office, and industrial applications. Silicon is typically used as the primary substrate material for MEMS fabrication, because of its unique electrical, thermal, and mechanical properties as well as its easy micromachining in batch processing and

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_43

potential incorporation with microelectronic circuitry mostly using conventional semiconductor manufacturing processes and tools. The resulting MEMS devices offer the advantages of smaller size, lighter weight, lower power consumption, and lower fabrication cost compared with existing macroscale systems offering similar functionality. With the advances of MEMS fabrication technology in recent decades, the MEMS market at component level now exceeds 5 billion, driving end-product markets of more than 100 billion [43.10].

Nevertheless, the road to commercialization of MEMS does not look as promising as expected. Many industrial companies have taken advantage of MEMS technology due to the high production volumes and high added value resulting from product integrations. Therefore, cost-efficiency has become the major factor driving MEMS toward commercialization. Several MEMS devices have been developed for and applied in the automotive industry and information technology field, dominating the MEMS market due to their high production volumes. However, custom-designed MEMS products remain very diverse, aiming for different applications, and their initial costs in small- to medium-scale production are still much higher than market-acceptable levels. In this regard, high packaging and testing costs have hindered MEMS commercialization. Furthermore, based on past experience in the IC industry, the cost of packaging processes is about 30% and can sometimes be more than 70% of total production costs. MEMS packaging processes are expected to be even more costly, because of the challenging and stringent packaging issues related to the MEMS components, in addition to microelectronic circuitry, in a typical MEMS product [43.11].

43.1.1 MEMS Packaging Fundamentals

Sealing or encapsulation is an important step in either integrated circuit (IC) or MEMS packaging processes, to protect devices during operation. In traditional IC packaging procedures, the overall packaging steps often involve [43.12, 13]:

1. Wafer dicing
2. Pick-and-place
3. Electrical connections, such as wire bonding
4. Plastic molding or housing for the sealing process.

With the increasing requirement for high-performance and multifunctional consumer electronic products, IC packaging processes have incorporated more complex designs and advanced fabrication technologies, such as Cu interconnects [43.14], flip-chip bonding [43.15], ball grid array [43.16], wafer-level chip-scale packaging [43.17], 3-D packaging [43.18], etc. to satisfy the needs for high I/O density, large die area, and high clock frequency. The functions of conventional IC packaging are to protect, power, and cool the microelectronic chips or components and provide electrical and mechanical connection between the microelectronic part and the outside world. Unlike regular ICs, the diversity of MEMS products complicates the sealing issue. MEMS packaging processes cannot directly follow the procedures applied in the IC packaging industry due to inclusion of free-standing physical microstructures or chemical substances which cannot survive dicing or pick-and-place steps before the sealing process. Moreover, MEMS components may need to interface with the outside environment (for example, fluidic interconnectors [43.19]), while other components may need to be hermetically sealed (for example, accelerometers [43.5]) in addition to the need for electrical interconnects. Therefore, MEMS packaging processes may have to provide more functionality, including better mechanical protection, thermal management, hermetic sealing, and complex electrical and signal distributions.

It has been suggested that MEMS packaging should be incorporated in the device fabrication stage as part of the micromachining process. Although this approach solves the packaging requirement for some specific devices, it does not address it for general microsystems. In particular, many MEMS devices are now fabricated using various foundry services [43.20, 21], and there is a tremendous need for a uniform packaging process. Figure 43.1 shows a typical MEMS device being encapsulated by a packaging cap. The most fragile part of this device is the suspended mechanical sensor, which is a freestanding mass–spring microstructure. This mechanical part should be protected during packaging and handling processes. Moreover, vacuum encapsulation may be required for such microstructures in applications including resonant accelerometers [43.5] or gyroscopes [43.6, 22]. A *packaging cap* with a properly designed microcavity should be fabricated to encapsulate and protect such fragile MEMS structures as a first-level postpackaging process. The wafer can then be diced and well-established packaging technology from the IC industry applied to complete the packaging. Additionally, a common packaging requirement for MEMS devices is hermetic sealing and sometimes vacuum encapsulation. Hermetic sealing is important to ensure that no moisture or contamination can enter the package and affect the functionality of microstructures. This increases the difficulty of common IC packaging processes tremendously. Although most single-function MEMS chips can employ typical IC packaging techniques, such as die-attachment pro-

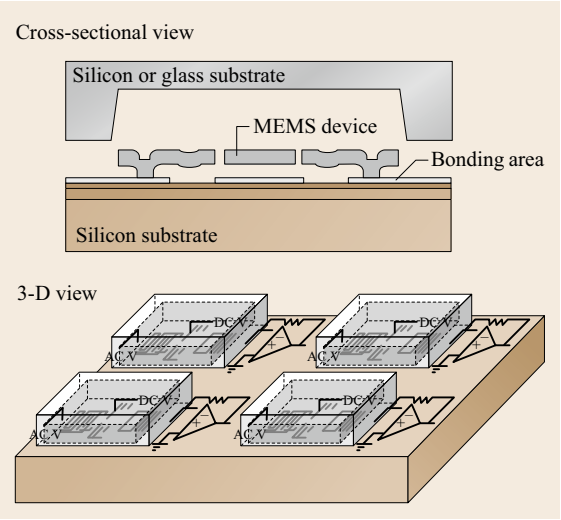


Fig. 43.1 Typical MEMS packaging with a MEMS structure encapsulated and protected by a packaging cap

cesses and wiring interconnects, using molded plastic, ceramic, or metal for packaging [43.12], the increasing complexity of MEMS devices requires more advanced packaging techniques, especially wafer-level packaging, for integration of multiple chips for multifunctional applications; for example, if chemical or biomedical substances are present [43.23], any sealing process must not exceed low processing temperatures. For optical devices [43.24], the sealing process should provide good optical paths. If mechanical resonators are included [43.25], vacuum sealing might be required to improve device performance, with the desired vacuum level depending on the device specification.

Before state-of-the-art MEMS packaging processes are discussed, several primary microfabrication processes for packaging applications are summarized. These processes include the flip-chip (FC) technique, ball grid array (BGA), through-wafer etching, and electroplating. Other silicon-based processes, such as thin-film deposition, wet and dry chemical etching, lithography, lift-off, and wire-bonding processes, are described in many textbooks [43.26].

Flip-Chip (FC) Technique

This technique is commonly used in the assembly process of a chip with microelectronics and a packaging substrate [43.15]. The microelectronic chip is *flip joined* with the packaging substrate, and metal solder bumps are used for both bonding and to form electrical paths between bond pads on the microelectronic chip and metal pads on the packaging substrate. Because the vertical bonding space may be very narrow, being controlled by the height of the solder bumps, and the bond pads can be distributed across the whole chip rather than only on the edge, this technique can provide high-density input/output (I/O) connections. In the FC technique, solder bumps are generally fabricated by means of electroplating. Before the bumping process, multiple metal layers, such as TiW-Cu, Cr-Cu, Cr-Ni, or TaN/Ta/Ni, are deposited to form a seed layer for electroplating and as a diffusion barrier to prevent diffusion of solder into underlying electrical interconnects.

Ball Grid Array (BGA)

This technology is very similar to the FC technique. An area array of solder balls on a single- or multichip module are used in the packaging process as electrical, thermal, and mechanical connects to join the module with the next-level package, usually a printed circuit board [43.16]. The major difference between typical BGA and FC chips is the size of the solder bumps; in a BGA chip, the bumps are in the order of 750 μm in diameter, which is 10 times larger than those commonly used in an FC chip.

Through-Wafer Etching

This is a chemical etching process to make through-wafer channels on a silicon substrate for fabrication of vertical through-wafer interconnects. The chemical etching can be either a wet or dry process. Anisotropic or isotropic etching solutions can be used in a wet etching process, while the dry etch process is based on plasma and ion-assisted chemical reactions, which can be either isotropic or anisotropic. To create high-density and high-aspect-ratio through-wafer vias, deep reactive-ion etching (DRIE) is typically used. Two popular DRIE approaches, Bosch and Cyro, are well described in literature [43.27].

Electroplating

Electroplating is another common microfabrication process. It can be used for deposition of an adherent metallic layer onto a conductive or nonconductive substrate. The process on a conductive substrate is called electrolytic plating, utilizing a seed layer as the anode to transfer metal ions onto the cathode surface when a DC current is passed through the plating solution. The plating process without use of an electrical current is called electroless plating and can be applied on both conducive and nonconductive surfaces. The electroless plating process requires a layer of noble metal such as Pd, Pt, or Ru on the substrate to catalyze the self-decomposition reaction in the plating solution. Electroplating processes are very important for fabrication of electrical interconnects and solder bumps for packaging applications, because of their low process temperature and cost. They are generally applied to pro-

vide electrical and thermal paths for various IC/MEMS packaging approaches.

43.1.2 Contemporary MEMS Packaging Approaches

Several MEMS packaging issues and approaches before 1985 were discussed in the book *Micromachining and Micropackaging of Transducers* [43.28], and researchers have been working on MEMS packaging approaches continuously since then; for example, *Senturia* and *Smith* [43.29] discussed packaging and partitioning issues for microsystems, and *Smith* and *Collins* [43.23] used epoxy to bond glass and silicon for chemical sensors. Several multichip module (MCM) methods have been proposed. *Butler* et al. [43.30] proposed adapting multichip module foundries using the chip-on-flex (COF) process. *Schuenemann* et al. [43.31] introduced a 3-D stackable packaging concept based on the top–bottom ball grid array (TBBGA) that includes electric, fluidic, optical, and communication interfaces. *Lee* et al. [43.32] and *Ok* and *Baldwin* [43.10] presented a direct-chip-attach MEMS packaging approach using through-wafer electrical interconnects. *Laskar* and *Blythe* [43.33] developed an MCM packaging process using epoxy. *Reichl* [43.34] discussed different materials for bonding and interconnection. *Grisel* et al. [43.35] designed a special process to package microchemical sensors. Special processes have also been developed for MEMS packaging, such as packaging for microelectrodes [43.36], packaging for biomedical systems [43.37], and packaging for space systems [43.38]. These specially designed, device-oriented packaging methods are aimed at individual systems. However, there are few reliable methods that would qualify as versatile MEMS postpackaging processes that meet the rigorous process requirements of low temperature, hermetic sealing, and long-term stability.

Several MEMS postpackaging processes have been proposed. *Butler* et al. [43.30] demonstrated an advanced MCM packaging scheme. It adopts the high-density interconnect (HDI) process consisting of embedding bare die into premilled substrates. Because MEMS structures have to be released after the packaging process, this scheme is undesirable for general microsystems. *Van der Groen* et al. [43.39] reported a transfer technique for complementary metal–oxide–semiconductor (CMOS) circuits based on epoxy bonding. This process overcomes the surface roughness problem, but epoxy is not a good material for hermetic sealing. In 1996, Cohn et al. demonstrated a wafer-to-wafer vacuum packaging process using silicon-gold eutectic bonding with a 2 μm-thick polysilicon microcap. However, experimental results showed substantial leakage after a period of 50 days. *Cheng* et al. [43.40] developed a vacuum packaging technology using localized aluminum/silicon-to-glass bonding. In 2002, *Chiao* and *Lin* [43.41] demonstrated vacuum packaging of microresonators by rapid thermal processing. These research efforts indicate the strong need for a versatile MEMS postpackaging process.

43.1.3 Bonding Processes for MEMS Packaging Applications

Silicon bonding technologies have been used in many MEMS fabrication and packaging applications, two types of which are commonly applied:

- Direct bonding processes such as anodic bonding and fusion bonding
- Bonding processes with intermediate layers, such as epoxy bonding, eutectic bonding, and solder bonding.

Direct wafer bonding processes are procedures that facilitate permanent attachment between two wafers without any intermediate layer. A permanent bond between two wafers can also be accomplished by using intermediate layers. Joining processes using intermediate layers have been extensively used in the ceramic industry to form metal-to-metal and metal-to-ceramic joints [43.42, 43] and can be characterized as [43.44]:

1. Fusion or melting of two materials to form a stable intermediate compound which facilitates the bond
2. Diffusion, in which pressurized joint parts are heated to 70% of the material's melting temperature to form a stable intermediate compound at the interface
3. Brazing, where a filler material is placed between the two parts to be joint, forming a stable intermediate compound upon heating.

These processes are commonly used when a lower bonding temperature or stronger bonding interface is required but cannot be achieved by a direct bonding process. Furthermore, the intermediate layers may reflow during the bonding process and fill the gaps between two bonding surfaces to overcome the surface roughness problem commonly encountered during direct wafer bonding processes. As such, the requirement for fine surface roughness for direct wafer bonding processes can be greatly relieved by using such intermediate layers.

Many MEMS applications have used both direct bonding as well as other bonding processes based on intermediate layers; For example, devices such as pressure sensors, micropumps, and biomedi-

cal or chemical sensors require mechanical interconnection when bonded onto the substrate [43.7, 19, 45]. Glass has been commonly used as the bonding material for anodic bonding at temperature of about 300–450 °C [43.46, 47]. *Klaassen* et al. [43.48] and *Hsu* and *Schmidt* [43.49] demonstrated different types of silicon fusion and Si–SiO_2 bonding processes at very high temperatures above 1000 °C. *Ko* et al. [43.28], *Tiensuu* et al. [43.50], *Lee* et al. [43.51], and *Cohn* et al. [43.52] used eutectic bonding for different applications. All of these bonding techniques have different mechanisms that determine the individual bonding characteristics and process parameters. This section discusses the details of these processes.

Fusion Bonding

Silicon fusion bonding is an important fabrication technique for silicon-on-insulator (SOI) technology. This method is based on the strong Si–O, Si–N, or Si–Si covalent bonds. However, very high bonding temperature (above 1000 °C) and flat bonding surfaces (below 6 nm) to ensure intimate contact are the two basic requirements for strong, uniform, and hermetic bonding. The common silicon-to-silicon fusion bonding process starts with wafer hydration (soaking in H_2O_2–H_2SO_4 mixture, diluted H_2SO_4, or boiling nitric acid, or use of oxygen plasma) to create a hydrophilic top layer consisting of O–H bonds [43.53]. Prebonding is accomplished when the two wafers are brought into intimate contact and van der Waals forces create a bond between the two wafers. An annealing step at elevated temperature is required to strengthen the bond. Although hydrophilic surface treatment can lower the bonding temperature, annealing above 800 °C is still required to prevent bubble formation at the bonding interface. *Bower* et al. [43.54] proposed that Si_3N_4 fusion bonding could be achieved at temperature below 300 °C. *Takagi* et al. [43.55] proposed that silicon fusion bonding could be achieved at room temperature by using Ar^+-beam treatment of the wafer surface with bond strength comparable to conventional fusion bonding. In summary, fusion bonding is a popular fabrication technique for MEMS fabrication and packaging.

Anodic Bonding

The invention of anodic bonding dates back to 1969, when *Wallis* and *Pomerantz* [43.56] found that glass and metal could be bonded together at about 200–400 °C below the melting point of glass with the aid of a high electrical field. This technology has been widely used for protecting onboard electronics in biosensors [43.57–59] and sealing cavities in pressure sensors [43.60]. Many reports have also discussed the possibility of lowering the bonding temperature using different approaches [43.61, 62]. Anodic bonding forms Si–O or Si–Si covalent bonds, some of the strongest chemical bonds available for silicon-based systems. The bonding process can be accomplished on a hot-plate at temperature of 180–500 °C in atmosphere or vacuum environment. When a static electrical field is applied between Pyrex glass and silicon, sodium ions in the glass migrate away from the silicon–glass interface, creating a locally high electrical field, and a bond forms due to electrochemical effects [43.56]. To create a high electrical field, a flat bonding surface with roughness of less than 50 nm is required. In addition, the electrical field required for bonding is larger than 3×10^6 V/cm [43.28]. Such a high electrical field is generated by a power supply of 200–1000 V. Figure 43.2 shows the setup for anodic bonding, where two bonding wafers are brought together and heated to an elevated temperature to supply the bonding energy. Care must be taken if there are free-standing, conductive micromechanical structures on either wafer to be bonded, as the high voltage tends to pull and may damage such structures. A thin metal film can be formed on the glass cap to provide shielding to solve this problem, as shown in Fig. 43.2. Furthermore, Corning 7740 Pyrex is commonly used in silicon-to-glass bonding systems, because it contains sodium ions and has a thermal expansion coefficient close to that of single-crystalline silicon in the range of 200–300 °C. The problem of induced residual stress can thereby be minimized in this temperature range. *Hanneborg* et al. [43.63] successfully bonded silicon with other thin solid films, such as silicon dioxide, nitride, and polysilicon, by using an intermediate glass layer with the anodic bonding technique. *Chavan* and *Wise* [43.64] reported absolute pressure sensors fabricated using the anodic bonding technique. In this process, a silicon cap with a thin, heavily doped boron layer and a recess cavity was bonded in vacuum environment to a glass substrate with prefabricated interconnection lines. However, the problem of oxygen outgassing due to the high electrical field in the anodic bonding process presents a challenge for the vacuum sealing process [43.65]. A thin Ti/Pt

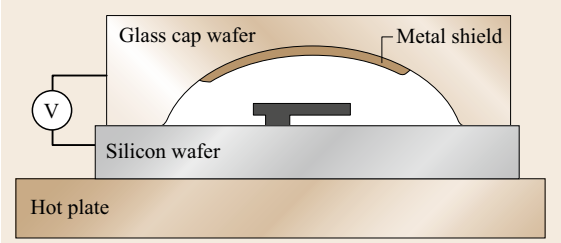


Fig. 43.2 Schematic diagram of setup for silicon-to-glass anodic bonding

layer predeposited on the glass surface has been shown to provide a good diffusion barrier, and the resulting pressure in the cavity can reach 200 mTorr [43.64]. In another example, microgyroscopes were fabricated using the anodic bonding technique by *Hara* et al. [43.66].

In practice, electrostatic bonding has become widely accepted for MEMS fabrication and packaging applications, as described above. Unfortunately, the possibility of contamination due to excessive alkali metal in the glass or damage to microelectronics due to the high electrical field, and the requirement for flat bonding surfaces limit application of anodic bonding for MEMS postpackaging [43.67].

Epoxy Bonding (Adhesive Bonding)

Epoxy comprises four major components: epoxy resin, filler-like silver slake, solvent or reactive epoxy diluent, and additives such as hardeners and catalysts [43.68, 69]. The bonding mechanism is very complicated, depending on the type of epoxy applied. In general, the main source of bonding strength is the van der Waals force. Because epoxy is a soft polymer material and its curing temperature for bonding is only around 150 °C, low residual stress and process temperature are the major advantages of epoxy bonding. However, the properties of epoxy can be easily changed by environmental humidity and temperature, so the bonding strength decays over time. In addition, epoxy bonding has low moisture resistance and is a dirty process due to its additives. These disadvantages make epoxy bonding unfavorable for special MEMS packaging requirements such as hermetic or vacuum sealing.

Eutectic Bonding

In many binary systems, there is a eutectic point corresponding to the alloy composition with the lowest melting temperature. If the environmental temperature is maintained above this eutectic point, two contacting surfaces containing the two elements with the eutectic composition can form liquid-phase alloy. Solidification of this eutectic alloy results in *eutectic bonding* at a temperature lower than the melting temperature of either element in the alloy. Eutectic bonding can form a strong metal bond; For example, in the case of the Au–Si alloy system, the eutectic temperature is only 363 °C for the composition with atomic ratio of 81.4% Au to 18.6% Si, and the bonding strength is above 5.5 GPa [43.70]. Because other alloy systems may have lower eutectic temperature than the Al–Si system, they have great potential for use in MEMS packaging applications. In addition to the Au–Si system, the Al–Ge, Au–SnSi, and Au–Ge–Si systems have been applied for MEMS packaging. Specifically, InvenSense has used the Al–Ge bonding process for their gyroscope products [43.71].

Solder Bonding

Solder bonding has been widely used in microelectronic packaging [43.72], offering the advantages of both low bonding temperature and high bonding strength for packaging applications. Furthermore, there are a variety of choices of solder material for specific applications. *Singh* et al. [43.73] successfully applied solder bump bonding for integration of electronic components and mechanical devices in MEMS fabrication [43.74]. In this case, indium metal was used to bond two separated silicon surfaces together by applying 350 MPa pressure, with bonding strength as high as 10 MPa. Glass frit can also be treated as a solder material and has been extensively used for vacuum encapsulation in the MEMS industry. Glass frits are ceramic materials that can provide strong bonding strength with silicon and good hermeticity. The bonding temperature is lower than 400 °C, suitable for electronic components. However, bonding width of greater than 200 μm is required to achieve good bonding results, which may become a drawback because area is a measure of manufacturing cost in the IC industry. Nevertheless, glass frit is the most popular bonding process used in current MEMS products.

Localized Heating and Bonding

Low bonding temperature and short process time are desirable process parameters for MEMS packaging fabrication, to decrease the thermal budget and increase throughput. However, most chemical bonding reactions require a minimum thermal energy to overcome the reaction energy barrier, or activation energy, to start the reaction and form a strong bond. As a result, high bonding temperature generally results in shorter processing time to reach the same bonding quality at a lower bonding temperature [43.75]. The common limitations of the bonding techniques described above are their individual bonding characteristics and temperature requirements. In general, MEMS packaging requires good bonding for hermetic sealing, while the processing temperature must be kept low at wafer level to have less thermal effects on devices that are already present; For example, a MEMS device may have prefabricated circuitry, biomaterial or other temperature-sensitive materials such as organic polymer, magnetic metal alloy, or piezoceramic. Since the packaging step comes after the MEMS device fabrication processes, the bonding temperature should be kept low to avoid effects of high temperature on the system. Possible temperature effects include residual stress due to thermal expansion coefficient mismatch between bonding materials and substrates, electrical contact failure due to atomic interdiffusion at the interface, and contamination due to outgassing or evaporation of materials. In addition to control of

Table 43.1 Summary of bonding mechanisms

Bonding method	Temperature	Roughness	Hermeticity	Postpackaging	Reliability
Fusion bonding	Very high	Highly sensitive	Yes	Yes by LH[a]	Good
Anodic bonding	Medium	Highly sensitive	Yes	Difficult	Good
Epoxy bonding	Low	Low	No	Yes	–[b]
Integrated process	High	Medium	Yes	No	Good
Low-temp. bonding	Low	Highly sensitive	–[b]	No	–[b]
Eutectic bonding	Medium	Low	Yes	Yes by LH[a]	–[b]
Brazing	Very high	Low	Yes	Yes by LH[a]	Good

[a] Localized heating
[b] No conclusive data

the bonding temperature, the magnitude of the force applied to create intimate contact for bonding and control of the atmospheric environment are other factors that should be considered. Based on heat transfer simulations [43.76], it is possible to confine the high-temperature area to a small region by using localized heating without heating the whole substrate. Therefore, assembly steps can always be applied after device fabrication without having detrimental effects. As such, localized heating and bonding techniques have been introduced and implemented as postprocessing approaches for fabrication of MEMS packaging [43.75, 77].

Table 43.1 summarizes these MEMS packaging technologies and their limitations, including the localized heating and bonding approach. The localized heating approach introduces several new opportunities. First, better and faster temperature control can be achieved. Second, higher temperature can be applied to improve the bonding quality. Third, new bonding mechanisms that require high temperature such as brazing [43.78] may now be explored for use in MEMS applications. As such, this approach has potential applications for a wide range of MEMS devices and is expected to advance the field of MEMS packaging.

43.2 Hermetic and Vacuum Packaging

Hermetic packaging is beneficial because it provides a moisture-free environment to avoid charge separation in capacitive devices, corrosion of metallization, and electrolytic conduction, thereby prolonging the lifetime of electronic circuitry. Especially for MEMS packaging, hermeticity is desirable in most cases since one of the main failure mechanisms for MEMS devices is humidity, and the surface tension of water can cause stiction of micromechanical structures, leading to malfunction. In several device applications, vacuum encapsulation is necessary but can be costly. Many surface-micromachined resonant devices need vacuum to improve their performance, such as comb-shaped μ-resonators and ring-type μ-gyroscopes, which have very high surface-to-volume ratio and vibrate in a very tight space [43.22, 79]. Two major approaches for hermetic and vacuum packaging of MEMS have been demonstrated and are discussed in this section:

- The integrated encapsulation approach
- The postprocess packaging approach.

Moreover, vacuum encapsulation by means of localized heating and bonding is discussed separately as another example of issues related to hermetic and vacuum packaging.

43.2.1 Integrated Micromachining Processes

Several hermetic and vacuum packaging processes for MEMS have been demonstrated based on integrated micromachining processes, where construction of sealing or protection caps is integrated with the MEMS device manufacturing process. This integrated approach has the advantage of sealing mechanical components in situ prior to the chip dicing and handling steps, to avoid contamination. An integrated vacuum sealing process using low-pressure chemical vapor deposition (LPCVD) is presented here as an illustrative example. This integrated process can encapsulate comb-shaped microresonators [43.80] in vacuum at wafer level. Figure 43.3 presents a cross-sectional view of the manufacturing process. First, a standard surface micromachining process [43.81] is conducted by using four masks to define a first polysilicon layer, anchors to the substrate, dimples, and a second polysilicon layer, as shown in Fig. 43.3a. The process so far is similar to the polyMUMPs process (MUMP: multi-user MEMS process) [43.21], and comb-shaped microstructures are fabricated at the end of these steps. In the standard surface micromachining process, the sacrificial layer (oxide) would be etched away to release the microstruc-

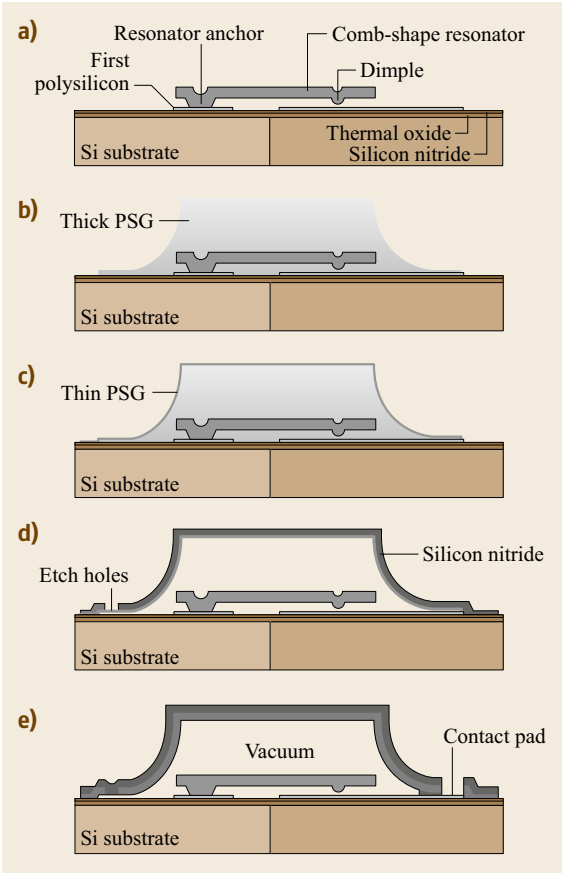


Fig. 43.3a–e Integrated vacuum encapsulation process using LPCVD nitride sealing to package micromechanical resonators. **(a)** Surface micromachining; **(b)** microshell; **(c)** etch channels; **(d)** etching and drying; **(e)** sealing (after [43.76])

tures. In the MEMS postpackaging process, a thick (7 μm) layer of phosphorus-doped glass (PSG) is deposited to cover the microstructure and patterned by using 5 : 1 buffered HF (BHF) to define a microshell area, as shown in Fig. 43.3b. A thin PSG layer of 1 μm is then deposited and defined to form etch channels, as illustrated in Fig. 43.3c. The microshell material, low-stress silicon nitride, is now deposited with thickness of 1 μm. Etch holes are defined and opened on the silicon nitride layer by using a plasma etcher. Silicon dioxide inside the packaging shell is now etched away using concentrated HF, and the wafer is dried by using the supercritical CO_2 drying process [43.82]. The result after these steps is shown in Fig. 43.3d. A 2 μm-thick LPCVD low-stress nitride is then deposited at pressure of 300 mTorr to seal the shell in the vacuum condition. Finally, the contact pads are opened, as shown in Fig. 43.3e. Figure 43.4 shows a scanning electron

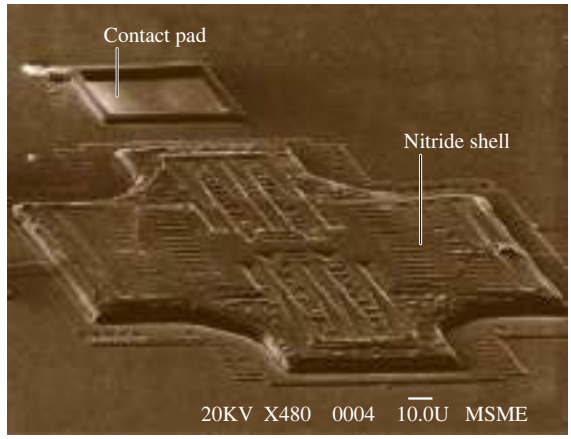


Fig. 43.4 SEM image showing a MEMS mechanical comb-shaped resonator vacuum-packaged using integrated LPCVD sealing as depicted in Fig. 43.3 (after [43.76])

microscopy (SEM) image of a finished device with a protective microshell on top. The total packaging area (microshell) has dimensions of about $400 \times 400\,\mu m^2$. A contact pad is shown with the covering nitride layer removed. The shape of the microresonator, with beams 150 μm long and 2 μm wide, is reflected on the surface of the microshell due to the integrated packaging process. The total height of the nitride shell is 12 μm, as seen standing above the substrate. Spectral measurements of the comb resonator inside the packaging revealed that a vacuum level of about 200 mTorr is accomplished using this method [43.83].

Similarly, *R. Aigner* et al. [43.84] reported a BiCMOS-compatible, integrated vacuum sealing process to package a polysilicon microaccelerometer. The protective shell was a polysilicon layer with supporting pillars anchored on the structural polysilicon. Release was achieved using a HF gas-phase etching process to remove the sacrificial oxide layer, and the release holes were sealed in a vacuum environment. The device was then injection molded into a plastic package at pressure of 100 bar. The supporting pillars were strong enough to hold the polysilicon shell under this high-pressure molding process.

An integrated sealing process based on evaporation of aluminum has also been reported [43.85]. A silicon substrate was deposited with a 4 μm-thick epitaxial n-type silicon layer. A controlled plasma etching and oxidation process formed a sharp tip, and a layer of borophosphosilicate glass (BPSG) was used to fill the trench as a sacrificial layer. The nitride sealing cap was deposited and patterned, and a 290 nm-thick PSG sacrificial release via was deposited and patterned, followed by deposition and patterning of the polysilicon anode. After the release etching process, aluminum evapora-

tion was applied in a 2×10^{-6} Torr vacuum chamber to deposit an 800 nm-thick aluminum layer to seal the release via. The resulting pressure was estimated as 1 mPa by measuring vacuum diode characteristics.

Other similar processes based on the integrated encapsulation concept have been demonstrated; For example, *Sniegowski* et al. [43.86] developed a reactive sealing method to seal vibratory micromachined beams; *Ikeda* et al. [43.79] used epitaxial silicon to seal microstructures; *Mastrangelo* et al. [43.87] used silicon nitride to seal mechanical beams as light sources; *Smith* et al. [43.88] used the approach of embedding microstructures and CMOS circuitry. All of these approaches integrate the encapsulation process within the MEMS fabrication process. The typical advantage of this approach is that such devices are ready for standard IC packaging processes such as dicing, pick-and-place, etc. once the wafer-level integrated sealing processes are completed. Specifically, SiTime has used a similar vacuum integration process to make their time-reference resonators [43.89].

Although the above vacuum sealing processes can successfully achieve MEMS hermetic and vacuum packaging, they suffer from some drawbacks; For example, these postpackaging processes are highly process dependent and are not suitable for generic MEMS postpackaging. MEMS companies or researchers have to adapt these postpackaging processes for their own device manufacturing process. Currently, standard MEMS foundry services do not support any of these integrated processes. Also, integrated encapsulation does not enable control of the cavity pressure, although it can achieve low pressure by wafer-level fabrication and provide lower manufacturing cost.

43.2.2 Postpackaging Processes

The second approach is defined as postpackaging processing. Such a packaging process starts when the device fabrication processes are completed, so this approach has high flexibility for various microsystems; For example, Fig. 43.5 shows a common industrial hermetic postpackaging technique called dual-in-line packaging (DIP) [43.90, 91]. A die is placed inside a ceramic holder covered by a sealing lid. Solder or ceramic joining is generally used for assembling the lid and holder under a pressure-controlled environment. High cost is the major drawback of this method, because of the expensive ceramic holder and low fabrication throughput. Another example postpackaging method is based on wafer bonding techniques combined with microshell encapsulation. Devices are sealed by stacking another micromachined silicon or glass substrate, as illustrated in Fig. 43.1. Integrated microsystems and

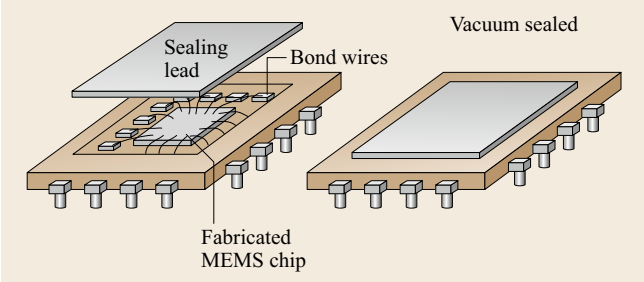


Fig. 43.5 Schematic of industrial post-packaging (DIPs) using a ceramic holder to be covered by a sealing lid

protection shells are fabricated on different wafers, either silicon or glass, at the same time. After the two substrates are assembled together using silicon fusion, anodic, or low-temperature solder bonding to achieve the final encapsulation, these microshells provide mechanical support, thermal paths, or electrical contact for the MEMS devices. Low packaging cost can be expected due to wafer-level processing.

A special heating method using rapid thermal processing (RTP) for wafer bonding applications is explained here to illustrate the roles of various control parameters such as temperature, time, and intermediate bonding materials. *Chiao* and *Lin* [43.92] reported a wafer bonding process by melting an intermediate filler material to facilitate sealing of micromechanical structures. Figure 43.6a shows the concept of the bonding and sealing scheme using an aluminum-to-glass bonding system, while Fig. 43.6b shows the experimental setup for aluminum-to-nitride bonding with integrated comb resonators inside. Aluminum with thickness of 3–4 μm was patterned to form sealing rings that surround the micromechanical structures on the device wafer. The width of a typical aluminum sealing ring was 100–200 μm, and the sealing area was $600 \times 600\,\mu m^2$. A glass (Pyrex, Corning 7740) wafer was used as the cap to cover the MEMS devices. The heating and bonding energy was provided by RTP; a typical heating history is shown in Fig. 43.7, where the overall heating process can be completed in 1 min, during which the temperature rises from room temperature to 990 °C then decreases to 350 °C. The aluminum-to-glass bonding and joining process was accomplished by heating at 990 °C for 2 s in the RTP chamber. It was shown that aluminum could extract oxygen to form aluminum oxide to create the bond [43.93]. Figure 43.8 shows a microheatuator successfully packaged using this bonding process; the surrounded liquid, isopropanol alcohol (IPA) in this case, was sealed without penetrating inside the package.

Other material systems have also been bonded using this RTP process, e.g., aluminum-to-silicon nitride

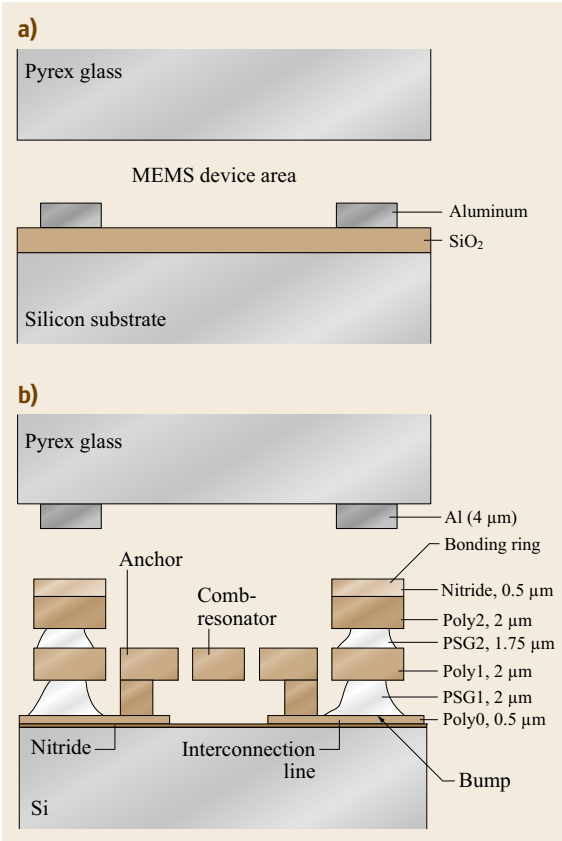


Fig. 43.6a,b Schematics of RTP bonding experiments: **(a)** the concept of aluminum-to-glass bonding, and **(b)** aluminum-to-nitride bonding with comb resonators

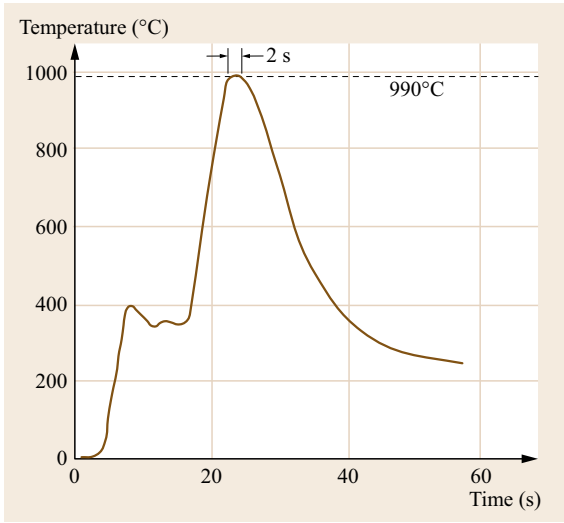


Fig. 43.7 Temperature history in an RTP bonding experiment

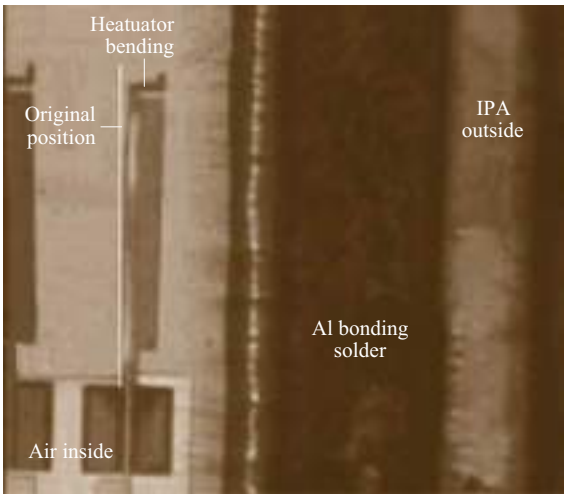


Fig. 43.8 A hermetically packaged microheatuator operating when immersed in liquid

joining (Fig. 43.6b) [43.94]. In this case, a 5000 Å-thick LPCVD silicon nitride layer was deposited and patterned on top of sealing ring structures that encompassed surface-micromachined comb-shaped resonators [43.81]. Using RTP with process time of 10 s at the peak temperature of 750 °C, a stable bond was formed at the aluminum–nitride interface. Figure 43.9 shows a packaged comb resonator that resonated at 19.6 kHz when immersed in deionized (DI) water, as seen under an optical microscope. The aluminum-to-nitride seal successfully blocked water from entering the package. To examine the bonding strength, the package was forcefully broken, as shown in Fig. 43.10. The glass debris was attached to the sealing ring surrounding the comb-drive resonator on the silicon substrate. This shows that the bonding strength of the aluminum-to-nitride system was greater than the glass fracture strength, which is estimated at around 270 MPa [43.95].

A vacuum sealing process using RTP bonding is discussed here in detail to address the technical issues in vacuum sealing processes. *Chiao* and *Lin* [43.96] reported a vacuum sealing process based on aluminum-to-nitride bonding using RTP. The RTP bonding process was conducted in a vacuum quartz tube, as shown in Fig. 43.11. Both device and cap wafers must be baked in vacuum at 300 °C for at least 4 h to drive out water and gas species that may adhere at the wafer surface [43.97]. This prebaking process in vacuum was necessary to minimize the outgassing effect during the bonding process in order to achieve high-quality vacuum. Afterwards, the device and cap wafers were immediately flip-chip-assembled, loaded into a sample holder, and placed inside a quartz chamber, as shown in Fig. 43.11. The system was then placed inside the RTP

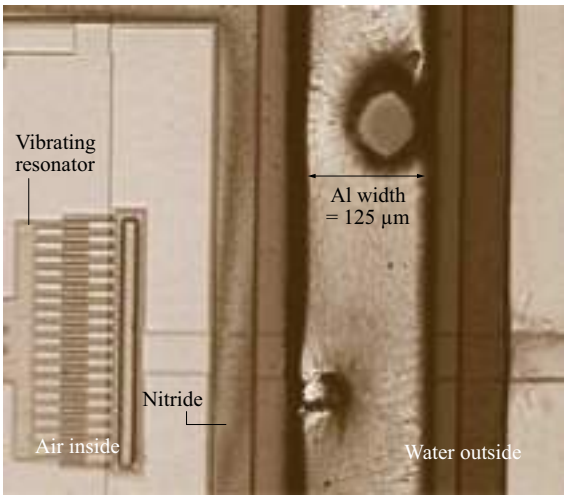


Fig. 43.9 A comb-drive resonator sealed in a a package chip and operating when the package was immersed in water

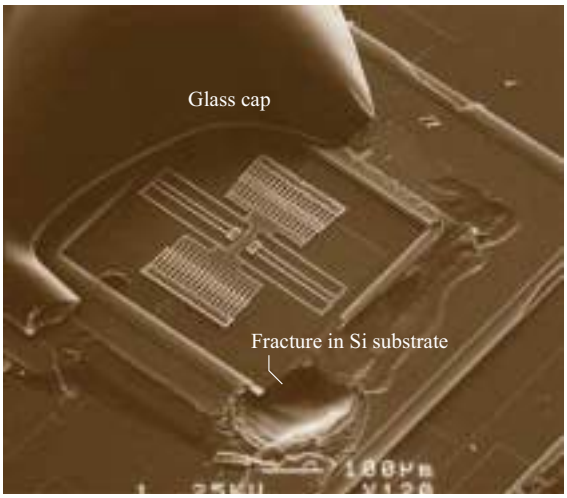


Fig. 43.10 SEM micrograph of a silicon substrate after forcefully breaking the aluminum-to-nitride bond. Glass debris was found attached to the silicon substrate

equipment, and the base pressure was pumped down to about 1 mTorr using a turbopump. The vacuum was held steady for 4 h to drive out gas trapped inside the package cavity [43.98]. The bonding and vacuum sealing process was achieved by RTP heating for 10 s at 750 °C to complete the bonding process.

Figure 43.12 shows the spectrum of a vacuum-packaged, double-folded beam comb-drive resonator measured using a microstroboscope. The central resonant frequency is at about 18 625 Hz, and the quality factor is extracted as 1800 ± 200, corresponding to a pressure level of about 200 mTorr inside the package [43.98]. This type of postpackaging process at wafer level has become the preferred approach for hermetic encapsulation, because it can provide lower cost and greater process flexibility. However, the packaging process relies on *good* bonding techniques. A strong and reliable bond between the two substrates should be provided, and the bonding procedure should be compatible with the other microsystem fabrication processes.

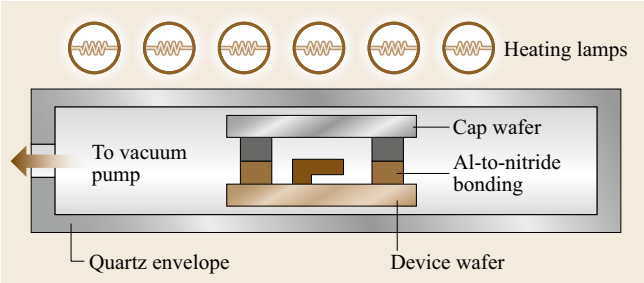


Fig. 43.11 Vacuum packaging apparatus for aluminum-to-nitride bonding using RTP

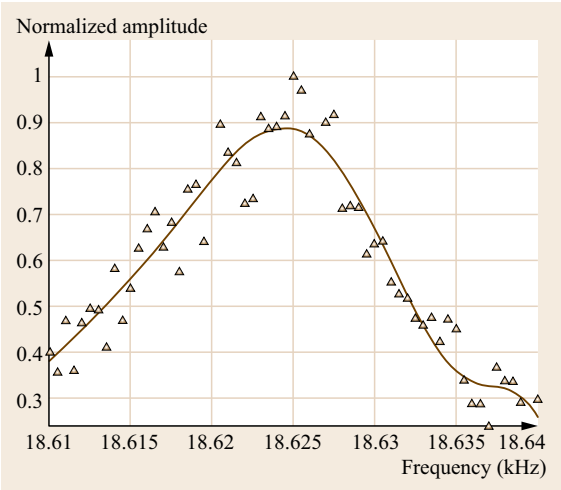


Fig. 43.12 Spectrum of a comb-shaped resonator vacuum-encapsulated by aluminum-to-nitride bonding using RTP (after [43.41])

43.2.3 Localized Heating and Bonding Processes

The approach of MEMS postpackaging by localized heating and bonding is proposed to address the problems of global heating effects. In this section, resistive microheaters are used as an example to provide localized heating, although several other means of localized heating have also been demonstrated, including laser welding [43.99], inductive heating [43.100], and ultrasonic bonding [43.101]. The principle of localized

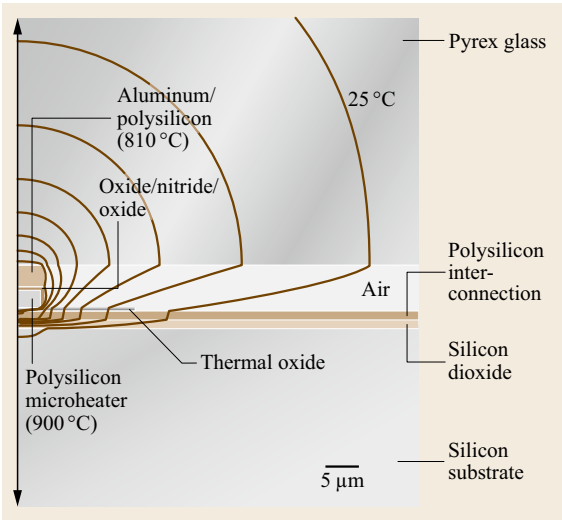


Fig. 43.13 Schematic of the 2-D heat transfer model with geometry and boundary conditions

heating is to achieve high temperature for bonding locally while maintaining low temperature globally at wafer level. Resistive heating by using microheaters on top of the device substrate is applied to form a strong bond with the silicon or glass cap. According to the results of two-dimensional (2-D) heat conduction finite element analysis (Fig. 43.13), the steady-state heating region of a 5 µm-wide polysilicon microheater capped with a Pyrex glass substrate can be confined locally as long as the bottom of the silicon substrate is constrained to ambient temperature. The physics of the localized heating behind this design can be understood by solving the governing heat conduction equations for the device structure without a cap [43.98]. As long as the width of the microheater and the thickness of the silicon substrate are much smaller than the die size and a good heat sink is placed underneath the silicon substrate, heating can be confined locally. The temperature of the silicon substrate can be kept low or close to room temperature. Several localized resistive heating and bonding techniques have been successfully developed for packaging applications, including localized silicon-to-glass fusion bonding, gold-to-silicon eutectic bonding, and localized solder bonding. Several solder materials have been successfully tested, including PSG, indium, and aluminum alloy [43.102].

The vacuum packaging example presented in this section is based on the localized aluminum/silicon-to-glass solder bonding technique. Built-in folded-beam comb drive μ-resonators are used to monitor the pressure of the package. Figure 43.14 shows the fabrication process for the package and resonators. Thermal oxide (2 µm) and LPCVD Si_3N_4 (3000 Å) are first deposited on a silicon substrate for electrical insulation, followed by deposition of 3000 Å LPCVD polysilicon. This polysilicon is used as both the ground plane and the electrical interconnect to the μ-resonators, as shown in

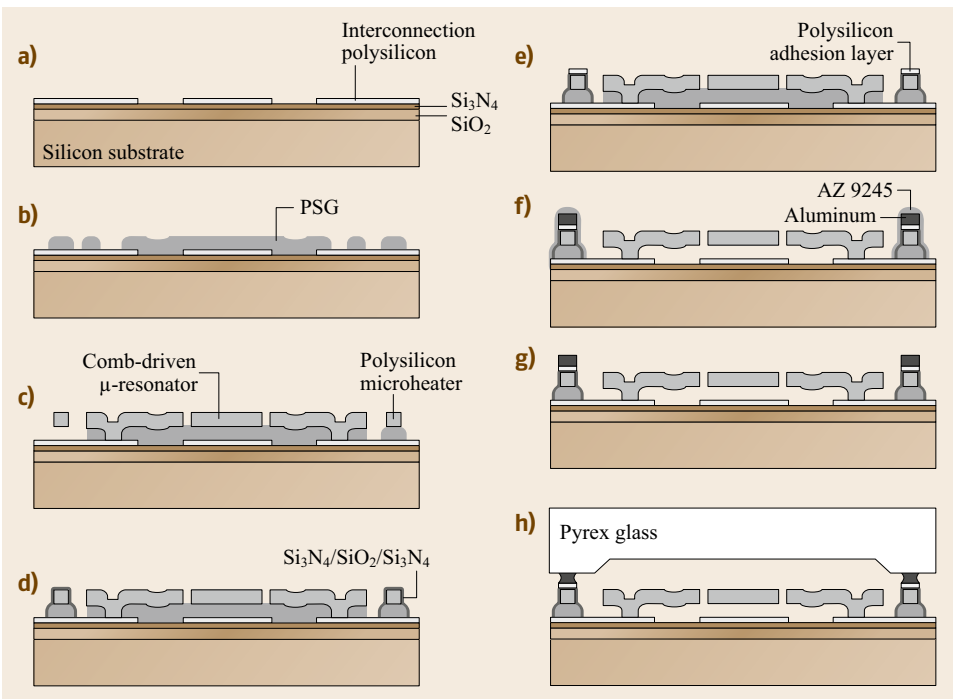


Fig. 43.14a–h Schematic process flow of vacuum encapsulation using localized aluminum/silicon-to-glass bonding. **(a)** Deposition and patterning; **(b)** sacrificial layer; **(c–g)** surface micromachining; **(h)** vacuum packaging

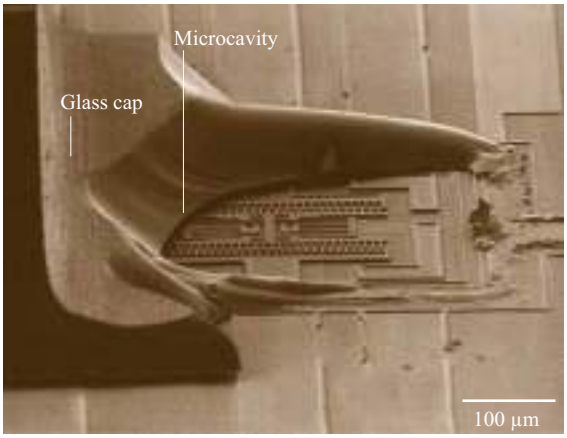


Fig. 43.15 SEM micrograph of encapsulated microresonators after the glass cap was forcefully broken

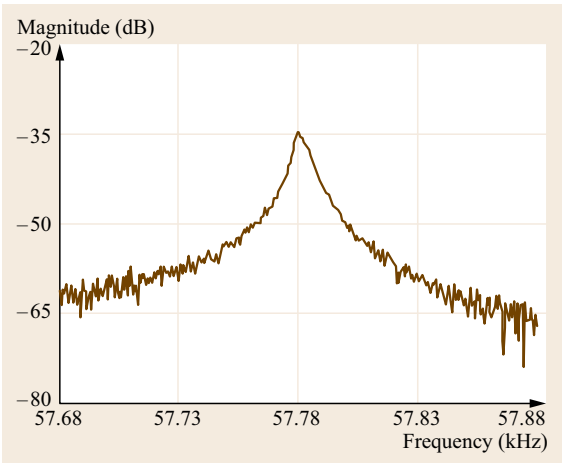


Fig. 43.16 Transmission spectrum of a glass-encapsulated μ-resonator after 120 min pump-down time in vacuum environment ($Q = 9600$)

Fig. 43.14a. Figure 43.14b shows a 2 µm LPCVD SiO_2 layer that is deposited and patterned as a sacrificial layer for fabrication of the polysilicon μ-resonators using a standard surface micromachining process. 2 µm-thick phosphorus-doped polysilicon is used as both the structural layer for the microresonators and the on-chip microheaters. This layer is formed over the sacrificial oxide in two steps to achieve a uniform doping profile. Lower input power and better process compatibility are two major advantages of using such on-chip microheaters for glass packages. The resonators are separated from the heater by a short distance of 30 µm, to effectively prevent their exposure to the high heater temperature, as shown in Fig. 43.14c. This concludes the fabrication of the μ-resonators.

To prevent the current supplied to the microheater from leaking into the aluminum solder during bonding, a LPCVD Si_3N_4 (750 Å)/SiO_2 (1000 Å)/Si_3N_4 (750 Å) sandwich layer is grown and patterned on top of the microheater, as shown in Fig. 43.14d. Figure 43.14e,f show how the aluminum (2.5 µm) and polysilicon (5000 Å) bonding materials are deposited and patterned. Sacrificial release is the final step to form the free-standing μ-resonators. Figure 43.14f shows how thick AZ 9245 photoresist is applied to cover the aluminum/silicon-to-glass bonding system to ensure that the system can withstand attack from concentrated hydrofluoric acid. After 20 min of sacrificial release in concentrated HF, the system as shown in Fig. 43.14g is ready for vacuum packaging. A Pyrex glass cap with a 10 µm-deep recess is then placed on top with applied pressure of ≈ 0.2 MPa under 25 mTorr vacuum, and the heater is heated using 3.4 W input power (the exact value depending on the design of the microheaters) for 10 min to complete the vacuum packaging process, as shown in Fig. 43.14h.

To evaluate the integrity of resonators packaged using such localized aluminum/silicon-to-glass solder bonding, the glass cap was forcefully broken and removed from the substrate. No damage was observed on the μ-resonator. A part of the microheater was stripped away, as shown in Fig. 43.15, demonstrating that a strong and uniform bond could be achieved

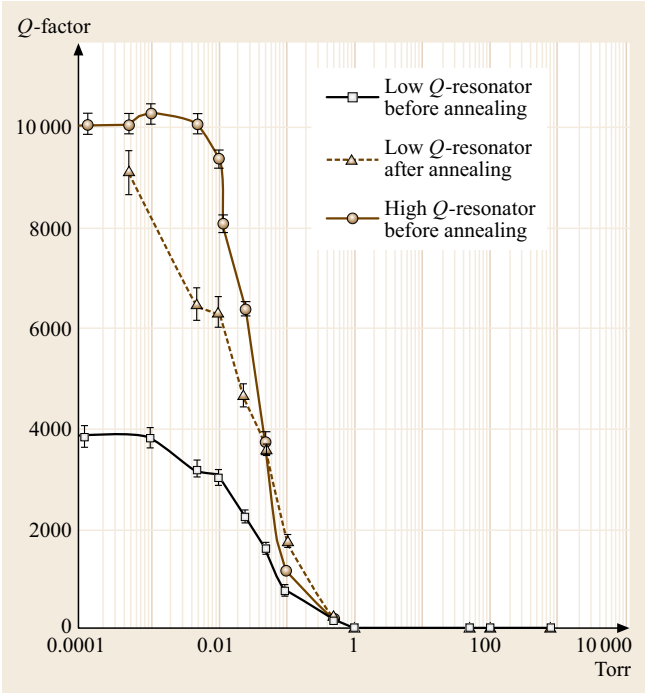


Fig. 43.17 Measured Q-factor versus pressure for unpackaged μ-resonators

without detrimental effects on the encapsulated device. Figure 43.16 shows a vacuum-encapsulated unannealed μ-resonator ($\approx$ 57 kHz) after 120 min of wait time. The Q-factor measured after packaging was 9600. Q versus pressure measurements for a high-Q unpackaged μ-resonator (Fig. 43.17) demonstrated that the pressure inside the package was comparable to the vacuum level in the packaging chamber.

Postprocess packaging using the localized heating and bonding technique includes four basic components:

1. An electrical and thermal insulation layer such as silicon dioxide or silicon nitride for localized heating
2. Resistive microheaters to provide the heating source for localized bonding
3. Bonding materials such as metal or polysilicon to provide good bonding and hermeticity with the silicon or glass substrate
4. A good heat sink under the device substrate to ensure localized heating during the bonding experiments.

MEMS devices are fabricated on the device chip then hermetically sealed in a cavity formed by the device chip, resistive microheaters, and protective cap. The process can be either die level or wafer level. A schematic design for the wafer-level packaging process is shown in Fig. 43.18. The resistive microheaters are parallel to each other and connected together to ensure that identical current density is applied for individual packages at the same time. These heaters can be fabricated on either the chip or protection cap and can be built on a larger wafer for current inputs. The interconnections for these packaging cavities can be built in the dicing area, such that no extra space is required for the packaging process.

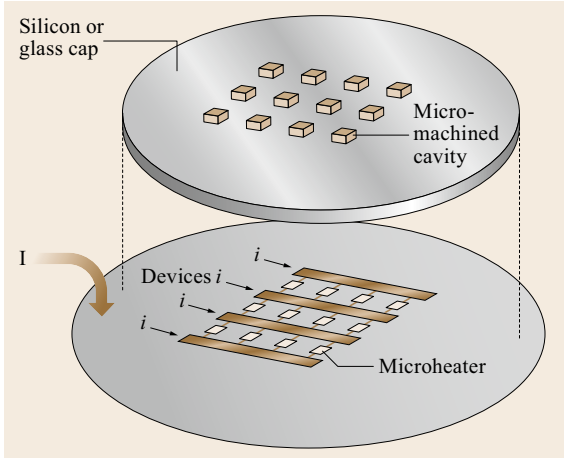


Fig. 43.18 Illustration of wafer-level vacuum packaging using the localized heating and bonding technique

43.3 Emerging Packaging Approaches

Wafer-level packaging and 3-D packaging have undoubtedly been the hottest emerging packaging technologies over the past few years and will be the trend in the near future [43.103, 104]. The process of packaging a wafer before dicing is called wafer-level packaging. Modern MEMS packaging usually involves at least three wafers, namely the application-specific integrated circuit (ASIC) wafer, the MEMS device wafer, and the cap wafer. These wafers are fabricated separately and bonded layer by layer to form the complete package. It is advantageous to encapsulate the MEMS devices early in the wafer processing to protect them mechanically and prevent contamination during subsequent fabrication and packaging processes. In this way, extra packaging equipment and repeated assembly processes can be largely eliminated. By dicing the bonded wafers at the end of the processes, one can obtain a large number of packaged chips with desired small footprint. For production of practical MEMS devices, this can reduce the cost of packaging and testing, and also improve yield and reliability. Meanwhile, 3-D packaging enables the combination of dissimilar classes of materials and components into a single system. Therefore, high-performance materials and subsystems can be combined in ways that would otherwise not be possible, thereby forming complex and highly integrated micro- or nanosystems; For example, polymers are being actively used for various MEMS sensors and actuators [43.105]. They provide many advantages in terms of cost, materials properties, and ease of processing. Integration of polymers for structural or functional purposes into microsystems is also a promising emerging technology with great potential for use in MEMS packaging.

43.3.1 Wafer-Level Packaging

Wafer-level packaging approaches for various applications can be classified into two categories:

- Interfacial bonding or bonding with intermediate melting materials
- Deposition sealing.

Mechanical protection of MEMS devices is realized by bonding a cap wafer on top of the device wafer, or by sealing of surface-micromachined cavities by deposition. Meanwhile, electrical feedthrough is needed for interconnection with the encapsulated MEMS. Normally, metal is used as the feedthrough material, while highly doped silicon can also be employed. In general, this is realized by either lateral feedthrough on the chip surface or vertical feedthrough using a through-hole via.

Wafer-Level Packaging by Interfacial Bonding

Transparent Pyrex glass wafers have thermal expansion coefficient similar to silicon wafers, and they are frequently used as cap wafers. In addition, interfacial bonding between metal and metal can also be employed. Electrical feedthrough is usually achieved using vias in the cap wafer. If silicon is used as the cap wafer, the thermal stress caused by thermal expansion mismatch between the device and cap wafers can be reduced, and holes in the cap wafer can be made using DRIE of silicon; For example, the fabrication process developed for film bulk acoustic resonator (FBAR) filters (RF MEMS devices) is shown in Fig. 43.19 [43.106]. Gold is electroplated using a photoresist mold, and the silicon substrate is etched by DRIE. The silicon cap wafer is bonded to the MEMS device wafer by Au–Au thermocompression bonding, and the etched holes are exposed by grinding the cap wafer. Finally, the packaged wafer is diced into individual chips, and wire bonding is carried out inside the exposed holes. In addition to Au–Au thermocompression bonding [43.107, 108], similar Al–Al [43.109] and Cu–Cu [43.110] thermocompression bonding can also be employed to achieve wafer-level packaging. Metal bonding surfaces are brought into contact with the application of force and heat simultaneously. Due to surface oxidation, high temperature is usually required to achieve high bonding quality, but MEMS packaging can be achieved at adequately low temperatures of about 300 °C or below. It has been demonstrated that Al-to-Al bonding can achieve wafer-level hermetic sealing and 3-D interconnects of MEMS devices [43.109]. Figure 43.20 shows wafer-level packaging processes employing Al-to-Al bonding and through-silicon vias (TSVs). After thermal oxidation for electrical isolation of the vias, highly doped poly-Si is deposited to fill the vias completely to form electrical signal paths. On the cap wafer, 2 μm-thick Al (with 2% Cu) is patterned at the perimeter of the individual dies as well as the input/output bond pads. On the MEMS device wafer, the seal rings and bond pads are also patterned using Al as described above. After cavity etching, the cap wafer is bonded to the MEMS wafer using Al bonding, creating both a hermetic seal around each die and conducting paths from the MEMS devices to the vias. The two wafers were bonded at ≈ 450 °C with various bond forces up to 80 kN. Leak detection on the capped device showed hermeticity of $\approx 10^{-12}$ cm^3 atm/s He leak rate with Al seal width as narrow as 3 μm. After back-grinding and polishing the bonded pair to expose the vias, new bond pads are formed on top of the cap wafer using standard passivation and redistribution layer (RDL) metallization. The electrical contact resistance of the Al-to-Al bonded interface was measured to be less than 1 Ω.

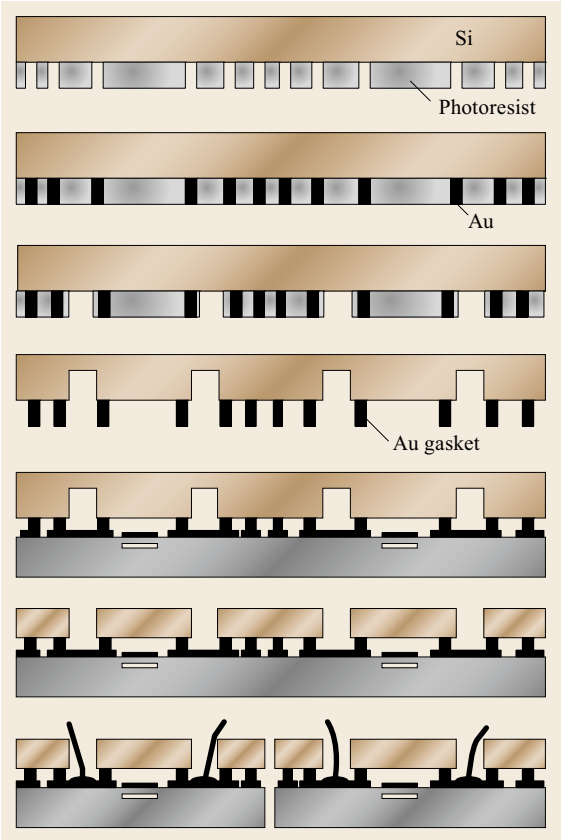


Fig. 43.19 Illustration of wafer-level packaging using metal-to-metal bonding (after [43.106])

Wafer-Level Packaging by Deposition Sealing

Narrow gaps for electrical feedthrough made between the MEMS wafer and the cap wafer can be sealed by depositing materials. Previously, an integrated process using surface-micromachined microshells was developed [43.111]. This process applies the concepts of sacrificial layer and low-pressure chemical vapor deposition (LPCVD) sealing to achieve wafer-level packaging. Similar processes have been

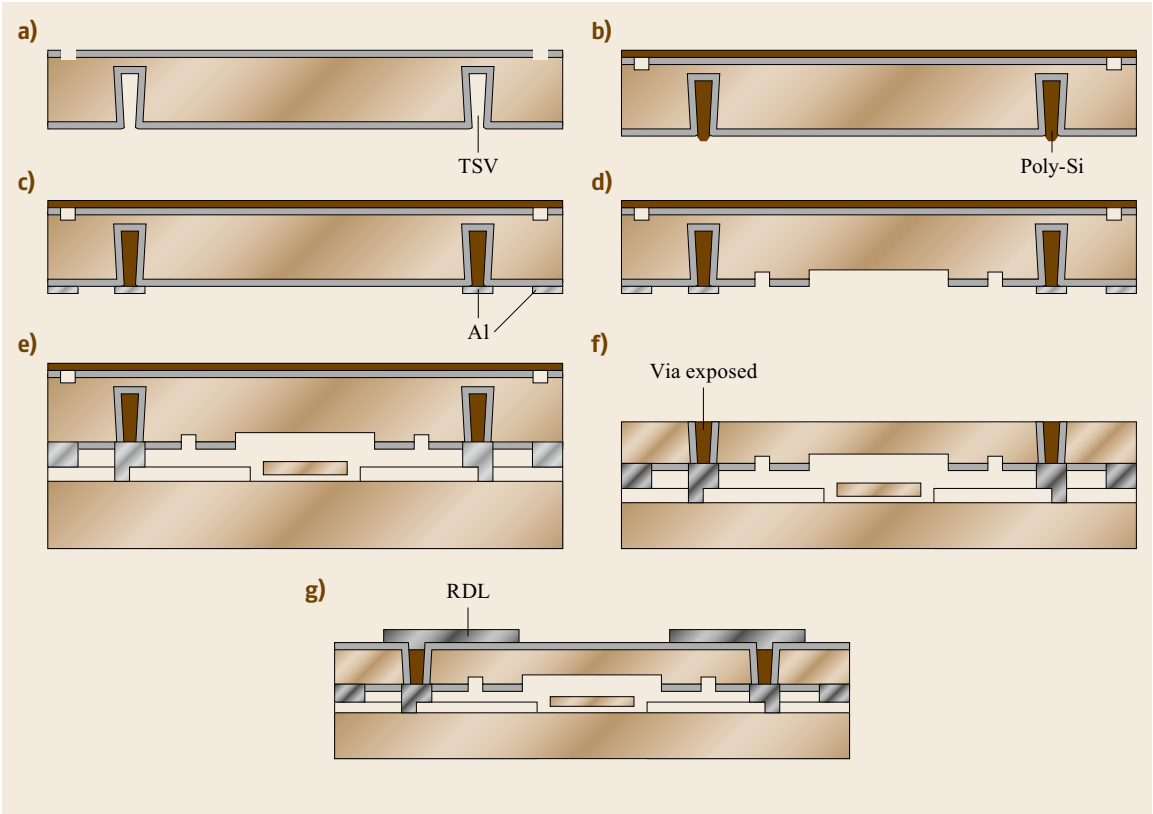


Fig. 43.20a–g Illustration of wafer-level hermetic packaging and 3-D interconnection using metal-to-metal bonding. **(a)** Etching and isolation; **(b)** filling and polishing; **(c)** deposition and patterning; **(d)** etching; **(e)** Al to Al thermocompression bonding; **(f)** back-grinding and polishing; **(g)** passivation and interconnection (after [43.109])

demonstrated; For example, *Guckel* et al. [43.112] and *Sniegowski* et al. [43.86] developed a reactive sealing method to seal vibratory micromachined beams. *Ikeda* et al. [43.79] adopted epitaxial silicon to seal microstructures. *Mastrangelo* and *Muller* [43.113] used silicon nitride to seal mechanical beams as light sources. *Smith* et al. [43.88] presented a new fabrication technology by embedding microstructures and complementary metal–oxide–semiconductor (CMOS) circuitry. All of these methods integrate the MEMS process with the packaging process so that no extra bonding process is required. However, these schemes are highly process dependent and not suitable for prefabricated circuitry.

Sacrificial etching can also be performed through porous materials such as porous poly-Si [43.115, 116] or through a gas-permeable membrane. Sealing can be achieved by depositing material on the porous material; For example, a permeable membrane is used in Fig. 43.21 for the following process sequence [43.114]: After forming the MEMS in the silicon wafer by etching, they are covered with patterned photoresist and a polymer overcoat. Since the photoresist thermally decomposes at 200 °C, the decomposed gas diffuses out through the polymer overcoat on heating. Finally, a thin metal layer is deposited on the surface of the polymer overcoat for hermetic sealing. In addition to deposition sealing, soldering can also be employed to seal MEMS; For example, a technique to premold and transfer lead-free solder balls for MEMS/electronics packaging applications has been demonstrated [43.117].

Electrical Feedthrough

Electrical feedthrough is needed for interconnection with the MEMS encapsulated in a cavity. Normally, metals are used as the feedthrough material, because of their low resistance, while highly doped silicon can also be used to reduce potential thermal stresses. Stray capacitance and stray inductance have to be taken into account for RF MEMS. Lateral feedthrough on a chip can be realized by embedding a thin metal layer in silicon dioxide and sealing the surface by anodic bonding with a glass cap wafer [43.118]. After etching the silicon dioxide on a silicon wafer, the etched groove is

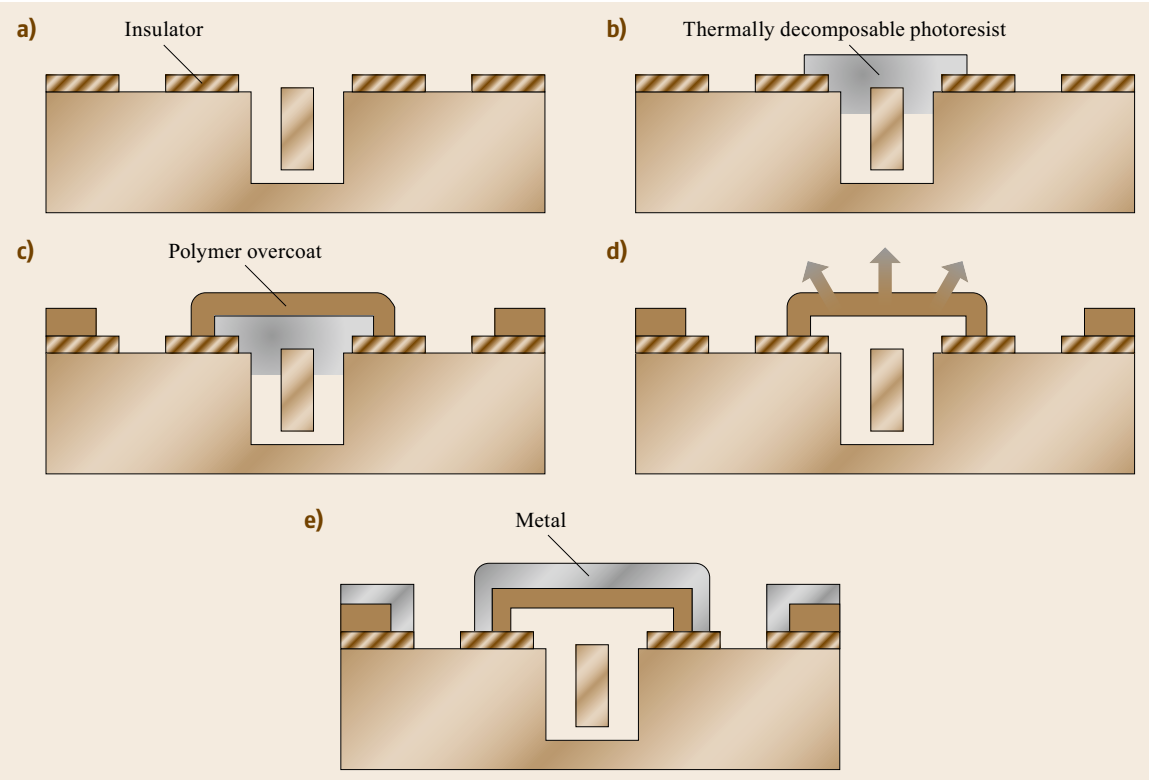


Fig. 43.21a–e Illustration of wafer-level packaging by deposition sealing. **(a)** Isolation; **(b)** cavity formation; **(c)** cap formation; **(d)** decomposition; **(e)** metallization (after [43.114])

filled with Cr–Al. The surface is coated with spin-on glass (SOG) to insulate and to make a planar surface, and silicon is sputter-deposited on it. This is anodically bonded to a Pyrex glass wafer. It was found that this lateral feedthrough is not as practically applicable compared with the vertical feedthrough described below, because the lateral feedthrough has to be fabricated on the same wafer as the MEMS.

Electrical interconnection can also be achieved using a through-hole via in the wafer, as shown in Fig. 43.22 [43.103]. The advantage of vertical feedthrough in the cap wafer is that the electrical interconnection can be fabricated on a different wafer from the MEMS wafer, so process compatibility is not required. For a glass cap wafer, it is not easy to make through holes. Meanwhile, vertical feedthrough can be achieved in a silicon wafer on which MEMS are fabricated [43.119].

43.3.2 3-D Packaging

Three-dimensional (3-D) packaging is an emerging technology that vertically stacks and interconnects multiple materials, technologies, and functional devices to form highly integrated micro- and nanosystems. Such work in the third dimension will allow exten-

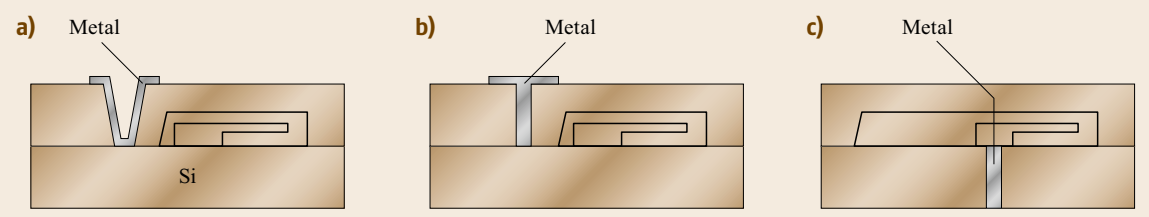


Fig. 43.22a–c Illustration of electrical feedthrough using a through-hole via in the wafer. **(a,b)** From top; **(c)** from bottom (after [43.103])

sion of Moore's law to higher density, functionality, and performance, as well as integration of more diverse materials and devices at lower cost. The potential benefits also include reduced power, small form factor, reduced packaging, and increased yield and reliability. It is recognized as an enabling technology for integration of MEMS, micro-opto-electro-mechanical systems (MOEMS), photonics, electronic ICs, and emerging NEMS.

Modern 3-D packaging technologies employ wafer-to-wafer joining, processing, and interconnection materials and components prepared using different processes. Typically they are classified into two categories: via-first approaches and via-last approaches. In via-first approaches, the vias establishing electrical (or alternatively optical) contacts between components on different substrates are defined during the bonding process. In contrast, in via-last approaches, the components are first bonded to each other, then vias to establish electrical (or alternatively optical) contacts between components on different substrates are defined. The advantages, limitations, and technological challenges of the via-first and via-last approaches are introduced in the following sections.

3-D Packaging by Via-First Approaches

Figure 43.23 shows the conceptual schemes of two major via-first 3-D packaging approaches. The bump bonding can be implemented by using, for example, solder bonding [43.120], eutectic bonding [43.121] or direct metal thermocompression bonding [43.122]. Typical dimensions of such bump-bonded metal vias are in the order of $100 \times 100\,\mu m^2$. At wafer level, it is challenging to obtain highly reliable processes for vias with dimensions below $20 \times 20\,\mu m^2$. In the technique shown in Fig. 43.23a, a completely processed and packaged MEMS device wafer containing metal bumps is bonded to a target wafer, typically a CMOS-based ASIC wafer. This technique is simply an extension of conventional chip-to-chip bump bonding to wafer-level bump bonding. Due to the minimum size of the metal bumps and the thickness of the wafer with the packaged MEMS devices, this technique allows only limited miniaturization and integration densities for the final system. In the technique shown in Fig. 43.23b, a partly processed MEMS wafer containing metal bumps is bonded to the target wafer. Thereafter, the components are further processed at wafer level to complete the MEMS devices using, e.g., etching and/or thinning processes. Although this technique allows only limited integration densities due to the minimum dimensions of the metal bumps, it has been proposed for a number of MEMS devices [43.123] and successfully implemented for commercial MEMS products [43.124].

The via-first 3-D packaging approach is being used for manufacture of gyroscopes [43.125], and combined three-axis accelerometers and three-axis gyroscopes integrated on a single chip. These devices are sold at very high volumes for consumer products such as motion controls in gaming. Taking the example of a gyroscope chip, the monocrystalline silicon capacitive gyroscope sensor and parts of the gyroscope package are prefabricated together, then bonded to the ASIC wafer that contains an etched cavity. Bonding and sealing is carried out directly to the top CMOS Al metal layer using an Al/Ge eutectic bonding process. Figure 43.24 shows a cross-sectional and top view of a gyroscope packaged using this approach [43.124].

3-D Packaging by Via-Last Approaches

In the technique shown in Fig. 43.25a, a completely processed and packaged MEMS device wafer is bonded to an ASIC wafer. Due to the thickness of the wafer containing the packaged MEMS devices, this technique allows only limited integration densities. In the techniques shown in Fig. 43.25b, a wafer containing the unpatterned MEMS device material(s) is bonded to the ASIC wafer. Thereafter, the handle wafer is released from the donor wafer. After the MEMS device material has been transferred to the ASIC wafer, it can be further processed and patterned. The vias between the devices and target wafer can then be formed. Finally, the sacrificial bond layer is removed by a selective etch. This technique has the advantage that accurate substrate-to-wafer alignment is not needed during the bonding process. The via and component positions on the target wafer are exclusively defined by the lithography and etching processes after bonding. Therefore, critical device dimensions and overlay accuracies in the nanometer range can be achieved, which is not possible with other techniques. The first MEMS device including fully functional CMOS ICs manufactured using very large-scale heterogeneous integration with a via-last approach was a 1-megapixel monocrystalline silicon micromirror array [43.126]. A polymer adhesive is used as the bonding layer, in combination with sputter-deposited aluminum vias [43.127], electroplated gold vias [43.128], or electroless-plated nickel vias [43.129]. Bonding with a polymer adhesive has the advantage that the MEMS device wafer and ASIC wafer can be bonded with very high yield and without any surface pretreatment or surface planarization. Planarized SiO_2 layers have also been proposed as an intermediate bonding layer for heterogeneous integration [43.130]. The tilting monocrystalline silicon mirrors are integrated on top of fully functional high-voltage CMOS drive electronics. The mirror array has resolution of 1 megapixel and pixel pitch of $16 \times 16\,\mu m^2$. The sil-

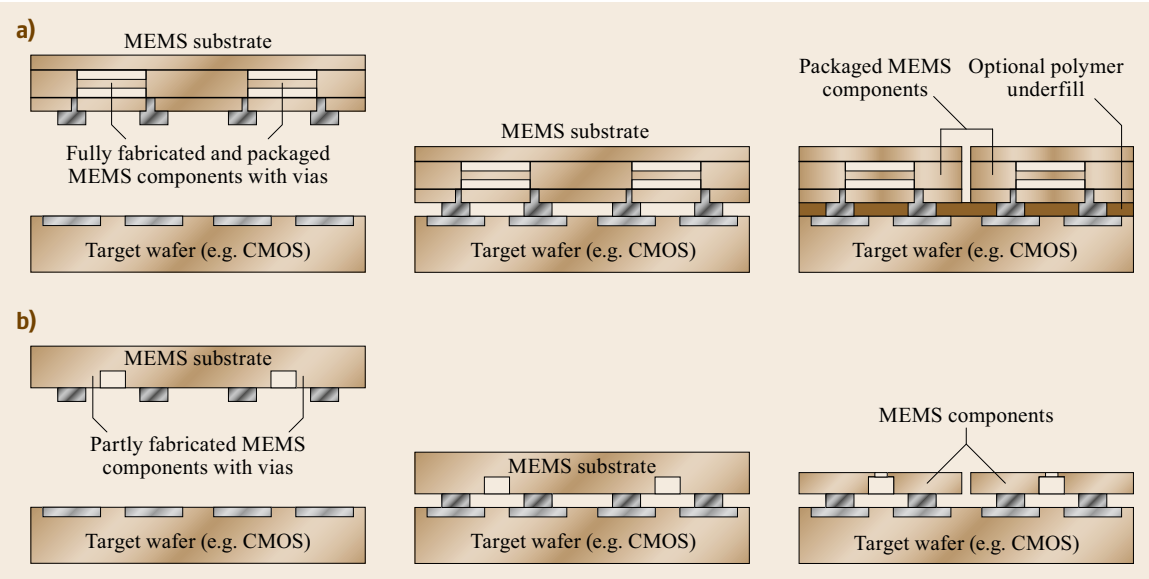
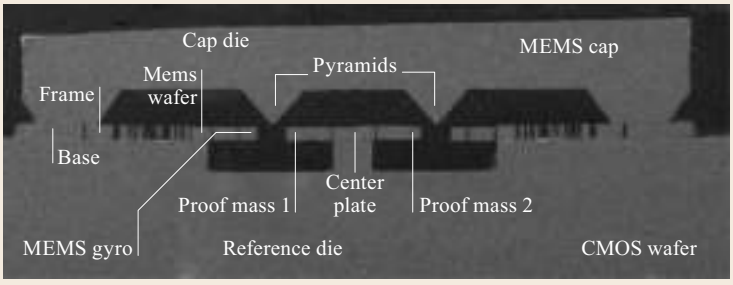


Fig. 43.23a,b Conceptual schemes of via-first 3-D packaging approaches. **(a)** Integration of fully processed MEMS components; **(b)** integration of partly processed and/or packaged MEMS components with subsequent wafer level processing (after [43.104])



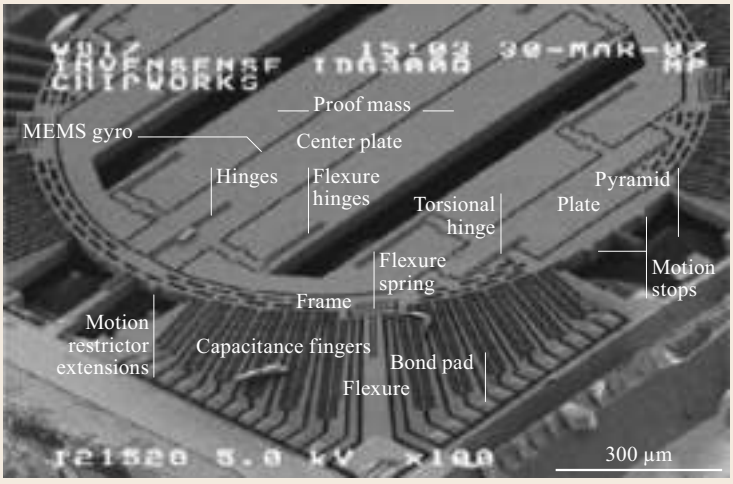


Fig. 43.24a,b Illustration of a gyroscope packaged with a via-first approach. **(a)** Cross-sectional view; **(b)** 3-D view (after [43.124], courtesy of Chipworks)

icon mirror membranes are 340 nm thick and have an extremely well-defined distance of 700 nm to the addressing electrodes on the underlying CMOS ICs. The mirror vias have diameter of 2 μm, and the torsional mirror hinges are 600 nm wide.

43.3.3 Polymer–MEMS Packaging

By using microheaters to provide heat locally for the bonding process, the global temperature can be significantly reduced to satisfy the strict low-temperature processing requirement for a variety of biomaterials and polymers. Polymeric materials are extensively utilized in microfluidic systems for applications in life sciences [43.131–133]. The advantages of polymer-MEMS include a broad range of choices of material properties, low raw material cost, and feasibility of mass production. To facilitate fabrication and application of polymer-MEMS, effective assembly and packaging processes are highly desired. In this section, two localized bonding and assembly schemes are introduced, employing soft thermoplastic materials with low glass-transition temperature as intermediate layers to demonstrate packaging of various systems, including polymer to silicon, polymer to glass, and polymer to polymer. The two packaging schemes are illustrated

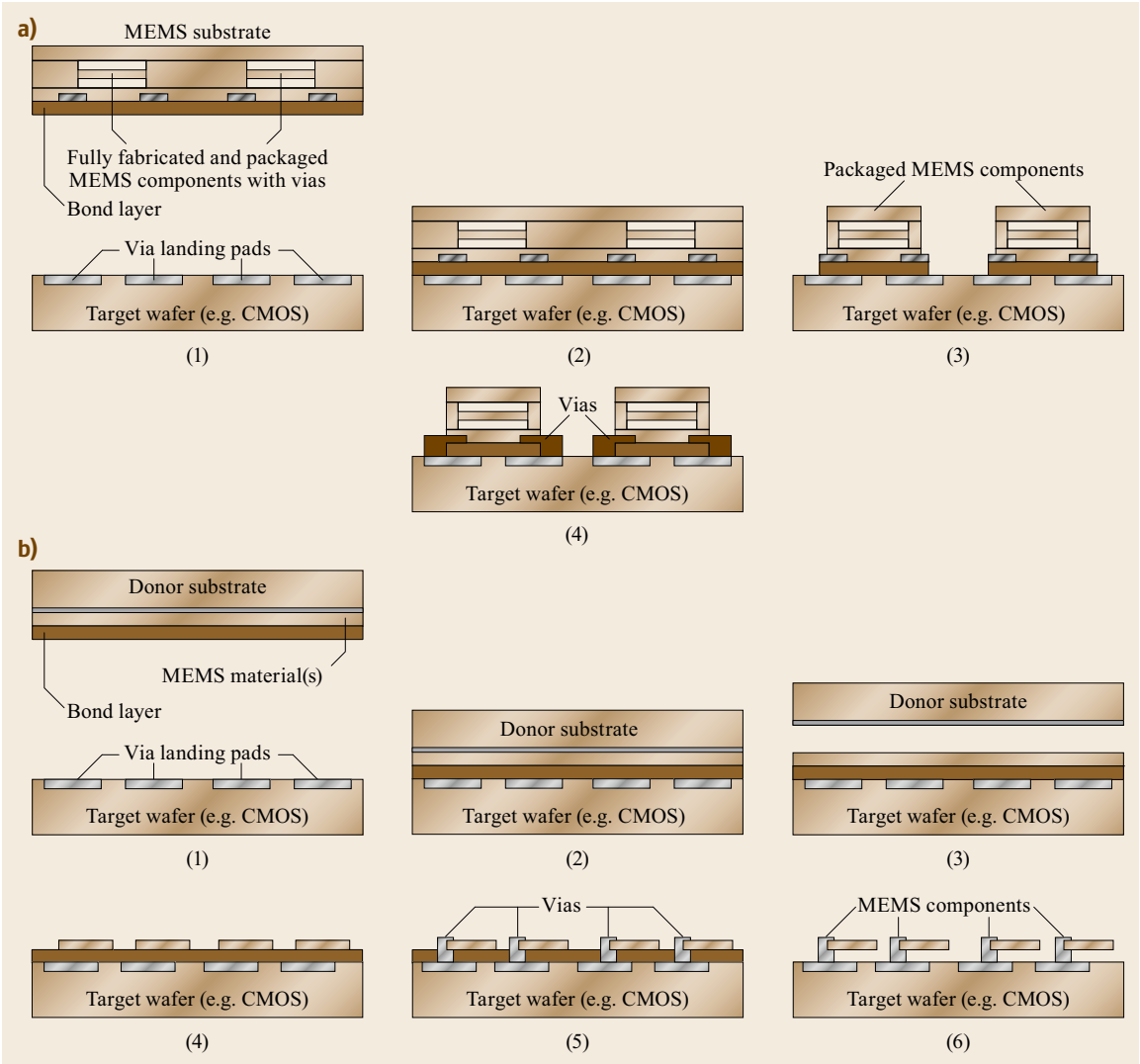


Fig. 43.25a,b Conceptual schemes of via-last 3-D packaging approaches. **(a)** Integration of fully processed and packaged MEMS components; **(b)** integration of MEMS material(s), release from donor wafer and subsequent processing of components (after [43.104])

in Fig. 43.26: (a) bonding by built-in-type heaters, and (b) bonding by external, reusable heaters.

A Mylar film (DuPont Teijin films) coated with a thermoplastic polyvinylidene chloride (PVDC) copolymer layer on one surface was bonded to either silicon, glass, or another polymeric substrate. The scheme with built-in microheaters is suitable for bonding thick polymeric materials with various types of substrate including silicon, glass, and polymer, because microheaters can easily provide the activation energy locally for bonding. Meanwhile, the scheme with external, reusable heaters is most suitable for bonding of thin polymer films to various substrates. Heat is transferred through the thin polymer film, and a heated zone is generated locally in the bonding interface where the bond is formed. In this case, it is preferable that the thin polymer film have two layers. The top layer should have high melting temperature to prevent bonding with the external reusable heaters, so that the heaters can be easily removed after completion of the bonding process and used repeatedly. Experimentally, aluminum wires of 30–70 μm width and 3 μm thickness were fabricated by a lift-off process and employed as the heating elements to form bonding loops of 1–2 mm in diameter on either silicon or glass substrates. Because of the small bonding area (less than $10^{-6}\,m^2$), about 0.4 N of force is sufficient to provide bonding pressure of 0.4 MPa. These aluminum heaters have resistance of approximately 0.8 Ω, and when voltages of 3.5 V are applied, about 15 W of power is generated locally to increase the temperature for bonding within about 0.25 s.

To test the quality of these bonding processes, several experiments were designed. In the first experiment, bonded systems were placed in a vacuum chamber and observed under an optical microscope. It was found that the top of the encapsulated chamber expanded to form a dome shape due to the 1 atm pressure difference across the membrane. The diameter of the bonding ring is 1.4 mm in this case, which corresponds to an effective area of $1.6\times10^{-6}\,m^2$, and the air permeability of Mylar is $8\,cc/m^2$ day atm. Based on a simplified approximation, it takes about 18 days for the air inside the cavity to diffuse out, and the membrane should return to flat. Figure 43.27 shows a SEM image of the result of Mylar bonded on PMMA. The dome shape can be observed as proof that a good seal was achieved. Afterwards, the polymer-to-polymer (Mylar-to-PMMA) bond was forcefully broken to examine the bonding interface under SEM (Fig. 43.28). It was observed that the bonding result was uniform, and part of the Mylar film was attached to the bonding substrate. Another experiment showed that direct encapsulation of water using localized heating can be achieved as a proof-of-concept demonstration of low-temperature processing. In this case, Mylar-to-PMMA bonding was performed with reusable heaters to encapsulate 0.18 μl of water in the cavity, as shown in Fig. 43.29. The close-up view at the top shows two small bubbles in the water-filled encapsulated chamber. These bubbles provide an easy way to verify the presence of water in the chamber. No leakage path can be identified, and water inside the package escapes mainly by evaporation and diffusion through the top Mylar membrane as well as the plastic substrate.

In summary, among the tested bonding systems using built-in aluminum heaters, it is found that plastic-to-plastic bonding shows the highest bonding strength. In the bonding systems using silicon or glass as the

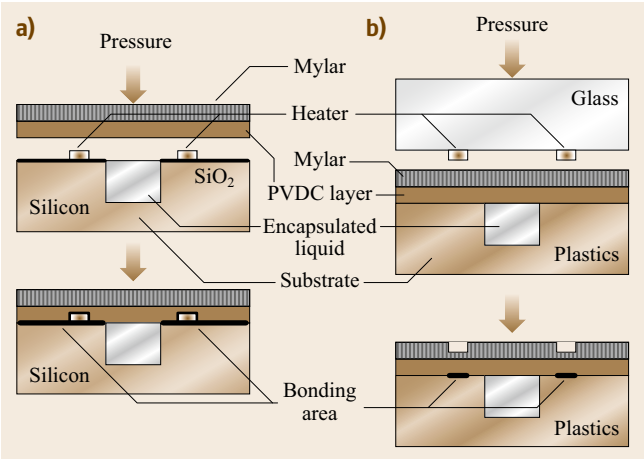


Fig. 43.26a,b Schematic illustration of polymer bonding processes in (a) built-in heater configuration and (b) reusable heater configuration

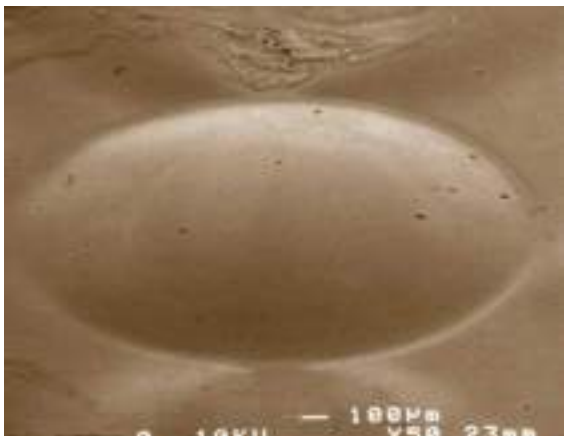


Fig. 43.27 SEM image of dome shape in Mylar membrane, demonstrating good sealing with poly(methyl methacrylate) (PMMA) (after [43.133])

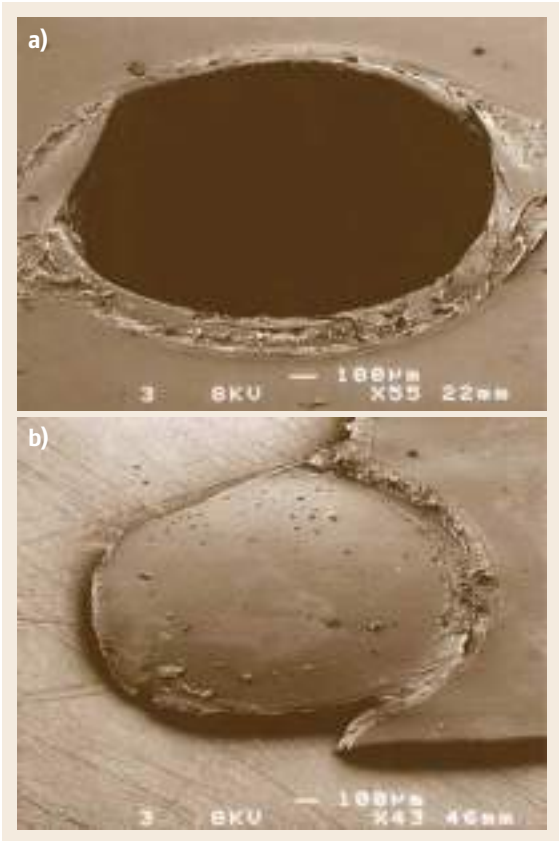


Fig. 43.28a,b SEM micrographs of bonding interface: **(a)** PMMA substrate, **(b)** Mylar layer (after [43.133])

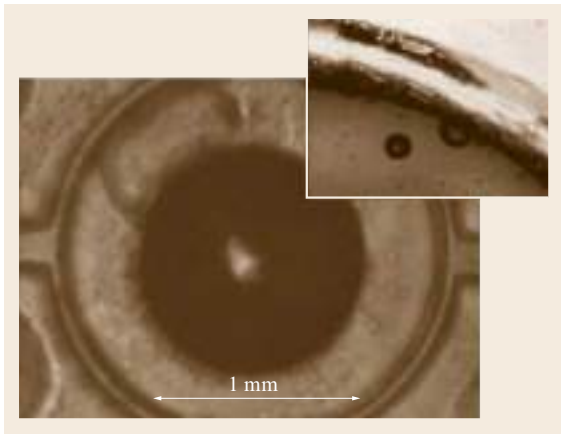


Fig. 43.29 Water encapsulation result (after [43.133])

bonding substrate, the bonding interface lies between the plastic and the aluminum heaters, and the bonding strength is lower than for plastic-to-plastic bonding but higher than that of plastic-to-silicon or plastic-to-glass bonds. Meanwhile, the plastic-to-plastic bonding process using reusable heaters shows the best bonding results as compared with plastic-to-silicon or plastic-to-glass systems. In the case of using external, reusable heaters for plastic-to-silicon and plastic-to-glass bonding systems, it is suggested that thin plastic films with (1) high adhesion chemistry with silicon and glass and (2) low melting temperature should be employed as the intermediate bonding layer to achieve high bonding quality.

43.4 Thermal Issues and Packaging Reliability

43.4.1 Thermal Issues in Packaging

The two key thermal issues related to MEMS packaging are:

- Heat dissipation from actuators and integrated circuitry components
- Thermal stress generated during the packaging process.

These two topics are discussed separately.

Heat Dissipation Issues

In an IC, heat dissipation becomes a serious problem as the size of transistors continues to shrink and their density on the chip keeps increase with advances in IC fabrication technology. The trend of packing increasing power into smaller packages has exacerbated thermal management challenges [43.134]. The electrical characteristics of transistors change with working temperature, so inefficient power dissipation that raises the working temperature can affect device performance. Present MEMS devices do not need high-power, high-performance microprocessors, so power dissipation is not a problem. Nevertheless, some functional components in packaged MEMS, such as biomaterials or laser diodes, are very sensitive to temperature variations. Several MEMS chemical sensors and other applications such as micro polymerase chain reaction (PCR) chambers for DNA replication actually require elevated temperature for operation, and microthermal platforms are built for these devices. Thermal management to maintain the working temperature of these chips for stable operation is still an essential packaging consideration. The geometrical complexity of MEMS resulting from packing various functional components into a tight space increases the difficulty of thermal

management. As the MEMS packaging integration process becomes more complex, fabrication constraints on the packaging process will have greater impact on the heterogeneous integration in front-end MEMS and IC processes; For example, a low temperature requirement for the packaging process generally limits the possible choice of materials in the back-end process. In general, conventional IC packaging employs a heat sink attached to the chip to remove heat. The heat sink is generally made of a copper or stainless-steel bar with an array of fin structures on one side for better natural or forced heat convection to dissipate heat to the environment. In addition to heat sinks, thermal vias, heat pipe cooling, immersion cooling, and thermoelectric cooling can also be used for effective heat removal. Because most MEMS packages still follow the typical IC packaging architecture, one promising thermal management method, viz. the heat pipe, is discussed for possible MEMS packaging applications.

A heat pipe is a sealed slender tube containing a wick structure and a working fluid, typically water for electronics cooling. It is composed of three sections: the evaporator section at one end, the condenser section at the other end, and an adiabatic section in the middle. In the evaporator section, heat is absorbed by the working fluid via phase transformation from liquid to vapor. In the condenser section, heat dissipates to the outside environment, and thus, the fluid returns to liquid phase. The vapor phase is in a high-pressure and high-temperature state, which forces the vapor to flow into the condenser section at a lower temperature. Once the vapor condenses and gives up its latent heat, the condensed fluid is then pumped back to the evaporator section by the capillary force developed in the wick structure. Therefore, the middle adiabatic section contains two phases: the vapor phase in the core region and the liquid phase in the wick, flowing in opposite directions and with no significant heat transfer between the fluid and the surrounding medium. Silicon has good thermal conductivity (1.41 W/(cm K)) and is easily micromachined to fabricate such heat pipes. Therefore, there is great potential for implementation of silicon micro heat pipes in IC and MEM packaging, and several approaches have been proposed on this topic [43.135–137].

Packaging-Induced Thermal Stresses

Thermal-based bonding processes have been used in MEMS packaging applications for many years, as described previously in this chapter. Thermal management is extremely important during the bonding process to avoid fracture of the substrate or MEMS devices themselves. Extremely high temperatures or rapid cooling conditions may result in damage and should be carefully evaluated both analytically and experimentally. There are many ways to provide heating energy, including electrical resistive heating, oven heating, or induction heating [43.100]. These bonding processes may be put into two categories: localized bonding, where heat is applied directly only to the adhesive material used to bond the package to the MEMS device, and global heating, where the entire system (MEMS device, adhesive, and packaging material) is heated to bond the materials, being the common approach for all MEMS packaging processes. Therefore, this section focuses on the thermal stress effects in MEMS packaging during heating and cooling procedures. The aluminum-to-glass bonding process using RTP is used as a specific example for the discussion of thermal stresses [43.92]. The bonding process heats up the packaging system to 750 °C for 10 s, then cools it back to room temperature. To simulate this process, an ANSYS program [43.138] was established to examine the shear stress due to coefficient of thermal expansion (CTE) variations in the bonding system as a result of temperature changes. The shear stress was recorded from the ANSYS analysis on the aluminum/Pyrex glass interface and the aluminum/silicon interface.

Two different models were analyzed. The first was for the quartz–aluminum–silicon bonding system, and the second was for the Pyrex glass–aluminum–silicon bonding system [43.139]. The results of the ANSYS analysis were then analyzed and compared with experimental observations. Figure 43.30 shows the ANSYS results for the Pyrex glass bonding system with aluminum solder width of 100 μm; the maximum residual stress is 60 MPa in the glass, slightly lower than the fracture strength of Pyrex glass of 70 MPa. It was discovered that increasing the aluminum width led to lower residual stresses. This likely occurred because the

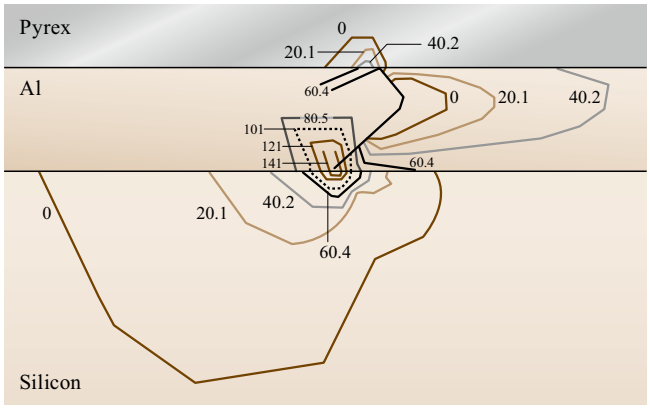


Fig. 43.30 Residual stress (GPa) for aluminum solder width of 100 μm in silicon–aluminum–glass bonding using RTP

length of the Pyrex glass, quartz, and silicon remained constant. As a result, it will always want to contract by the same amount for a constant temperature change, independent of the aluminum width. However, when increasing the aluminum width, the stress did not occur in such a concentrated area and therefore decreased; For example, the maximum residual stress analyzed from ANSYS in the Pyrex glass bonding system was 74.5, 58, and 60 GPa for aluminum width of 30, 50, and 100 μm, respectively [43.139]. Pyrex glass has documented strength of around 69 GPa [43.140]. Fracture should always occur for aluminum width of 30 μm or less, according to the ANSYS analysis. Fracture may occur sporadically at width of 50 or 100 μm, depending on the amount and magnitude of the flaws in the Pyrex glass. Experiments were carried out on the Pyrex glass bonding system with a width of 100 μm. The samples were heated up to 750 °C, then cooled down by taking them out of the oven. In all four experimental cases, small cracks were observed in the Pyrex glass, as shown in Fig. 43.31. These cracks may have occurred consistently for several reasons. First, they may be a result of handling the Pyrex glass before bonding. The Pyrex glass samples were kept in containers with each other, which may have resulted in abrasive contact and possibly caused flaws in the material. These flaws could result in reduced strength below the value of 69 GPa, so the predictions based on the ANSYS analysis could be correct. Second, it was observed that the cracks were small, only occurring tens of microns away from the aluminum and not propagating completely through the Pyrex glass. These cracks could be caused by the high stress applied, but the cracks did not reach a critical size and therefore did not propagate completely through the Pyrex glass. Therefore, the strength remained at the theoretical value of 69 GPa, and the Pyrex glass only partially cracked. Experimental analysis by *Chiao* and *Lin* [43.92] showed that fracture was not observed when using aluminum width greater than 150 μm. This is consistent with the results of the ANSYS analysis showing that, as the width of the aluminum is increased, the residual stress decreases.

ANSYS calculations were also carried out for the quartz bonding system, predicting maximum stress of 207, 117, and 100 GPa, for aluminum width of 30, 50, and 100 μm, respectively. All three of these stress values are much larger than the theoretical strength of quartz at 48 GPa, so fracture should always occur. Quartz has much higher coefficient of thermal expansion (CTE) than silicon, which explains this prediction. Experimentally, a quartz substrate was used to test silicon–aluminum–quartz bonding; the result is shown in Fig. 43.32. It was observed that cracks occurred all over the sample, causing serious damage to the quartz wafer. These cracks could be the failure mechanism of the hermetic package. Therefore, Pyrex glass was identified as a better bonding substrate than quartz.

The thermal stresses generated in the packaging process with quartz are much larger than for Pyrex glass, because of the different CTE mismatches in these two systems. Quartz has low CTE ($0.54\times10^{-6}\,\mathrm{K}^{-1}$) compared with aluminum ($23\times10^{-6}\,\mathrm{K}^{-1}$) or silicon ($3.5\times10^{-6}\,\mathrm{K}^{-1}$). On the other hand, Pyrex glass has a CTE ($3.2\times10^{-6}\,\mathrm{K}^{-1}$) much closer to those of silicon and aluminum, resulting in smaller stresses. The practical implications of the ANSYS results and the information presented above are that materials must be chosen carefully when carrying out bonding. To ensure that fracture will not occur, materials with CTE much higher or lower than that of silicon should not be used. This finding is a valuable feature of this particular packaging system and should benefit other packaging processes involving bonding of MEMS packaging, because it supports the prediction that Pyrex glass is an excellent material to bond with silicon in MEMS packaging, as long as sufficiently wide adhesive material is used.

43.4.2 Packaging Reliability

Packaging is one of the key issues to be addressed in evaluation of the reliability of MEMS products. Any

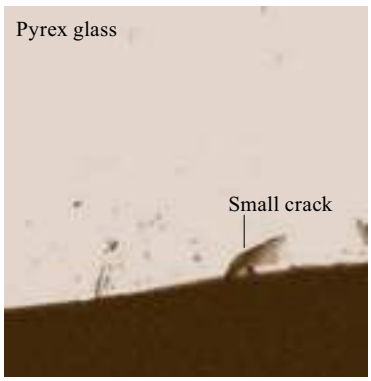


Fig. 43.31 Micrograph of experimental result for Pyrex glass–aluminum–silicon system. Small cracks can be observed



Fig. 43.32 Bonding result of the quartz–aluminum–silicon system; fracture can be observed

defects created during the sealing and packaging process may result in immediate device failure or may degrade device performance over time; For example, microaccelerometers that are used to deploy airbags in automobile safety applications require excellent reliability. If any leakage path is created during the sealing process at the two bonding interface, moisture may be able to enter the sealed microcavity and cause device failure over time. Thermal stress induced by the CTE mismatch is one of the main factors affecting packaging reliability. In fact, stress formation can occur not only during the packaging process but also during device operation. In particular, during device operation, the package will go through various temperature cycles because of environmental changes. Such temperature variation can cause expansion of packaging materials when constrained in the packaged assembly. As a result of such thermal mismatch, significant stresses are induced in the package and may finally cause device failure. In addition to thermal mismatch, corrosion, creep, fracture, fatigue crack initiation and propagation, and delamination of thin films are all possible factors that can cause failure of packaged devices [43.141]. These failure mechanisms can be prevented or deferred by using proper packaging designs; For instance, thermally induced strain inside the packaging material is generally below the tolerance of the material and cannot cause immediate catastrophic damage. However, cyclic loading can generate and accumulate stresses and eventually cause failure. Several common designs have been used in IC packaging to prolong device lifetime; For example, the strain in solder interconnects of BGA or flip-chip packaging can be effectively reduced by introducing a polymer underfill material between the chip and substrate for effective distribution of thermal stresses induced by CTE mismatch [43.142]. Such strain can be further reduced if excellent thermal paths are built around interconnects to diminish thermal stresses originated from the temperature gradient between the ambient and operation temperature. Delamination is another source of reliability problems, occurring at the interface between adjacent material layers. In MEMS, components made of dissimilar materials are commonly bonded together to provide specific functions. Delamination can result in electrical or mechanical failure of devices in the package, such as mechanically cracking through the electrical via wall to form an electrical open because of propagation of the delamination of the metal line from the dielectric layer or overheating of the die because of delamination of the die from the underlying layer to form a gap in the heat dissipation path. Because of the complex stress and thermal loading, geometry, and material properties of MEMS, development of packaging designs to increase reliability is very important and requires more extensive investigation.

Reliability testing is required before a new device can be delivered to the market. Test results can provide information for subsequent improvement of the packaging design and fabrication processes. Hence, the approach used to analyze failure data, known as *reliability metrology*, is very important in the packaging industry. This analysis method uses the mathematical tools of probability and statistical distributions to evaluate data, understand failure patterns, and identify sources of failure; For example, the failure density function is defined as the time derivative of the cumulative failure function

$$f(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t} \tag{43.1}$$

$$F(t) = \int_0^t f(s)\,\mathrm{d}s\,. \tag{43.2}$$

The cumulative failure function $F(t)$ is the fraction of a group of original devices that has failed at time t. The Weibull distribution function is one of the analytical mathematic models commonly used in packaging reliability evaluation to represent the failure density function [43.12].

$$f(t) = \frac{\beta}{\lambda}\left(\frac{t}{\lambda}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right], \tag{43.3}$$

where β and λ are the Weibull parameters. The parameter β is called the shape factor and measures how the failure frequency is distributed around the average lifetime. The parameter λ is called the lifetime parameter and indicates the time at which 63.2% of the devices have failed. By integrating both sides of the equation, $F(t)$ becomes

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right]. \tag{43.4}$$

Using the Weibull distribution function with the two parameters extrapolated from experimental data, one can estimate the number of failures at any time during a test. Moreover, knowing the meaning and values of these parameters, one can compare two sets of test data; For example, higher λ indicates that a set of samples has longer lifetime. Because all such mathematical models are statistical approximations based on real experimental data, use of more testing samples can provide more accurate estimations.

43.4.3 Long-Term and Accelerated MEMS Packaging Tests

The ability to estimate the reliability or lifetime of a device provides valuable information for the manufacturer to maximize the profit margin by balancing the cost and quality of the product. Moreover, the warranty period given by the manufacturer has to be determined based on product reliability information. The reliability of MEMS packages is best characterized using long-term tests with statistical data analyses. However, it is very difficult to measure the reliability or lifetime of a device in a real-time fashion, because testing during a prolonged time period may be required to prompt many devices to fail. To evaluate the reliability of a device in a timely fashion, accelerated testing is normally conducted to speed up the device aging process and thus shorten the total testing time required. Accelerated testing, from the packaging and sealing point of view, is a testing method that emphasizes failure of the seal when foreign elements leak inside the microcavity, which may affect device performance. For a hermetic package, the lifetime of a MEMS device is essentially an estimation of the time required for water to penetrate into the package. For vacuum-encapsulated MEMS devices, in addition to water penetration through the seal, gas penetration or outgassing from within packaging materials such as the substrate, cap, and seal over time can degrade the vacuum level and thus device performance. Therefore, the lifetime of a vacuum-encapsulated MEMS package can be evaluated using the time for gas to evolve into the package from either the seal or device materials, whichever occurs first. Unfortunately, there are not many research publications that deal with such long-term and accelerated testing to evaluate MEMS packaging reliability. In the conventional IC packaging industry, reliability estimation is carried out using accelerated testing and statistical predictions [43.12]. Accelerated tests often utilize high temperature and high humidity, e.g., autoclave tests [43.143], to speed up corrosion of the sealing boundary and thereby accelerate package failure. The MEMS industry could use very similar accelerated tests to estimate the lifetime of a MEMS package, because the basic assumptions regarding the failure mode and humidity issues are similar to those for conventional IC packages. Several research groups have reported reliability studies on MEMS packages formed using different bonding methods and materials [43.93, 94, 144]. In this section, two MEMS packaging examples that aim to address long-term and accelerated testing are discussed.

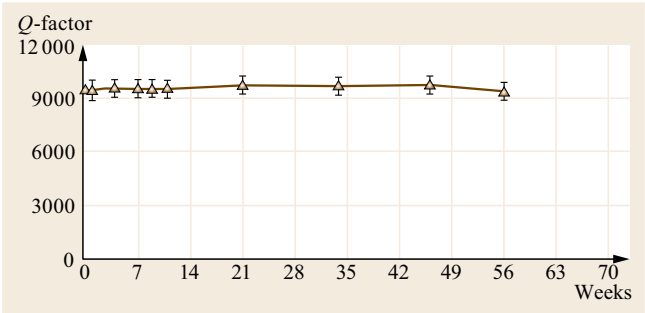


Fig. 43.33 Long-term measurement of encapsulated μ-resonators. No degradation of Q-factor is found after 56 weeks

Figure 43.33 shows long-term measurements of the Q factor of vacuum-packaged μ-resonators obtained using localized aluminum/silicon-to-glass bonding [43.93]. The vacuum encapsulation process is described in detail in Fig. 43.8. It was found that the vacuum package obtained by localized heating and bonding provided a stable vacuum environment for the μ-resonator with quality factor of 9600, showing no degradation over at least one year. Since the performance of high-Q μ-resonators is very sensitive to environmental pressure, as shown in Fig. 43.17, any leakage can be easily detected. The fact that this high Q value can be maintained for one year indicates that the packaging process was performed well and that both aluminum and Pyrex glass are suitable materials for use in vacuum packaging applications. According to a previous study of hermeticity in different materials, metal has lower permeability to moisture than other materials such as glass, epoxy, and silicon. With width of 1 μm, metal can effectively block moisture for more than 10 years [43.12]. In this vacuum packaging system, the bonding width is 30 μm, such that it can sufficiently block the diffusion process of moisture. On the other hand, the effects of diffusion of air molecules into these tiny cavities have not been studied extensively, and design guidelines for vacuum encapsulation are not clearly defined. Further investigations are needed in this area, and the example presented here serves as a good starting point.

On the other hand, accelerated testing involves placing a large number of samples in a harsh environment, such as elevated temperature, elevated pressure, and 100% humidity, to accelerate the corrosion process. Statistical failure data are then gathered and analyzed to predict the lifetime of packages under normal usage environment. As a result, the long-term reliability of the package can be predicted without going through true long-term tests. Unfortunately, accelerated testing is an area that has not been addressed in MEMS research papers. Although the MEMS industry must have done some extensive reliability tests, they do not publish the results, probably due to liability concerns. Among the

very limited publications, this section uses a specific MEMS packaging system that has gone through accelerated tests as an illustrative example [43.41].

The considered MEMS package was fabricated using RTP bonding, as described previously in this chapter. The goal of the accelerated testing was to examine the failure rate at the bonding interface. The accelerated test was started by placing the packaged samples into an autoclave chamber filled with high-temperature (130 °C) pressurized (2.7 atm) steam at 100% relative humidity for accelerated testing. Pressurized steam can penetrate small crevasses if any defect is present at the bonding interface [43.59]. The elevated temperature and humid environment speed up the corrosion process. A package was considered to have failed if water condensed or diffused into the package. The statistical data gathered from this accelerated test can be categorized as right-censored data [43.145]. Statistical failure data were gathered every 24 h by optical examination for a period of 864 h, during which new failures were seldom observed (therefore, right-censored on time axis). In practice, this method was easier and more economical to implement than other methods. Owing to the robustness of the samples, it was difficult to complete the tests to the point where all packages failed. The cumulative failure function $F(t)$ is defined as

$$F(t) = \frac{\text{Number of cumulative failures}}{\text{Number of samples, } N}\,, \qquad (43.5)$$

where N is the sample size at the beginning of the test. A package was considered to have failed if water condensed inside or diffused into the package; For example, water was found to diffuse into the cavity after 240 h of testing in Fig. 43.34. However, no leakage path could be identified under optical microscopy in this case. Figure 43.35 shows the function $F(t)$ (in %) plotted versus the logarithm of time. In general, most failures occurred in the first 96 h ($\ln(t) \approx 4.56$); this high number of early failures reflects a yield issue of the sealing process. Moreover, packages with smaller bonding width and larger bonding area showed higher failure percentages. Both statistical models, i. e., Weibull and lognormal [43.145], were used and compared to analyze the collected data to predict the lifetime of the packages, using the least-squares fit method to determine the best fitting model. It was found that the R^2, the coefficient of determination [43.145], values were generally in the range of 0.8 when using the lognormal model as compared with 0.5 when using the Weibull model. Therefore, the lognormal model was used to predict the lifetime of the packages.

Figure 43.36 shows the inverse standard normal distribution function versus ln(time). The maximum-likelihood estimator (MLE) was then used to predict the mean, standard deviation, and mean time to failure (MTTF). Table 43.2 presents the MLE calculation results for the MTTF. The wide confidence interval results from the fact that only a small number of samples had failed by the end of the test. It was also observed that packages with larger bonding width and smaller bonding area had longer MTTF values. The lower bound of the MTTF provides the worst-case scenario; For example, only 4 out of 31 samples had failed by the end of the test in the case of ring width of 200 µm and sealing area of $450 \times 450\,\mu m^2$. The MTTF predicts, in the worst-case scenario, a 90% chance that a package will fail in 0.57 years in the autoclave environment.

It is widely accepted that the acceleration factor (AF) for autoclave tests follows the Arrhenius equa-

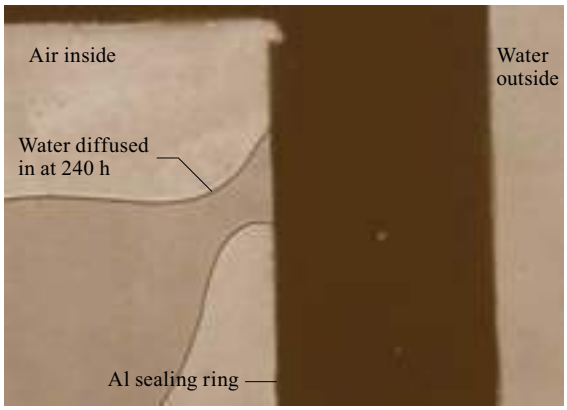


Fig. 43.34 A particular device that failed at 240 h of testing time

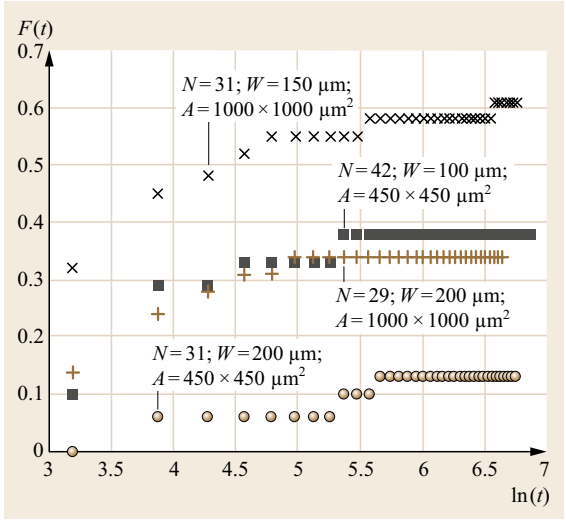


Fig. 43.35 Cumulative failure data

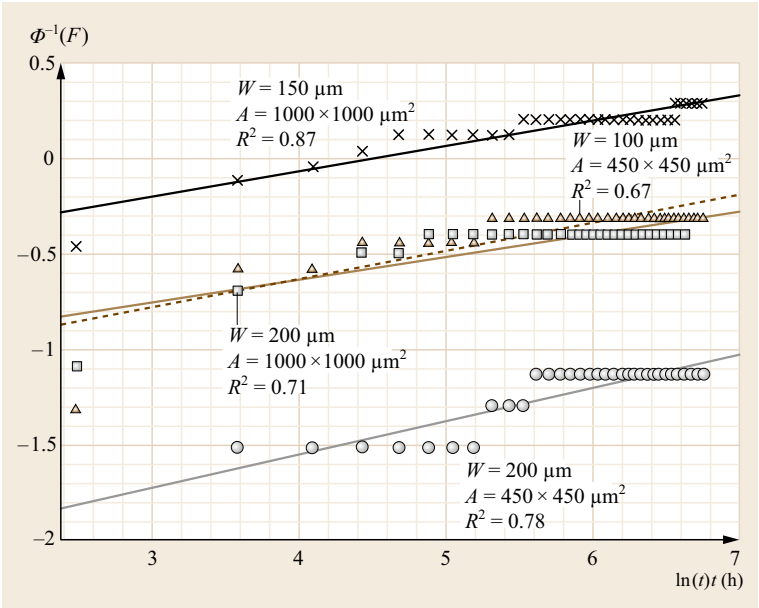


Fig. 43.36 Life data fitted by log-normal distribution. R^2 is the coefficient of determination

Table 43.2 Maximum-likelihood estimation of mean time to failure (MTTF)

MLE calculation results for MTTF[a]				
Bonding width, W (µm)	**Area, A (µm²)**	**MTTF**		**Worst case in jungle condition (years)**
		UB (years)	**LB (years)**	
200	450 × 450	1.8×10^7	0.57	1700
100	450 × 450	5.3	0.10	300
200	1000 × 1000	6.5×10^3	0.09	270
150	1000 × 1000	0.50	0.017	50

[a] UB is the upper bound and LB is the lower bound of the 90% confidence interval, respectively. The MTTF LB times AF is the worst case MTTF used in jungle condition

tion [43.12] and can be modeled as

$$\mathrm{AF} = \frac{\left(\mathrm{RH}^{-n} \mathrm{e}^{\Delta E_\mathrm{a}/k_\mathrm{B}T}\right)_\mathrm{normal}}{\left(\mathrm{RH}^{-n} \mathrm{e}^{\Delta E_\mathrm{a}/k_\mathrm{B}T}\right)_\mathrm{accelerated}}, \tag{43.6}$$

where RH is the relative humidity (85%, RH = 85), k_B is the Boltzmann constant, and T is the absolute temperature. The recommended value for n, an empirical constant, is 3.0 [43.146], while that for ΔE_a, the activation energy, is 0.9 eV for a plastic dip package and 0.997 eV for an anodically bonded glass-to-silicon package [43.59]. Using $\Delta E_\mathrm{a} = 0.9$ eV, an AF of about 3000 is estimated for the accelerated testing condition as compared with the jungle condition (35 °C, 1 atm, and 95% RH); the corresponding worst-case lifetime values in jungle condition are also listed in Table 43.2. The high values of estimated MTTF in jungle condition could be a result of overestimation of the AF because the plastic dip package may have lower AF compared with glass packages. Nevertheless, these data and analyses provide important guidelines in the area of accelerated testing of MEMS packages.

For vacuum-packaged MEMS devices, the lifetime can be evaluated by monitoring the quality factor of microresonators inside sealed cavities. Again, vacuum-packaged MEMS resonators obtained by aluminum-to-nitride bonding using RTP are discussed in detail here to illustrate the various factors involved in reliability. It was found that, under normal condition of room-temperature storage, the quality factor of resonators remained constant after 37 weeks, as shown in Fig. 43.37. Furthermore, the vacuum quality in harsh environment was characterized by placing a vacuum-encapsulated comb resonator into an autoclave chamber

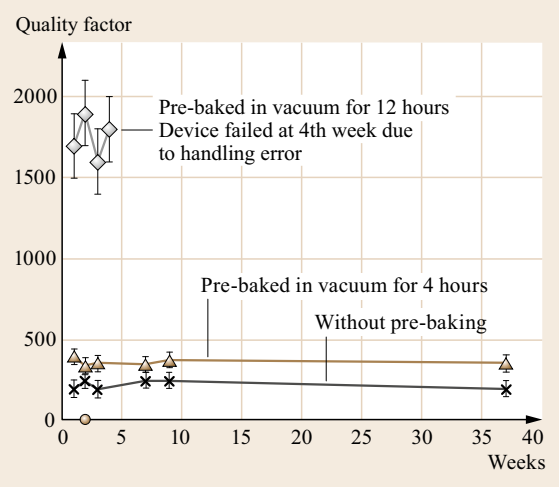


Fig. 43.37 Long-term stability tests to 37 weeks. The Q factor increased with prebaking time

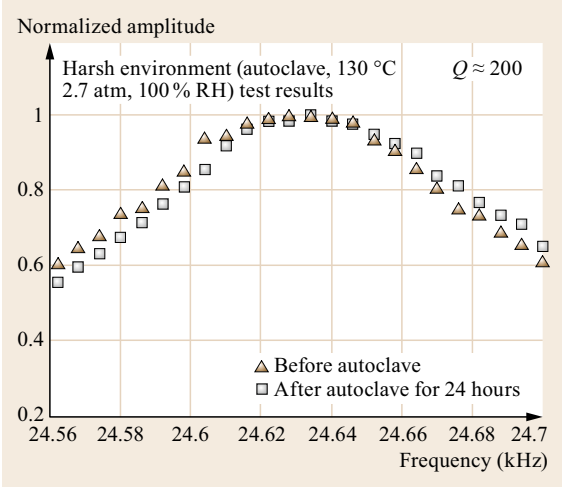


Fig. 43.38 Spectrum measured before and after accelerated testing for 24 h

(130 °C, 2.7 atm, and 100% RH) for accelerated testing. The result is shown in Fig. 43.38; the quality factor remained at 200 after 24 h in the autoclave testing chamber. Since the slight differences between the two spectra are within normal experimental errors, it can be concluded that the harsh environment in the test did not affect the vacuum seal.

To characterize the vacuum lifetime of the packages, two vacuum-packaged comb resonators were placed in the harsh environment for continuous testing for up to 1008 h; the results are summarized in Table 43.3. The quality factor of each package was measured in every 24 h interval, being found to remain at 400 and 200, respectively, before failure. The first packaged resonator with Q of 200 had aluminum sealing ring width of 75 μm and sealing area of $650 \times 650\,\mu m^2$. The second packaged micro resonator had Q value of 400 and aluminum sealing ring width of 200 μm and sealing area of $550 \times 550\,\mu m^2$. If the accelerated testing results on water penetration [43.94] are applied here for gas penetration as a measure of the vacuum sealing characteristics, the accelerated lifetime of the first and second packaged resonator would fall in the range of 0.017–0.1 years (149–876 h) and 0.09–0.57 years (769–4993 h), respectively. Experimentally, the first packaged resonator failed at 576 h into the test in the autoclave chamber. This corresponds well to the lifetime prediction from the previous work [43.94]. The second packaged resonator survived in the autoclave chamber for more than 1008 h (the device did not fail), and this result also verifies the prediction made from the previous work. However, note that these results are preliminary data, and more tests on more packaged devices should be conducted to enable meaningful statistical analyses.

Table 43.3 Summary of accelerated testing results on two vacuum-packaged resonators

	Package 1	Package 2
Q	400	200
Time under testing (h)	> 1008	576
MTTF, lower bound (h)	769	149

43.5 Future Trends and Summary

In the past, development of MEMS packaging mainly originated from IC packaging advancement, because existing packaging techniques could significantly reduce the development cost of MEMS. However, it is expected that this situation will change very soon such that MEMS packaging approaches will assist IC packaging development. Recent progress in IC packaging has aimed to provide high I/O density and greater chip integration capability to meet needs for higher speed and higher data communication rates. To satisfy these requirements, several packaging concepts and techniques have been developed, including 3-D packaging, wafer-level packaging, BGA, and flip-chip technique. Although all of these concepts and methods can provide packages with greater I/O density, flexibility in terms of chip integration, and lower manufacturing cost for IC fabrication, they are still insufficient to provide solutions for future applications,

because of the increasing complexity and requirements of MEMS packaging. In contract, with the progress of MEMS fabrication technologies, several key processes such as deep reactive-ion etching (DRIE), wafer bonding, and thick photoresist processes [43.147] have been utilized for IC packaging fabrication. Therefore, technologies developed for MEMS fabrication can also assist development of new IC packaging approaches.

To address future needs for process integration, adaptive multichip module (MCM) [43.31] or 3-D packaging combined with vertical through-substrate interconnects [43.10, 148] are promising approaches for development of future MEMS packaging processes. Based on low-temperature flip-chip solder bonding technique, these packaging methods can provide greater flexibility in terms of device fabrication and packaging. Devices can be fabricated before they are integrated together to form microsystems, thereby dramatically reducing packaging costs. Vertical through-substrate interconnects can achieve higher I/O density as well as lower resistance, parasitic capacitance, and mutual inductance. Although this approach offers many possible advantages, technical challenges remain; For instance, metal is commonly used as the filler material inside vertical vias to form electrical interconnects, which can introduce large thermal mismatch with respect to silicon substrate and generate huge thermal stresses that cause packaging reliability problems. Moreover, filling materials into these high-aspect-ratio vias will be an interesting engineering challenge.

The future development of MEMS packaging depends on successful implementation of various unique techniques:

1. Development of mechanical, thermal, and electrical models for packaging designs and fabrication processes
2. Wafer-level, chip-scale packaging with low packaging cost and high yield
3. Effective testing techniques at wafer level to reduce testing costs
4. Device integration by vertical through interconnects as an interposer [43.1] to avoid thermal mismatch problems.

In addition to these approaches and challenges, there are many other possibilities that have not been listed but that also require dedicated investigation; For example, several key nanotechnologies have been introduced in previous chapters, but packaging solutions for such NEMS devices have not been addressed. Because it is feasible to use MEMS as a platform for NEMS fabrication, all the packaging issues discussed in this chapter apply directly to NEMS devices as well. On the other hand, nanotechnology may introduce new opportunities for MEMS/NEMS packaging applications by providing superior electrical, mechanical, and thermal properties [43.149–152]; For example, carbon nanotubes have very high thermal conductivity [43.151] and may be applicable to enhance thermal cooling effects for improved IC/MEMS/NEMS packaging applications.

In summary, this chapter has introduced MEMS packaging issues in the areas of fabrication, application, reliability, and future development. Design and modeling, material selection, process integration, and cost are main issues to be considered when developing a new MEMS packaging process.

References

43.1 M. Matsuo, N. Hayasaka, K. Okumura, E. Hosomi, C. Takubo: Silicon interposer technology for high-density package. In: *Proc. IEEE Electron. Compon. Technol. Conf.* (2000) pp. 1455–1459

43.2 K.E. Peterson: Silicon as a mechanical material, Proc. IEEE **70**, 420–457 (1982)

43.3 L. Lin, H.-C. Chu, Y.-W. Lu: A simulation program for the sensitivity and linearity of piezoresistive pressure sensors, IEEE/ASME J. Microelectromech. Syst. **8**, 514–522 (1999)

43.4 Y.C. Tai, R.S. Muller: Lightly-doped polysilicon bridge as a flow meter, Sens. Actuators **15**, 63–75 (1988)

43.5 D. Hicks, S.-C. Chang, M.W. Putty, D.S. Eddy: Piezoelectrically activated resonant bridge microacceleromenter. In: *Tech. Digest Solid-State Sens. Actuators Workshop* (1994) pp. 225–228

43.6 J. Berstein, S. Cho, A. King, A. Kourepenis, P. Maciel, M. Weinberg: A micromachined comb-drive tuning fork rate gyroscope. In: *6th IEEE Int. Conf. MEMS* (1993) pp. 143–148

43.7 M. Madou: Compatibility and incompatibility of chemical sensors and analytical equipment with micromachining. In: *Tech. Digest Solid-State Sens. Actuators Workshop* (1994) pp. 164–171

43.8 M.J. Zdeblick, J.B. Angell: A microminiature electric-to-fluidic valve. In: *Proc. Trandsucers'87, 4th Int. Conf. Solid-State Transducers Actuators* (1987) pp. 827–829

43.9 J.H. Tsai, L. Lin: A thermal bubble actuated micro nozzle-diffuser pump, IEEE/ASME J. Microelectromech. Syst. **11**, 665–671 (2002)

43.10 S.J. Ok, D. Baldwin: High density, aspect ratio through-wafer electrical interconnect vias for low

cost, generic modular MEMS packaging. In: *8th IEEE Int. Symp. Adv. Packag. Mater.* (2002) pp. 8–11
43.11 H. Reichl, V. Grosser: Overview and development trends in the field of MEMS packaging. In: *14th IEEE Int. Conf. MEMS* (2001) pp. 1–5
43.12 R.R. Tummala, E.J. Rymaszewski, A.G. Klopfenstein: *Microelectronics Packaging Handbook, Semiconductor Packaging* (Springer, Dordrecht 1997)
43.13 C.A. Harper (Ed.): *Electronic Packaging and Interconnection Handbook* (McGraw-Hill, New York 1991)
43.14 P. Kapur, P.M. McVittie, K. Saraswat: Technology and reliability constrained future copper interconnects–Part II: Performance implication, IEEE Trans. Election Devices **49**, 598–604 (2002)
43.15 J. Lau: *Flip Chip Technologies* (McGraw-Hill, New York 1996)
43.16 R. Prasad: *Surface Mount Technology: Principles and Practice*, 2nd edn. (Chapman Hall, New York 1989)
43.17 M. Töpper, J. Auersperg, V. Glaw, K. Kaskoun, E. Prack, B. Keser, P. Coskina, D. Jäger, D. Petter, O. Ehrmann, K. Samulewicz, C. Meinherz, S. Fehlberg, C. Karduck, H. Reichl: Fab integrated packaging (FIP) a new concept for high reliability wafer-level chip size packaging. In: *IEEE Electron. Compon. Technol. Conf.* (2000) pp. 74–81
43.18 S. Savastiouk, O. Siniaguine, E. Korczynski: 3-D wafer level packaging. In: *Int. Conf. High-Density Interconnect Syst. Packag.* (2000) pp. 26–31
43.19 J.H. Tsai, L. Lin: Micro-to-macro fluidic interconnectors with an integrated polymer sealant, J. Micromech. Microeng. **11**, 577–581 (2001)
43.20 T.A. Core, W.K. Tsang, S. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. **36**, 39–47 (1993)
43.21 K. Koester, R. Majedevan, A. Shishkoff, K. Marcus: *Multi-User MEMS Processes (MUMPS) Introduction and Design Rules, Rev. 4* (MCNC MEMS Technology Applications Center, Research Triangle Park 1996)
43.22 R. Lengtenberg, H.A.C. Tilmans: Electrically driven vacuum-encapsulated polysilicon resonator, Part I: Design and fabrication, Sens. Actuators A **45**, 57–66 (1994)
43.23 R.L. Smith, S.D. Collins: Micromachined packaging for chemical microsensors, IEEE Trans. Electron Devices **35**, 787–792 (1988)
43.24 M.C. Wu: Micromachining for optical and optoelectronic systems, Proc. IEEE **85**(11), 1833–1856 (1997)
43.25 M. Putty, K. Najafi: A micromachined gyroscope. In: *Solid-State Sens. Actuators Workshop* (1994) pp. 212–220
43.26 S. Wolf: *Silicon Processing for the VLSI Era, Vol I: Process Technology* (Lattice, Sunset Beach 1995)
43.27 J.K. Bhardway, H. Ashraf: Advanced silicon etching using high density plasmas, SPIE Micromach. Fabr. Technol. **2639**, 224–233 (1995)
43.28 W.H. Ko, J.T. Suminto, G.J. Yeh: Bonding techniques for microsensors. In: *Micromachining and Micropackaging for Transducers*, ed. by C.D. Fung, P.W. Cheung, W.H. Ko, D.G. Fleming (Elsevier, Amsterdam 1985)
43.29 S.D. Senturia, R.L. Smith: Microsensor packaging and system partitioning, Sens. Actuators **15**, 221–234 (1988)
43.30 J.T. Butler, V.M. Bright, P.B. Chu, R.J. Saia: Adapting multichip module foundries for MEMS packaging. In: *IEEE Int. Conf. Multichip Modul. High Density Packag.* (1998) pp. 106–111
43.31 M. Schuenemann, A.J. Kourosh, R.V. Grosse, R. Leutenbauer, G. Bauer, W. Schaefer, H. Reichl: MEM modular packaging and interfaces. In: *IEEE Electron. Compon. Technol. Conf.* (2000) pp. 681–688
43.32 D.W. Lee, T. Ono, T. Abe, M. Esashi: Fabrication of microprobe array with sub-100nm nano-heater for nanometric thermal imaging and data storage. In: *Proc. IEEE Micro Electro Mech. Syst. Conf.* (2001) pp. 204–207
43.33 A.S. Laskar, B. Blythe: Epoxy multichip modules, a solution to the problem of packaging and interconnection of sensors and signal-processing chips, Sens. Actuators A **36**, 1–27 (1993)
43.34 R. Reichl: Packaging and interconnection of sensors, Sens. Actuators A **25–27**, 63–71 (1991)
43.35 A. Grisel, C. Francis, E. Verney, G. Mondin: Packaging technologies for integrated electrochemical sensors, Sens. Actuators **17**, 285–295 (1989)
43.36 J.L. Lund, K.D. Wise: Chip-level encapsulation of implantable CMOS microelectrode arrays. In: *1994 Solid-State Sens. Actuators Workshop* (1989) pp. 29–32
43.37 T. Akin, B. Siaie, K. Najafi: Modular micromachined high-density connector for implantable biomedical systems. In: *Micro Electro Mech. Syst. Workshop* (1996) pp. 497–502
43.38 L. Muller, M.H. Hecht, L.M. Miller, H.K. Rockstad, J.C. Lyke: Packaging qualification for MEMS-based space systems. In: *Micro Electro Mech. Syst. Workshop* (1996) pp. 503–508
43.39 S.V. Groen, M. Rosmeulen, P. Jansen, K. Baert, L. Deferm: CMOS compatible wafer scale adhesive bonding for circuit transfer. In: *Int. Conf. Solid-State Sens. Actuators, Transducers'97* (1997) pp. 629–632
43.40 Y.T. Cheng, Y.T. Hsu, L. Lin, C.T. Nguyen, K. Najafi: Vacuum packaging using localized aluminum/silicon-to-glass bonding using localized aluminum/silicon-to-glass bonding. In: *14th IEEE Int. Conf. MEMS* (2001) pp. 18–21
43.41 M. Chiao, L. Lin: Vacuum packaging of microresonators by rapid thermal processing. In: *Proc. SPIE Smart Electron., MEMS, Nanotechnol* (2002) pp. 17–21
43.42 J.H. Partridge: *Glass-to-Metal Seals* (The Society of Glass Technology, Chapeltown 1949)
43.43 P. Kumar, V.A. Greenhut: *Metal-to-Ceramic Joining* (Minerals, Metals and Materials Society, Pittsburgh 1991)
43.44 M.G. Nicholas, D.A. Mortimer: Ceramic/metal joining for structural applications, Mater. Sci.

Technol. **1**, 657–665 (1985)
43.45 M. Esashi, S. Shoji, A. Nakano: Normally closed microvalve and micropump fabricated on a silicon wafer, Sens. Actuators **20**, 163–169 (1989)
43.46 M.E. Poplawski, R.W. Hower, R.B. Brown: A simple packaging process for chemical sensors. In: *Solid-State Sens. Actuators Workshop* (1994) pp. 25–28
43.47 S.F. Trautweiler, O. Paul, J. Stahl, H. Baltes: Anodically bonded silicon membranes for sealed and flush mounted microsensors. In: *Micro Electro Mech. Syst. Workshop* (1996) pp. 61–66
43.48 E.H. Klaassen, K. Petersen, J.M. Noworolski, J. Logan, N.I. Malfu, J. Brown, C. Storment, W. McCulley, G.T.A. Kovac: Silicon fusion bonding and deep reactive ion etching: A new technology for microstructures, Sens. Actuators A **52**, 132–139 (1996)
43.49 C.H. Hsu, M.A. Schmidt: Micromachined structures fabricated using a wafer-bonded sealed cavity process. In: *Solid State Sens. Actuators Workshop* (1994) pp. 151–155
43.50 A.L. Tiensuu, J.A. Schweitz, S. Johansson: In situ investigation of precise high strength micro assembly using Au-Si eutectic bonding. In: *Int. Conf. Solid-State Sens. Actuators, Eurosens. IX* (1995) pp. 236–239
43.51 A.P. Lee, D.R. Ciarlo, P.A. Krulevitch, S. Lehew, J. Trevino, M.A. Northrup: Practical microgripper by fine alignment, eutectic bonding and SMA actuation. In: *Int. Conf. Solid-State Sens. Actuators, Eurosens. IX* (1995) pp. 368–371
43.52 M.B. Cohn, Y. Liang, R. Howe, A.P. Pisano: Wafer to wafer transfer of microstructures for vacuum package. In: *Solid State Sens. Actuators Workshop* (1996) pp. 32–35
43.53 Q.-Y. Tong, U. Gosele: *Semiconductor Wafer Bonding, Science and Technology* (Wiley, New York 1999)
43.54 R.W. Bower, M.S. Ismail, B.E. Roberds: Low temperature Si_3N_4 direct bonding, Appl. Phys. Lett. **62**, 3485–3487 (1993)
43.55 H. Takagi, R. Maeda, T.R. Chung, T. Suga: Low temperature direct bonding of silicon and silicon dioxide by the surface activation method. In: *Int. Conf. Solid-State Sens. Actuators, Transducer 97*, Vol. 1 (1997) pp. 657–660
43.56 G. Wallis, D. Pomerantz: Filed assisted glass-metal sealing, J. Appl. Phys. **40**, 3946–3949 (1969)
43.57 L. Bowman, J. Meindl: The packaging of implantable integrated sensors, IEEE Trans. Biomed. Eng. **33**, 248–255 (1986)
43.58 M. Esashi: Encapsulated micro mechanical sensors, Microsyst. Technol. **1**, 2–9 (1994)
43.59 B. Ziaie, J. Von Arx, M. Dokmeci, K. Najafi: A hermetic glass-silicon micropackages with high-density on-Chip feedthroughs for sensors and actuators, J. Microelectromech. Syst. **5**, 166–179 (1996)
43.60 Y. Lee, K. Wise: A batch-fabricated silicon capacitive pressure transducer with low temperature sensitivity, IEEE Trans. Electron Devices **29**, 42–48 (1982)
43.61 S. Shoji, H. Kicuchi, H. Torigoe: Anodic bonding below 180 degree C for packaging and assembling of MEMS using lithium aluminosilicate-beta-quartz glass-ceramic. In: *Proc. 10th Annu. Int. Workshop Micro Electro Mech. Syst.* (1997) pp. 482–487
43.62 M. Esashi, N. Akira, S. Shoji, H. Hebiguchi: Low-temperature silicon-to-silicon anodic bonding with intermediate low melting point glass, Sens. Actuators A **23**, 931–934 (1990)
43.63 A. Hanneborg, M. Nese, H. Jakobsen, R. Holm: Silicon-to-thin film anodic bonding, J. Micromech. Microeng. **2**, 117–121 (1992)
43.64 A.V. Chavan, K.D. Wise: Batch-processed vacuum-sealed capacitive pressure sensors, ASME/IEEE J. Microelectromech. Syst. **10**(4), 580–588 (2001)
43.65 H. Henmi, S. Shoji, Y. Shoji, K. Yoshimi, M. Esashi: Vacuum packaging for microsensors by glass-silicon anodic bonding, Sens. Actuators A **43**(1–3), 243–248 (1994)
43.66 T. Hara, S. Kobayashi, K. Ohwada: A new fabrication method for low-pressure package with glass-silicon-glass structure and its stability. In: *10th Int. Conf. Solid-State Sens. Actuators, Transducers'99 Digest Tech. Papers*, Vol. 2 (1999) pp. 1316–1319
43.67 S.A. Audet, K.M. Edenfeld: Integrated sensor wafer-level packaging. In: *Int. Conf. Solid-State Sens. Actuators, Transducers'97* (1997) pp. 287–289
43.68 R.C. Benson, N. de Haas, P. Goodwin, T.E. Phillips: Epoxy adhesives in microelectronic hybrid applications, Johns Hopkins APL Tech. Dig. **13**, 400–406 (1992)
43.69 M. Shimbo, J. Yoshikawa: New silicon bonding method, J. Electrochem. Soc. **143**, 2371–2377 (1996)
43.70 P.M. Zavracky, B. Vu: Patterned eutectic bonding with Al/Ge thin film for MEMS, Proc. SPIE **2639**, 46–52 (1995)
43.71 J. Seeger, M. Lim, S. Nasiri: Development of high-performance, high-volume consumer MEMS gyroscopes. In: *Solid-State Sens., Actuators, Microsyst. Workshop* (2010) pp. 61–64
43.72 G. Humpston, D.M. Jacobson: *Principles of Soldering and Brazing* (ASM International, Materials Park 1993) pp. 241–244
43.73 A. Singh, D. Horsely, M.B. Cohn, R. Howe: Batch transfer of microstructures using flip-chip solder bump bonding. In: *Int. Conf. Solid State Sens. Actuators, Transducer 97*, Vol. 1 (1997) pp. 265–268
43.74 M.M. Maharbiz, M.B. Cohn, R.T. Howe, R. Horowitz, A.P. Pisano: Batch micropackaging by compression-bonded wafer-wafer transfer. In: *12th Int. Conf. MEMS* (1999) pp. 482–489
43.75 Y.T. Cheng, L. Lin, K. Najafi: Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging, IEEE/ASME J. Microelectromech. Syst. **9**, 3–8 (2000)
43.76 L. Lin: *Selective Encapsulations of MEMS: Micro Channels, Needles, Resonators, and Electromechanical Filters*, Ph.D. Thesis (Univ. California, Berkeley 1993)

43.77 Y.C. Su, L. Lin: Localized plastic bonding for micro assembly, packaging and liquid encapsulation. In: *Proc. IEEE Micro Electro Mech. Syst. Conf.* (2001) pp. 50–53

43.78 M. Schwartz: *Brazing* (Chapman Hall, London 1995)

43.79 K. Ikeda, H. Kuwayama, T. Kobayashi, T. Watanabe, T. Nishikawa, T. Oshida, K. Harada: Three dimensional micromachining of silicon pressure sensor integrating resonant strain gauge on diaphragm, Sens. Actuators A **21–23**, 1001–1010 (1990)

43.80 L. Lin, R.T. Howe, A.P. Pisano: Microelectromechanical filters for signal processing, IEEE/ASME J. Microelectromech. Syst. **7**, 286–294 (1998)

43.81 W.C. Tang, C.T.-C. Nguyen, R.T. Howe: Laterally driven polysilicon resonant microstructures, Sens. Actuators A **20**, 25–32 (1989)

43.82 G.T. Mulhern, D.S. Soane, R.T. Howe: Supercritical carbon dioxide drying of microstructures. In: *7th Int. Conf. Solid State Sens. Actuators* (1993) pp. 296–299

43.83 M. Judy: *Micromechanisms Using Sidewall Beams*, Ph.D. Thesis (EECS Department, Univ. California, Berkeley 1994) p. 162

43.84 R. Aigner, N.K.-G. Opperman, H. Kapels, S. Kolb: Cavity-micromachining technology: Zero-package solution for inertial sensors. In: *TRANSDUCERS '01. EUROSENSORS XV. 11th Int. Conf. Solid-State Sens. Actuators. Dig. Techn. Papers* (2001) pp. 186–189

43.85 M. Bartek, J.A. Foerster, R.F. Wolffenbuttel: Vacuum sealing of mcirocavities using metal evaporation, Sens. Actuators A **61**, 364–368 (1997)

43.86 J.J. Sniegowski, H. Guckle, R.T. Christenson: Performance characteristics of second generation polysilicon resonating beam force transducers. In: *IEEE Solid-State Sensor Actuator Workshop* (1990) pp. 9–12

43.87 C.H. Mastrangelo, R.S. Muller, S. Kumar: Microfabricated incandescent lamps, Appl. Optics **30**, 868–873 (1993)

43.88 J. Smith, S. Montague, J. Sniegowski, L.R. Manginel, P. McWhorter, R. Huber: Characterization of the embedded micromechanical device approach to the monolithic integration of MEMS with CMOS, Proc. SPIE **2879**, 306 (1996)

43.89 A. Partridge, M. Lutz, B. Kim, M. Hopcroft, R.N. Candler, T.W. Kenny, M. Esashi: MEMS resonators: Getting the packaging right. In: *Proc. 9th SEMI Microsyst./MEMS Seminar (SEMICON)* (2005) pp. 55–58

43.90 A.M. Leung, J. Jones, E. Czyzewska, J. Chen, B. Woods: Micromachined accelerometer based on convection heat transfer. In: *11th Int. Conf. MEMS* (1998) pp. 627–630

43.91 D.R. Spark, L. Jordan, J.H. Frazee: Flexible vacuum-packaging method for resonating micromachines, Sens. Actuators A **55**, 179–183 (1996)

43.92 M. Chiao, L. Lin: Hermetic wafer bonding based on rapid thermal processing, Sens. Actuators A **91**, 398–402 (2001)

43.93 Y.T. Cheng, L. Lin, K. Najafi: Fabrication and hermeticity testing of a glass-silicon packaging formed using localized aluminum/silicon-to-glass bonding, IEEE/ASME J. Microelectromech. Syst. **10**, 392–399 (2001)

43.94 M. Chiao, L. Lin: Accelerated hermeticity testing of a glass-silicon package formed by rapid thermal processing aluminum-to-silicon nitride bonding, Sens. Actuators A **97/98**, 405–409 (2002)

43.95 M.K. Keshavan, G.A. Sargent, H. Conrad: Statistical analysis of the hertzian fracture of pyrex glass using the Weibull distribution function, J. Mater. Sci. **15**, 839–844 (1980)

43.96 M. Chiao, L. Lin: A wafer-level vacuum packaging process by RTP aluminum-to-nitride bonding. In: *Tech. Dig. Solid-State Sens., Actuators Microsyst. Workshop* (2002) pp. 81–85

43.97 F. Rosebury: *Handbook of Electron Tube and Vacuum Techniques* (Addison-Wesley, Boston 1965)

43.98 Y.T. Cheng, W.T. Hsu, K. Najafi, C.T. Nguyen, L. Lin: Vacuum packaging technology using localized aluminum/silicon-to-glass bonding, J. Microelectromech. Syst. **11**(5), 556–565 (2002)

43.99 C. Luo, L. Lin: The application of nanosecond-pulsed laser welding technology in MEMS packaging with a shadow mask, Sens. Actuators A **97/98**, 398–404 (2002)

43.100 A. Cao, M. Chiao, L. Lin: Selective and localized wafer bonding using induction heating. In: *Tech. Dig. Solid-State Sens. Actuators Workshop* (2002) pp. 153–156

43.101 J.B. Kim, M. Chiao, L. Lin: Ultrasonic bonding of In/Au and Al/Al for hermetic sealing of MEMS packaging. In: *Proc. IEEE Micro Electro Mech. Syst. Conf.* (2002) pp. 415–418

43.102 Y.T. Cheng, L. Lin, K. Najafi: Localized bonding with PSG or indium solder as intermediate layer. In: *12th Int. Conf. MEMS* (1999) pp. 285–289

43.103 M. Esashi: Wafer level packaging of MEMS, J. Micromech. Microeng. **18**, 073001 (2008)

43.104 M. Lapisa, G. Stemme, F. Niklaus: Wafer-level heterogeneous integration for MOEMS, MEMS, and NEMS, IEEE J. Sel. Top. Quantum Electron. **17**, 629–644 (2011)

43.105 C. Liu: Recent developments in polymer MEMS, Adv. Mater. **19**, 3783–3790 (2007)

43.106 R.C. Ruby, A. Barfknecht, C. Han, Y. Desai, F. Geefay, G. Gan, M. Gat, T. Verhoeven: High-Q FBAR filters in a wafer-level chip-scale package. In: *IEEE Int. Solid-State Circuit Conf.* (2002) pp. 184–185

43.107 H. Shao, D. Kumar, S.A. Feld, K.L. Lear: Fabrication of a Fabry-Perot cavity in a microfluidic channel using thermocompressive gold bonding of glass substrates, J. Microelectromech. Syst. **14**, 756–762 (2005)

43.108 A. Decharat, M. Boers, F. Niklaus, G. Stemme: Novel room-temperature wafer-to-wafer attachment and sealing of cavities using cold metal welding. In: *Tech. Dig. MEMS* (2007) pp. 385–388

43.109 C.H. Yun, J.R. Martin, E.B. Tarvin, J.T. Winbigler: Al to Al wafer bonding for MEMS encapsulation and

3-D interconnect MEMS2008. In: *Tech. Dig.* (2008) pp. 810–813
43.110 K.-N. Chen, S.H. Lee, P.S. Andry, C.K. Tsang, A.W. Topol, Y.-M. Lin, J.-Q. Lu, A.M. Young, M. Ieong, W. Haensch: Structure, design and process control for Cu bonded interconnects in 3-D integrated circuits IEDM. In: *Tech. Dig.* (2006) pp. 367–370
43.111 L. Lin, K. McNair, R.T. Howe, A.P. Pisano: Vacuum encapsulated lateral microresonators. In: *7th Int. Conf. Solid State Sens. Actuators* (1993) pp. 270–273
43.112 H. Guckel: Surface micromachined pressure transducers, Sens. Actuators A **28**, 133–146 (1991)
43.113 C.H. Mastrangelo, R.S. Muller: Vacuum-sealed silicon micromachined incandescent light source. In: *IEEE IEDM* (1989) pp. 503–506
43.114 P. Monajemi, P.J. Joseph, P.A. Kohl, F. Ayazi: Wafer-level MEMS packaging via thermally released metal–organic membranes, J. Micromech. Microeng. **16**, 742–750 (2006)
43.115 K.S. Lebouitz, R.T. Howe, A. Pisano: Permeable polysilicon etch-access windows for microshell fabrication. In: *Dig. Tech. Papers, Transducers* (1995) pp. 224–227
43.116 R. He, C.J. Kim: On-wafer monolithic encapsulation by surface micromachining with porous polysilicon shell, J. Microelectromech. Syst. **16**, 462–472 (2007)
43.117 B.H. Stark, K. Najafi: A mold and transfer technique for lead-free fluxless soldering and application to MEMS packaging, J. Microelectromech. Syst. **15**, 849–857 (2006)
43.118 M. Esashi, N. Ura, Y. Matsumoto: Anodic bonding for integrated capacitive sensors. In: *Proc. Micro Electro Mech. Syst.'92* (1992) pp. 43–48
43.119 C.H. Yun, J.R. Martin, T. Chen, D. Davis: MEMS wafer-level packaging with conductive vias and wafer bonding. In: *Dig. Tech. Papers, Transducers'07* (2007) pp. 2091–2094
43.120 S. Nasiri: Wafer level packaging of MOEMS solves manufacturability challenges in optical cross connect. In: *IEEE Photonic Devices Syst. Packag. Symp* (2003)
43.121 P. Chen, C. Lin, C. Liu: Amorphous Si/Au wafer bonding, Appl. Phys. Lett. **90**(13), 132120 (2007)
43.122 P.R. Morrow, C.-M. Park, S. Ramanathan, M.J. Kobrinsky, M. Harmes: Three-dimensional wafer stacking via Cu–Cu bonding integrated with 65-nm strained-Si/low-k CMOS technology, IEEE Electron Device Lett. **27**(5), 335–337 (2006)
43.123 I.W. Jung, Y.-A. Peter, E. Carr, J.-S. Wang, O. Solgaard: Single-crystal-silicon continuous membrane deformable mirror array for adaptive optics in space-based telescopes, IEEE J. Sel. Top. Quantum. Electron. **13**(2), 162–167 (2007)
43.124 MEMS Process Review Report: *InvenSense IDG-300 Dual-Axis Angular Rate Gyroscope Sensor* (Chipworks, Ottawa 2007) MPR-0702-0801
43.125 S. Nasiri: New innovations in MEMS fabrications are responsible for meeting the demand for low-cost inertial sensors for consumer markets. In: *IMAPS Proc* (2007) pp. 407–412
43.126 F. Zimmer, M. Lapisa, T. Bakke, M. Bring, G. Stemme, F. Niklaus: One-megapixel monocrystalline-silicon micromirror array on CMOS driving electronics manufactured with very large-scale heterogeneous integration, J. Microelectromech. Syst. **20**(3), 564–572 (2011)
43.127 F. Zimmer, M.A. Lapisa, T. Bakke, M. Bring, G. Stemme, F. Niklaus: Very large scale heterogeneous system integration for 1-megapixel monocrystalline silicon micro-mirror array on CMOS driving electronics. In: *Proc. IEEE MEMS* (2011) doi:10.1109/MEMSYS.2011.5734530
43.128 F. Forsberg, N. Roxhed, P. Ericsson, S. Wissmar, F. Niklaus, G. Stemme: High-performance quantum-well silicon-germanium bolometers using IC-compatible integration for low-cost infrared imagers. In: *Proc. Transducers, Denver* (2009) pp. 2214–2217
43.129 M.A. Lapisa, F. Zimmer, A. Gehner, G. Stemme, F. Niklaus: Hidden-hinge micro-mirror arrays made by heterogeneous integration of two mono-crystalline silicon layers. In: *Proc. IEEE MEMS* (2011) doi:10.1109/MEMSYS.2011.5734520
43.130 A. Lapadatu, G. Kittilsland, A. Elfving, E. Hohler, T. Kvisterøy, T. Bakke, P. Ericsson: High-performance long wave infrared bolometer fabricated by wafer bonding, Proc. SPIE **7660**, 766016 (2010)
43.131 J. Voldman, M.L. Gray, M.A. Schmidt: Microfabrication in biology and medicine, Annu. Rev. Biomed. Eng. **1**, 401–425 (1999)
43.132 H. Becker, L.E. Locascio: Polymer microfluidic devices, Talanta **56**, 267–287 (2002)
43.133 Y.-C. Su, L. Lin: Localized bonding processes for assembly and packaging of polymeric MEMS, IEEE Trans. Adv. Packag. **28**, 635–642 (2005)
43.134 G. Thyrum, E. Cruse: Heat pipe simulation, a simplified technique for modeling heat pipe assisted heat sinks, Adv. Packag. **10**(12), 23–27 (2001)
43.135 G.P. Peterson, A.B. Duncan, M.H. Weichold: Experimental investigation of micro heat pipes fabricated in silicon wafers, J. Heat Trans. **115**, 751–756 (1993)
43.136 L. Jiang, M. Wong, Y. Zohar: Forced convection boiling in a microchannel heat sink, IEEE/ASME J. Microelectromech. Syst. **10**, 80–87 (2001)
43.137 F. Arias, S.R.J. Oliver, B. Xu, E. Holmlin, G.M. Whitesides: Fabrication of metallic exchangers using sacrificial polymer mandrils, IEEE/ASME J. Microelectromech. Syst. **10**, 107–112 (2001)
43.138 *ANSYS Modeling and Meshing Guide*, 3rd edn. (SAS IP, Canonsburg 2013)
43.139 D. Bystrom, L. Lin: Residual stress analysis of silicon-aluminum-glass bonding processes. In: *ASME 2002 Int. Mech. Eng. Congr. Expos., MEMS Symp.* (2002) pp. 45–50
43.140 H. Scholze: *Glass: Nature, Structure, and Properties*, 1st edn. (Springer, New York 1991)
43.141 R.R. Tummala: *Fundamentals of Microsystems Packaging* (McGraw-Hill, New York 2001)
43.142 S.J. Adamson: BGA, CSP, and flip chip, Adv. Packag. **11**(6), 21–24 (2002)

43.143 JEDEC Solid State Technol. Assoc.: *Accelerated Moisture Resistance-Unbiased Autoclave* (JEDEC, Arlington 2000), JESD22-A102C

43.144 M. Dokmeci, K. Najafi: A high-sensitivity polyimide capacitive relative humidity sensor for monitoring anodically bonded hermetic micropackages, IEEE/ASME J. Microelectromech. Syst. **10**(2), 197–204 (2001)

43.145 E.E. Lewis: *Introduction to Reliability Engineering*, 2nd edn. (Wiley, New York 1996)

43.146 W.D. Brown: *Advanced Electronic Packaging* (IEEE, Piscataway 1999)

43.147 F. Niklaus, P. Znoksson, E. Käluesten, G. Stemme: Void free full wafer adhesion bonding. In: *Proc. IEEE Micro Electro Mech. Syst. Conf.* (2000) pp. 241–252

43.148 C.H. Cheng, A.S. Ergun, B.T. Khuri-Yakub: Electrical through-wafer interconnects with sub-picofarad parasitic capacitance. In: *IEEE Electron. Compon. Technol. Conf.* (2002) pp. 18–21

43.149 D. Routkevitch, A.A. Tager, J. Haruyama, D. Almawlawi, M. Moskovits, J. Xu: Nonlithographic nano-array arrays: Fabrication, physics, and device applications, IEEE Trans. Electron Devices **43**, 1646–1658 (1996)

43.150 J. Gou, M. Lundstrom, S. Datta: Performance projections for ballistic carbon nanotube fieldeffect transistors, Appl. Phys. Lett. **80**, 3192–3194 (2002)

43.151 S.U.S. Choi, Z.G. Zhang, W. Yu, F.E. Lockwood, E.A. Grulke: Anomalous thermal conductivity enhancement in nanotube suspensions, Appl. Phys. Lett. **79**, 2252–2254 (2001)

43.152 K. Velikov, A. Moroz, A. Blaaderen: Photonic crystals of core-shell colloidal particles, Appl. Phys. Lett. **80**, 49–51 (2002)

Part J Nanotech

Part J Nanotechnology and Society and Education

44 Nanotechnologies in Societal Context
Barbara Herr Harthorn, Santa Barbara, USA

45 Environment, Health and Safety Issues in Nanotechnology
Rui Chen, Beijing, China
Chunying Chen, Beijing, China

46 Nanoscience and Nanotechnology Convergence
William S. Bainbridge, Arlington, USA

47 Global Perspectives of Nanotechnology Education
Kurt Winkelmann, Melbourne, USA
Bharat Bhushan, Columbus, USA

44. Nanotechnologies in Societal Context

Barbara Herr Harthorn

This chapter provides an overview of the past decade of research on the societal aspects and implications of nanotechnology in the USA. It starts by providing key terms and definitions and then outlines the contours of the social, ethical, governance, and participatory research in the USA, with key examples of nanoELSI work. The chapter argues that all these elements are different facets of responsible development and responsible innovation, and that the National Nanotechnology Initiative's investment in nanoELSI research, education and outreach has provided an unprecedented advance in scholarship and policy. The chapter proposes that nanoELSI has in some respects developed new forms of hybrid social science, ethics, historical, legal, sociological, psychological interdisciplinarity in addition to the interdisciplinary collaborations that form the basis of much nanoscale science and engineering innovation. Integration of the societal and the technical is an ongoing challenge, and the chapter cites some notable advances in this area as well.

44.1 **Assessing Technological Progress in Societal Terms** 1545
44.1.1 Approaches to nanoELSI/nanoELSA 1545
44.1.2 Responsible Development and Responsible Innovation 1547
44.2 **Nanotechnologies and Upstream Societal Engagement** 1547
44.3 **Ethics** 1549
44.4 **Governance, Law and Regulation** 1550
44.4.1 Standards 1551
44.4.2 EHS/OHS 1551
44.4.3 Voluntary Approaches/Soft Law 1551
44.4.4 Responsible Development/ Responsible Innovation 1552
44.5 **Public Perceptions and Participation in Decision-Making** 1552
44.6 **Integrating the Societal with the Technical** 1553
44.7 **Concluding Remarks** 1554
References 1554

This chapter provides an overview of research on the societal implications of nanotechnologies. It begins by defining key terms used by the societal research community and then discusses societal assessment of technological progress, nanotechnologies and *upstream* societal engagement, ethics, governance, law and regulation, public perceptions and participation in decision-making and the challenges and progress made in the past decade of work on integrating the societal with the technical. The US National Nanotechnology Initiative (NNI) has provided an unprecedented program in support of societal implications research, education and outreach and as such serves as the primary focus of this chapter, but comparative others from Europe, East Asia, and Latin America are included as appropriate.

This chapter takes up the critical issue of how to consider the crucial question of *what does the world need?* in terms of new technologies and innovation. Who decides, whose values matter? Determining that in a democratic society and responsible global innovation system requires systematic knowledge of that world, in all its social, economic, and political complexity – societal expertise, in addition to scientific/technical expertise – and it requires the ability to integrate the two. In this chapter I will address the current state of knowledge about these issues, using selective examples to illustrate.

The primary funding for societal research in the USA has been through cross-directorate support from the National Science Foundation (NSF), which received a small portion of the NNI budget from 2005–2015. In addition to dedicated funding of two societal implications Nanoscale Science and Engineering Centers (NSECs), the Centers for Nanotechnology in Society at the University of California at Santa Barbara (UCSB) and Arizona State University (ASU), and a number of other societal Nanoscale Interdisciplinary Research Team (NIRT) and Nanoscale Exploratory Research

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_44

(NER) awards, the NNI has also mandated Societal and Ethical Implications of Nanotechnology (SEIN) components in the other NSECs and the National Nanotechnology Infrastructure Network (NNIN). The NSECs and NNIN (now the National Nanotechnology Coordinated Infrastructure) societal programs are run under agency-directed cooperative agreements and can be seen as *program-initiated research*; the relevant NIRT and NER awards are *investigator-initiated*, but constitute a much smaller piece of the funding pie. Much of the work discussed herein derives from this critical support.

Many other chapters in this handbook address in detail the issues of definitions and standards for nanomaterials and nanoenabled technologies. US societal researchers have largely followed the NNI definitions of engineered materials at the nanoscale measuring between 1−100 nm [44.1]. For the purposes of societal research, there is a diversity of approaches with some researchers addressing *nanotechnology* in its generic form and a larger segment, especially in the second half of the *societal decade* from 2005–2015, that follows the UK Royal Society [44.2] which specifies the plural form as a necessary distinction due to the very large class of materials represented by the term and the variation in physical, chemical and optical properties within that class. In the societal research arena, these variations have also been critical dimensions across the material/product life cycle, and in the ethical, legal and societal issues they raise.

The *societal context* of new technologies is an idea that draws from the social sciences and science and technology studies (STS). In these fields, technologies are seen to be both produced by society (for example, by funding research, defining priorities, and training the scientific workforce) and in turn to have significant effects on and in society (for example, by changing whole systems of transportation, or construction materials, or how people communicate). Studying the societal context therefore means learning both how technologies are designed, innovated and produced, including historical, economic and political aspects of their creation, development and governance and their interactions with society, for example, the way new technologies might create new jobs or create technologies that replace jobs. The societal context is thus conceptualized as the encompassing framework for research and development (R&D) of new technologies and their consequences. It is a multiscalar idea – this context and its interactions can be studied at local (or community), regional (which in the USA might be state level or broader), national, and global levels. It operates across the life cycle of products from the upstream to the downstream, so it has temporal/spatial variance. And, it involves the individual and collective thoughts, values and behaviors of human beings, a notoriously tricky research subject. So technologies in society are also a highly dynamic system, changing over time in ways that are complex and difficult to track and model.

Given this understanding of the societal context, *societal implications* of nanotechnologies refers to all those potential interactions, near-, middle- and long-term, between the science and technology (S&T) enterprise and wider society, both in the innovation and production of new technologies and in their dissemination into society. In the NNI, this research has been implemented largely as a nanoELSI program, where ELSI refers to the ethical, legal and societal (or social) issues (or implications) of a new technology. This is modeled in concept after the main (and only) prior example in the USA, the ELSI program of the Human Genome Project implemented by the National Institutes of Health (NIH) in 1990 [44.3]. The NNI and NSF also refer to this program in parts of its documentation as societal and ethical implications of nanotechnology (SEIN), but it references the same focus on societal contexts. Until recently, the NNI has clearly differentiated nanoELSI from nanotechnology-related environmental health and safety (EHS) and nanoeducation programs in its guidelines and funding streams. A decade ago, the German philosopher, *Schummer* [44.4] explored the very diverse interpretations of the meaning of *societal and ethical implications of nanotechnology*, pointing to the *immature* state of nanoscience and the problems posed by its early stage of development for *ethical and sociological reflection* on its potential societal implications. This plurality of views is still evident today.

NanoELSI or SEIN is often differentiated from nanoELSA (ethical, legal and social aspects) as the latter term is used in Europe. Researchers there make a point of differentiating their ELSA focus on *aspects* versus the *implications* focus of US nanoELSI and human genome ELSI, because the US ELSI focus is argued to reflect an exclusive focus on *negative downstream* side-effects *at the expense of the broader set of upstream questions of the priorities and directions for research driven societal change* [44.5]. However, nanoELSI as implemented in the USA has arguably cast a very broad net in its scope and was also the earliest in the world to create a dedicated upstream societal program. The 10-year investment in the two NSECs, the Centers for Nanotechnology in Society at UCSB and ASU, represents the largest such investment in the world, and it has focused intently first on upstream and now midstream issues. However, the US societal research focus is now drawing to a close, just as societal issues related to nanotechnology reemerge as a concentrated focus in EU/EC programs [44.6].

As the scope of potential societal implications and effects of such a large class of new and emerging technologies, products, and systems is enormous, the expertise needed to address this research agenda is likewise very broad. The societal initiatives related to nanotechnology (ELSI and ELSA) in the USA and abroad have gathered together an unusually interdisciplinary group of scholars with a wide array of expertise in nanoenterprise/societal interface, including ethics/bioethics, cultural analysis, history of S&T, law and policy, regulation and governance of S&T, innovation systems and their global political economies, environmental studies, media and communication about new S&T, risk analysis and decision risk, public opinion, sustainability studies, social movements, public deliberation, democratic participation, to name only some of them. Nanoscale scientists and engineers, both practicing and in training, have also been key participants in the work. The two US societal centers, CNS-UCSB and CNS-ASU, as leaders of the NSF Network for Nanotechnology in Society, have also taken the lead, in conjunction with the NanoCenter at the University of South Carolina and its societal project, in founding in 2009 a new international professional society, the Society for the Study of Nanoscience and Emerging Technologies (the S.NET; recently renamed the Society for the Study of New and Emerging Technologies). The societal NSECs at UCSB and ASU have provided essential infrastructure and resources to this international network that NSF national Center funding has made possible. With broad international membership, the S.NET's participants reflect the diversity of approaches brought to this field; what follows can only provide a broad overview, with key examples.

44.1 Assessing Technological Progress in Societal Terms

The NNI was developed with four main goals:

1. To advance world-class nanotechnology R&D
2. To foster transfer of new technologies into products for commercial and public benefit
3. To develop and sustain educational resources, a skilled workforce and supporting infrastructure to advance nanotechnology
4. To support responsible development of nanotechnology [44.7].

The societal research within the NNI was designed to help meet these goals and has provided critical social, behavioral and economic analysis relevant to all four of them. This body of work uses a large and diverse methodological toolkit and focuses at scales from the local to the global and back again.

44.1.1 Approaches to nanoELSI/nanoELSA

Philosophical approaches. NanoELSI and SEIN have pursued a robust engagement with the philosophy of ethics and represents ... *the first major scientific endeavor that embodies a transformation in science with a major place for ethics* [44.8, p. 102]. Philosophers in collaboration with nanoscientists at the NanoCenter at University of South Carolina, for example, identified six main SEIN issues for nanotech development:

1. Environmental issues
2. Equity concerns
3. Legal, regulatory and insurance challenges
4. Privacy
5. Nanomedicine and medical ethics
6. *Hypertechnology*, their term for the rapid advance of these new technologies [44.8].

They also advocated for a full stakeholder engagement with *dialogue that is open, honest and attentive to interests and equity* [44.8, p. 101]. Other philosophers have added military applications, human enhancement and risk to the list of ethics concerns [44.9]. *Schummer* and *Baird* [44.10] explored the range of philosophical and ethical issues posed by nanotechnology, citing metaphysical issues with what *Nordmann* [44.11] called the *incredible tininess of nano*, and issues of truth and knowledge construction. *Lewenstein* [44.12] also explored the scope of nanoSEIN, as others in the volume reflected on issues of risk management, public discourse and anticipating unknown futures. Philosopher *Rosalyn Berne* also published that same year her book on *Nanotalk* [44.13], which reported on her conversations about ethics with Nanoscale Scientists and Engineers (NSE) that demonstrated both the unfamiliarity and importance of these issues. In addition, the NSEC at Northeastern University, for example, incorporated a societal implications component that included a partnership with an ethicist (*Sandler*) and integration of his work into the work of the center [44.14, 15].

Social Sciences Approaches

The core social sciences of anthropology, sociology and political science, along with communication, global studies and economics, have made signal contributions to nanoSEIN. The CNS at UCSB is built on such approaches that range from ethnographic studies of nanomaterials and toxicology labs [44.16], mul-

tilevel sociological and anthropological analysis of the global nanoenterprise [44.17, 18] and comparative global analysis of national investment programs in nanoR&D [44.19–21], to comparative nanoworkforce development analysis [44.22–24] and nanotech social movement analysis [44.25]. Political scientists and theorists have contributed essential theoretical and empirical understanding of collective action, democratic participation and governance [44.26–28]. For example, the team at the University of Wisconsin led by Scheufele and Brossard, along with ASU's Corley, has devoted itself to polling and political communication research about US public attitudes towards nanotechnology [44.29–31] that has culminated in two National Academy of Sciences (NAS) Sackler Colloquia on the Science of Science Communication (2012, 2013) [44.32] and an edited supplement in *Proceedings of the National Academy of Sciences* (PNAS) [44.33].

Science and Technology Studies

Science and technology studies (STS, also called Science, Technology and Society) is an interdisciplinary field focused on social, political and cultural analysis of scientific knowledge and technology in their societal contexts, both contemporary and historical. The CNS at UCSB has been notable in the SEIN for its focus on the *applied history* or *history in the making* of nanoscience and nanotechnology. For example, the work of these scholars has looked at how the history of the development of new instrumentation brought about changes in interdisciplinary collaboration among nanoscientists [44.34], at the sociological processes for developing and maintaining such new forms of interdisciplinarity at conferences and other meetings [44.35], and how more than 50 years of S&T development and citizen engagement have contributed to the current nanoenterprise [44.36]. STS also has a strong science policy focus as one of the critical points of connection between science and technology (S&T) and society. Science policy scholars at Georgia Tech have been particularly active contributors to nanoSEIN, notably *Shapira* and *Youtie* who, along with *Kay* (now at UCSB), have together mapped the global nanoinnovation system [44.37–40] for policy makers, and *Cozzens*, who has elevated developing world studies of equity issues [44.41]. ASU's combined CNS and Consortium for Science, Policy and Outcomes (CSPO) efforts have developed and staffed a science policy school in Washington, DC that is training a diverse rising generation of new technology policy scholars.

Risk Analysis and Perception

The NNI has demonstrated a keen interest in public risk perceptions about emerging nanotechnologies and particularly the potential for public backlash against nanotechnologies. Risk perception is an interdisciplinary field that combines approaches from social psychology, sociology and anthropology to systematically study the factors driving public perceptions and attitudes. Perceptions in this field are viewed as having both cognitive and affective components that vary based on knowledge, familiarity and other factors [44.42]. The processes that produce the social amplification of risk perception are of particular interest to policymakers and industry, as amplified risk perception can lead to stigmatization and backlash [44.43]. The CNS at UCSB has taken the lead on nanotech risk perception research, dedicating one of its three main streams of research (led by Harthorn, Pidgeon and Satterfield) to risk perception and social response in the USA and comparative UK publics. They organized an international specialist meeting on the topic in Santa Barbara in 2010, coedited a special issue of the journal *Risk Analysis* on nanotech risk perception in 2011 [44.44] and have also studied the risk views of scientists, toxicologists, regulators, and industry leaders in order to fully understand the current and projected future societal context for these emerging technologies [44.45–47].

Technology Assessment

Technology assessment (TA) is a Northern European-derived mode of studying and evaluating new technologies that serves as both an antecedent and parallel to many aspects of US SEIN research and projects. It has had well-developed programs across Northern Europe beginning in the 1960s. In particular, its form of Constructive Technology Assessment (CTA) developed in the Netherlands and Denmark in the mid-1980s [44.48] has served as the model for such US SEIN programs as CNS-ASU's real-time technology assessment (RTTA). One of CTA's key distinguishing elements from other European TA is the development of strategies and tools to contribute *feedback of TA activities into the actual construction of technology* [44.48, p. 252]. Indeed, Dutch chemist/philosopher and CTA author/promoter Arie Rip has played an active cross-national and intercontinental role in the development of US SEIN research, including at both CNSs and in the S.NET. Technological forecasting is a critical element of CTA, and this has been carried into the US SEIN through a number of pathways, most notably CNS-ASU's *anticipatory governance* framework [44.49] and the responsible innovation framework for governance [44.50].

Nanotech as Hybrid, new Path?

Just as NSE has necessitated and forged new Science and Engineering (S&E) interdisciplinarities [44.35], it

is arguable that SEIN has in some respects developed new forms of hybrid social science, ethics, historical, legal, sociological and psychological interdisciplinarity as well. The question of what is new about nanotechnologies is one researchers in this field direct at both our subjects of study and our methodological and reflexive practices in studying them [44.51].

44.1.2 Responsible Development and Responsible Innovation

As noted above, the fourth aim of the NNI is *to support responsible development of nanotechnology*. This may seem obvious – billions of dollars of public funds are being expended in the NNI, and new and potentially disruptive technologies have already or are predicted to result, and of course this would better be done responsibly than otherwise. However, in the nanoELSI research world, the specificities of this commitment are an important focus of both social and ethical inquiry; the mandate also means that the SEIN community shares in that responsibility and the problems of its implementation. Indeed, we are all engaged in a large societal experiment in advancing these goals – not surprisingly, the pathway forward here is not unified, nor even all proceeding in parallel. In simplest terms, responsible development involves a kind of cost-benefit analysis, where nanotech development has a normative/ethical obligation both to maximize the environmental, social, and economic benefits of the technologies developed, and to minimize their negative environmental, social and economic consequences. Societal research is essential to that assessment.

But of course different sectors of society often have very different values about what constitutes *benefit* and what risks or harms are acceptable to achieve those benefits. In the nanotech case, one of the readily agreed upon aspects of responsible development is the need for adequate EHS governance.

In the past half-decade, nanoELSI/ELSA researchers, primarily in Europe, have led the way to a focus on responsible innovation (RI). Although this approach shares many features with the responsible development aims of the NNI, its proponents have articulated a more integrated framework that both credits science and technological innovation with enormous transformative benefits, particularly in wealthy countries, but also lays responsibility with past S&T for unmet promises and unintended consequences. To create innovation that is good for society under conditions of uncertainty, unpredictability and ignorance, the RI framework stipulates 4 main features of innovation and development:

1. Responsive
2. Deliberative
3. Anticipatory
4. Reflective [44.50].

See below for more on this in relation to governance of nanotechnologies more specifically.

44.2 Nanotechnologies and Upstream Societal Engagement

Because the societal initiative within the NNI was implemented early in the technological development process, before many of the new materials and nanoenabled products and systems had been extensively scaled up in commercialization, this represented an unprecedented *upstream* societal effort. The mandate with upstream societal engagement is different from one in the midstream or downstream, when the R&D effort has matured, public awareness may be widespread, and effects on the ground may already be visible. Such engagement has been recurrently troubled by late-stage consultation, when the decisions have already been made and technologies are widely disseminated. If risk issues emerge at that stage, past efforts indicate that public dialogue is likely to prove frustrating at best, for example, in the GM Nation project in the UK [44.52]. The nanotechnology *upstream* conditions from over a decade ago created an opportunity, at least in theory, for a different kind of dialogue in society about technological development processes and decisions, one in which public views could in theory inform government, industry and S&E action [44.53].

Such public engagement and participation efforts are often seen differently among the different nanoR&D stakeholders. Government and industry often pursue public engagement in a potentially fraught effort to gain public acceptance. This is problematic in several ways:

1. Across studies of different technologies, public engagement has been shown to have highly variable effects on acceptability [44.54].
2. What constitutes acceptability is more complicated than it seems [44.55].
3. In order to be successful, public engagement and participation needs to include two-way communication, not just top-down S&E experts instructing the lay public in how to think [44.56].

There are three widely agreed upon reasons to pursue public participation and engagement in S&T decision making:

1. *Normative*: because it's the right or ethical thing to do, for example because the public have a right to know and to participate in decisions that may directly affect them
2. *Instrumental*: because it is often useful to do, for example to gain public buy-in on controversial decisions that affect them
3. *Substantive*: it can often provide important knowledge and perspectives that can produce better policy or decisions [44.56, 57].

However, most prior engagement has taken place further downstream in the S&T production life cycle, often after risk controversies have arisen. So, how to effectively engage diverse publics early in development, in an ethical and two-way mode, has occupied a great deal of the nanoELSI communities' thought and effort. First, in the upstream context, it is particularly important to see the results of engagement as provisional or contingent. Thus, for example, widespread benefit-centric public views on nanotechnology [44.46] should not be taken to mean publics share an unconditional approval of nanotechnologies. Rather, these upstream views in the nanotech case have been associated with very high levels of uncertainty and unwillingness to provide acceptance/rejection decisions, which scholars take to be a measure of potential malleability [44.47]. And both experimental survey and deliberative research have indeed shown that people's initial uncertainty about nanotechnologies can move toward either greater acceptance or greater risk concern or it can remain ambivalent following deliberative dialogue or experimental information provision in surveys [44.58, 59]. In order to fulfill the ethical, substantive and instrumental goals, *public engagement should be done early in the tech development process*, as we have seen in the nanotechnology case in the USA [44.45, 60], the UK [44.45, 61] and elsewhere in Europe and South America [44.61, 62], *and it should be done often*, which is far rarer [44.63, 64].

Who are the publics to be engaged? As indicated above, this is not a given. For democratic participation, *the public* is necessarily plural and obviously needs to be diverse enough to represent a range of social positions and locations. For a large multicultural nation like the USA, this is very difficult, although multisited efforts [44.60] and diversely drawn groups [44.45, 63] help. Cross-national comparisons have also contributed new understandings, particularly regarding the deeper social risks attributed to some new nanotechnologies, and shared by many people around the globe [44.45, 61]. In most of these efforts the publics invited to participate are selective – for research reasons they are not run as open community fora as self-selection may bring those more concerned or more knowledgeable than a representative sample of the full population [44.65]. Social science survey research, by contrast, often aims to reach truly representative samples of a total population. These people are recruited randomly or selectively by survey research firms for their demographic and locational attributes.

Have nanotech deliberative engagements been successful? A systematic review by *Corner* and *Pidgeon* [44.66] of 18 nanotech deliberation projects in North America and Europe found that a number of them were effective in producing deliberative, informed judgments, even among participants who walked in the door with little to no knowledge, and that benefit-centric views of nanotechnologies were widespread among participants, but there were also deep-seated reservations about government, corporations and responsibility [44.47].

Nongovernmental organizations (NGOs) or civil society organizations (CSOs) represent another kind of publics, sometimes referred to in STS scholarship as the *uninvited public* because their tactics and protest aims (needlessly) intimidate S&E, government and industry [44.67]. *Welsh* and *Wynne* have argued [44.68] that such policy maker anxiety about *unruly* publics is misplaced and that engagement is often constructive and beneficial to desired outcomes. In the nanotech case, NGOs have also often been reluctant to serve as token representatives of *the public* and fear becoming coopted by government and/or industry in multistakeholder meetings. In theory, NGOs are the main publics to be consulted downstream [44.69], but in the nanotech case, NGOs have also played significant roles, across nations, and studying them is an important part of the societal agenda [44.25]. They have a clear role to play in advocating for responsible development and innovation. One recent book-length study of the independent organization the Project on Emerging Nanotechnologies (PEN) provides a model demonstration of this argument [44.70].

Science museum audiences constitute a ready-made science-interested public for engagement. In 2005, along with the two societal NSECs at UCSB and ASU, the NNI SEIN generously funded the formation of a Nano Informal Science Education Network (NISEnet), a large national network of science museums, with the Boston Museum of Science as lead [44.71]. This multisited organization has engaged and educated a network of educators and the publics they serve. Although arguably designed in principle around a *deficit model* of public knowledge, the

NISEnet in practice has reached beyond one-way science education to engagement, incorporating societal aspects into their materials and programs directly and indirectly, demonstrating the synergies that can exist between nanoELSI and nanoISE.

Engagement and public participation have been integral to the organizations of both CNS-UCSB and CNS-ASU, in part in response to funder-driven criteria. Both centers have used large and creative toolkits to engage diverse publics via nanoscience cafes, series of deliberative workshops, Nano Days informal science education, citizen advisory boards, NGO conferences and meetings, multistakeholder workshops, web-based courses, blogs, Twitter, Facebook and myriad other activities. Other notable nanotech public engagement efforts have included the Citizens' School at University of South Carolina [44.72] and deliberative exercises at the University of Wisconsin [44.73]. Engagement assumes a relatively deep connection and two-way dialogue; public outreach activities, which often do not meet these engagement criteria, of course, are far more numerous throughout the NNI.

In spite of these many efforts and solid methodological advances, there have been impediments to fully realized public deliberation and engagement in the USA. The most serious of these is the lack of direct channels of communication to policymakers, an essential element of democratic participation, and a relatively common practice in EU citizen engagement. Center level funding for the societal work has enabled more information flow between engagement work and policy makers, but it is not institutionalized and hence very labor intensive over time and fragile as a process. In addition, scaling up S&T deliberative engagement to the national level would require substantial investment not yet in evidence for nanotechnologies, nor in the plans for converging technologies societal implications work.

44.3 Ethics

Like its predecessor ELSI program in the Human Genome Program [44.3], the nanoELSI or SEIN program in the USA has ethics and bioethics as a central part of its agenda. In addition to the domains of nanotechnology applications and effects of enduring interest to ethicists cited above, ethicists have had unusual opportunities in the context of nanoR&D in the USA and Europe to engage NSE but also social scientists, policy makers, and others about the ethical principles for technology assessment.

The German philosopher *Schummer* has argued compellingly that international reflections about nanoethics are necessarily complicated by the diversity of sociocultural issues (such as language, cultural heritage, economics, and politics) that affect the way people perceive the ethical issues in nanotechnology [44.74]. That is, he has raised the issue of ethical relativism in *the globalization of the ethical debate around nanotechnology* [44.74, p. 217]. In this, he is asking how individuals' or societies' moral order or intuition may be in conflict with given nanotechnologies. He clearly differentiates this from the idea of risks (although risk perceptions do incorporate moral judgments). Like many others in the nanoELSI arena, he uses *technology* to reference not just actual products but also the full range of R&D activities and their governing mechanisms.

Schummer points to the embedded normative (ethical) concepts built into goals, such as those articulated above as the goals of the NNI. Because such goals are prescriptive – they state what kind of technology should be developed – they encompass a *hidden normative debate about norms and values that are frequently expressed in the form of hopes and fears* [44.74, p. 220]. The fact that such normative debates should be explicit and deliberately in the public discourse is itself an ethical issue because [44.74, p. 220]

> *it smuggles in values in the disguise of definitions or forecasts of allegedly deterministic technological developments, which are kept apart from normative debates.*

Although these values can and do vary within societies at the level of individuals, they also reflect the patterns different countries tend to use. As one example of how cultural context may shape ethics and values in relation to nanotechnologies, Schummer points to German sensitivity to the idea of *human enhancement*, which in the USA has been a central goal for nanobiotechnology and also a vested interest of the military. In Germany, because of their Nazi legacy, there is strong abhorrence of any approach that could be used for eugenic purposes. Other concepts with strong cultural-specificity in normative ideas are those about human identity and integrity, privacy and ideas about the private sphere, national prestige objects, and underlying concepts of what constitutes *a good life*. Economic imperatives are also deeply implicated in ethical issues, but with significant differences across developing and developed countries because developing countries have a unique opportunity to catch up economically through entering new tech markets that may lead them to neglect ethi-

cal issues. However, patent policies in place since 1980 and changes in raw material dependency are likely to increase economic gaps more than previous technologies [44.74].

Thus, the list of potential ethical concerns that can or should be brought to bear on the case of nanotechnology development is lengthy [44.75]. The main ones that have been the focus of nanosocietal research and discussion in the past decade include the following:

- Distributive justice or fairness regarding the benefits and harms: Does/will the technology improve the lives of some but not others? Does/will the technology production harm some but not others? Are the ones who are harmed different from the ones who benefit [44.44, 45, 76]?
- Procedural justice: Are people informed about the technological effects? And who gets to make the decisions? Do people share in the decision making? Are there procedures in place to ensure a fair outcome [44.76, 77]?
- The precautionary principle(s) (PP): *Weckert* and *Moor* [44.78] have argued that in spite of its definitional issues (e.g., over the threshold of anticipated harm required to invoke it) the precautionary principle is valid and useful in the case of nanotechnologies. Ethicists have noted its particular applicability in environmental effects [44.79]. The EU has employed active precaution in its approach to nanomaterials regulation (e.g., through the REACH program (Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals), implemented by the European Union in 2007), putting the burden on industry to do safety testing before taking nanotechnologies to market; in the USA the National Institute of Occupational Safety and Health (NIOSH) adopted a *better safe than sorry* approach to nanomaterials industry guidance and precautionary information dissemination to proactively reduce potential worker exposures to nanomaterials like carbon nanotubes ahead of full hazard characterization [44.80]. And in the USA and Europe, NGOs have frequently been the stakeholders advocating most vocally for precautionary approaches, often including moratoria on development [44.81, 82].
- Involuntary exposure through lack of informed consent to risks and hazards, including such broader social and economic risks as invasion of privacy through use of nanosensors, accelerating inequality, loss of identity, loss of humanness, and loss of livelihood, in addition to exposures to toxins.
- Whether a technology is needed is also a concern in risk perception research and deliberations as well as larger responsible development/innovation contexts.

As can be seen from above, equity issues top the list and are the first-order concerns. These include access to beneficial technologies or economic gain as well as safety from potential harms. They also include issues of emerging technologies and global inequalities [44.83]. It is also clear that ethics issues are inextricably intertwined with both governance and risk, which will be discussed below. The ethical challenges in the past decade have in many cases been about anticipated rather than present issues and finding new methods to assess anticipated concerns has been on many researchers' minds [44.84].

44.4 Governance, Law and Regulation

Technological governance is a central ethical issue in all technological R&D, and nanoELSI researchers have been concerned with potentially ungovernable or unanticipated aspects of nanotech innovation through the past decade. The term *governance* refers to a much broader system of responsibility and political control of technologies than that of formal governmental regulation. Governance includes governmental R&D programs, subsidized industries, and restrictive or incentivized regulation. It also includes decisions on technology control made by corporations. Such decisions that are made without public information, or with secrecy, are likely to raise ethical issues, suspicion and mistrust [44.4]. Nanotechnology governance thus includes legal and regulatory aspects but refers as well to this wider scope of responsibility and power.

In his recent overview of nanotechnology governance since 2003, *Bosso* argues that

> *the incentives driving innovation and commercialization* [...] *always outpace incentives for more anticipatory and systematic oversight of a technology's potentially adverse environmental and health impacts.*

In addition, nanotechnology poses particular challenges for risk governance because of its framings as revolutionary and unfamiliar [44.69, p. 163].

One early (2006) anticipatory approach to governance issues of nanotechnology was undertaken by the International Risk Governance Council (IRGC) in Switzerland, in consultation with a number of lead-

ing international figures, and led by German sociologist *Renn* and US NNI leader *Roco* [44.85, 86]. In their IRGC report, they proposed a risk governance framework that divided two different risk governance concerns across present or near-term technologies, and longer-term technologies. They aimed to identify risk governance deficits (technical and organizational, institutional and risk communication) and make recommendations to address them. The Royal Society's Report on Nanotechnology in 2004 [44.2] similarly initiated a robust assessment project, including focused attention to societal context, and anticipated a range of potential problems.

Both of these approaches reflected strong upstream efforts to critically examine future directions. Both included the societal context for nanotechnologies as critical to governance, and both made a number of recommendations for successful governance and risk communication. However, in the intervening decade, neither has been carried forward in a cohesive or programmatic way by governments, industries or nanoELSI researchers. Instead, a more fragmented approach has emerged, without the critical coherence and power to effect needed change in the absence of crisis [44.69].

The following are the main examples of nanotechnology governance, regulation and legal change issues in the past decade.

44.4.1 Standards

Nanotechnology is not a technology but rather a very broad class of technologies defined through their scale, many with new and hitherto unknown characterization, and as such they have posed major challenges to classification and standardization. Although physical, technical classification is a technical rather than societal issue being addressed by the International Standards Organization (ISO), the Organization for Economic and Co-operation and Development (OECD) and the US National Institute of Standards and Technology (NIST), the lack of agreed upon standards, and the difficulty of deriving them, have produced many spillover effects on the societal side. For example, since with few exceptions in the USA engineered nanomaterials (ENMs) continue undifferentiated from their bulk forms from a regulatory standpoint, they are also untrackable from a governance standpoint. Efforts such as the California Department of Toxic Substance Control (DTSC) call-ins with legally required ENM reporting in 2010 (carbon nanotubes) and 2011 (quantum dots and nanometals/metal oxides) proved remarkably unsatisfactory for identifying or monitoring nanomanufacturers [44.87]. Thus firms that are producing or handling ENMs continue to fly largely under the radar, limiting governmental ability to ensure workplace, community or consumer safety. In a recent review, *Schulte* et al. at NIOSH outlined 15 years of occupational health and safety (OSH) of nanotechnology and noted that in spite of progress, *unified analytical standards and exposure characterization methods are still lacking* [44.88, p. 1] and that protection across the life cycle is still lacking.

44.4.2 EHS/OHS

Regulating and governing risks to EHS and OHS have been at the top of almost everyone's list as the leading regulatory concern for many aspects of the nanoenterprise. *Beaudrie* et al. [44.89] applied a gap analysis to regulatory oversight of nanomaterials across the life cycle and found many problematic areas where existing regulatory controls were simply not in place. They then went on to demonstrate in an expert survey that among diverse experts, nanoregulators had significantly lower confidence in the adequacy of the current regulatory apparatus to handle ENMs than did physical scientists or toxicologists [44.90].

Expert surveys have provided extensive understanding of how nanoscientists and nanotoxicologists view the hazards of ENMs, and how they handle such materials. Industry surveys have also provided insight into the ENM workplace and its safe handling and product stewardship practices, as well as the ENM industry's strong preference for autonomy from regulation [44.91, 92]. Thus, experts are a highly diverse group in their views on the risks of nanomaterials and on regulatory capacity.

44.4.3 Voluntary Approaches/Soft Law

Nanotechnology has witnessed several attempts at voluntary regulation, the first via the joint Environmental Defense Fund/DuPont NanoRisk Framework sponsored by Rice University's International Council on Nanotechnology (ICON) (2005) – it failed to generate sufficient active participation from the nanoindustry, as did the EPA's voluntary Nanomaterials Stewardship Program (NMSP). Even California's mandatory call in failed to generate significant response, and in the absence of regulation, DTSC personnel had few means to identify companies and businesses handling ENMs. *Bosso* [44.69] shows that as recognition of the shortfalls of current regulatory approaches has coalesced, there has been increasing advocacy for radical change to the current US risk management regime such as EPA's Toxic Substances Control Act (TSCA). Yet more recently most calls in the legal field are for address-

ing deficiencies in governance *by calls for more flexible combinations of* soft *and* hard *law approaches* [44.69, p. 163], including voluntary industry standards and codes of behavior, voluntary information sharing and other flexible and self-regulated approaches. However, such approaches have major acknowledged limitations. Bosso's analysis suggests that *where* in the upstream-midstream-downstream innovation stream technological governance occurs should have a major effect on what kind of approach is taken, with the upstream permitting public engagement and dialogue, and the midstream allowing soft law approaches with coordinating committees bridging industry and society, but that the downstream requires hard law and regulatory teeth, relatively less role for the public and elevated roles for NGOs and lawyers [44.69]. As one example of the uphill battle involved in formal regulatory change, he traces the ups and downs of attempts to reform and update TSCA, first in 2009–2010, then again in 2015 (finally signed into law on 22 June 2016).

There are numerous other legal and regulatory issues involved in nanotech R&D, too many to address in this brief overview. The innovation system has changed significantly and patent thickets and other problems abound. Privacy issues continue to mount in the USA and globally, in spite of legal protections in the USA under Health Insurance Portability and Accountability Act (HIPAA) and Federal Deposit Insurance Corporation (FDIC) that create national standards and regulate consumer financial information. Consumer safety and protection are regulated under the Consumer Product Safety Commission (CPSC) and agricultural uses are regulated by the US Department of Agriculture, while food and pharmacological products are regulated by the Food and Drug Administration. The complexity and division of duties and boundaries among all these US players creates a disjointed process for regulatory scrutiny of ENMs and their enabled products and inevitable gaps in coverage [44.89] as well as communication issues.

44.4.4 Responsible Development/ Responsible Innovation

As noted above, the responsible innovation framework defines responsibility in legal terms, but goes beyond legislated safety (e.g., general liability regimes and sector-specific regulation), which regimes are notably after-the-fact regulation of the products of innovation [44.93]. Such legal tools have been critical to safeguard health.

However, RI advocates for a prospective kind of responsibility, one that is *anticipatory* and can move alongside innovation's creative imagination and invention. It needs to be *reflective* about the purposes of innovation, the values underlying innovation. It also needs to be *deliberative* about engaging publics and others as a part of democratic participation, and *responsive* in being open to new information. *Owen* et al. define it as follows [44.93, p. 36]:

> *Responsible innovation is a collective commitment of care for the future through responsive stewardship of science and innovation in the present.*

The RI framework itself provides a normative statement of what innovation in society should be and do, and it has garnered interest and support. It has generated a growing literature that is a mix of theoretical treatises on the ethical and legal aspects of the framework and empirical case analyses of RI in action [44.94–96]. It has also spawned a new journal, *Responsible Innovation*, founded and edited until recently by ASU's Guston, and now by ASU's Fisher. The decade ahead will serve as a critical test of its ability to invoke the kind of care and responsibility for societal aspects of the global S&T innovation system.

44.5 Public Perceptions and Participation in Decision-Making

From the above, it should be clear that public attitudes, public risk perceptions and beliefs and values are of critical importance to the nanoELSI program. Both the nanoELSI centers and a substantial number of other leading societal researchers have contributed to a now substantial literature on this. These studies have used in-depth, qualitative research based on interviews and public deliberations; they have also conducted numerous public attitude, risk perception, and experimental surveys of US and other publics. Media studies based on both traditional print and new media have been important to understand emerging public attitude formation. NGOs have also been studied for their efforts to influence public views, journalists and policy makers. Risk has been a critical dimension of this work, but values, beliefs, and identities also play primary roles in public perceptions [44.44, 47].

As indicated above, there are strong ethical, instrumental and substantive principles for public participation in advanced democratic societies. The analytic-deliberative model for understanding risk explicitly invoked the need for both analytic assessments and

deliberative engagement in making risk decisions and became the normative model for understanding risk in the USA [44.97]. New forms of participation are also a part of the mix, exemplified by the United Kingdom Energy Research Centre (UKERC) project on energy technologies and public engagement [44.98].

Full democratic public participation in technological R&D in a large multicultural nation like the USA, not to mention the global community, is undoubtedly a significant challenge. The nanoELSI and responsible innovation work has created successful models that can now be deployed in the new and converging technology in society work ahead. However, public participation in the nanotechnology innovation system has just completed its upstream phase. Midstream engagement and participation should be of critical importance to scholars, scientists and engineers, industry and policy makers.

44.6 Integrating the Societal with the Technical

As indicated above, the generous funding of SEIN research in the NNI has resulted in a large body of scholarly research on ethical, legal, economic, environmental and social research on aspects of nanotechnology. We SEIN researchers have been challenged not to just research societal aspects of nanotechnologies but to get out of our familiar sandboxes to integrate those efforts within the NSE enterprise. The work has been shared widely, with major efforts to disseminate and publish in S&E publications, to disseminate to policy makers and to present to and engage in dialogue with industry (for a set of examples, see CNS-UCSB presentations 2005–2016 [44.99]). A number of collaborative research projects and programs have emerged in this context. Some of the main things we have learned in this process are discussed here.

The education of both society and the S&E enterprise is necessarily a mutual process and SEIN researchers need to reach out, forge partnerships, be willing to travel and to step outside the comfort zone of the social sciences to work in S&E contexts that may not start with a lot of space, intellectual or symbolic, for the societal aspects. This has also on occasion required a new kind of three-way communication between science–social science–society, where the societal researcher needs to serve as an essential broker and translator. The *trading zone* metaphor has been used to describe a creolized form of language and interaction to describe how scientists with different disciplinary backgrounds can meet and engage, particularly with the help of those with interactional expertise in more than one area [44.100], and this has been advocated within the NNI [44.101]. However, models based on scientists converging in the shared use of instrumentation, for example, may not be the most apt for societal science/nanoscience interface, where values may diverge more strongly and the impetus to *trade* may not be self-evident to all participants. More subtle and ethnographic approaches to traversing cultural divides, particularly where asymmetrical power relations apply, may well be called for. Below are several examples of successful approaches to this challenge of integration between nanoscience and society.

NanoELSI researchers with both societal and materials training and expertise have been important full partners in the SEIN context. For example, CNS-UCSB has had a political scientist and STS scholar who is also an electrical engineer (Bimber) on its leadership team from the start, and the history of science research group in the center is led by a materials-trained historian (McCray). The risk perception group includes a faculty member (Kandlikar) who is both a materials and science policy trained researcher, who has trained students into both approaches (e.g., Beaudrie). CNS-ASU is similarly diverse with a number of faculty and researchers in past or present leadership positions with S&E backgrounds and science policy training and interests (e.g., Sarewitz, Miller and Bennett)

Collaborations of societal researchers with S&E have proliferated in the USA under NSF SEIN funding. In CNS-UCSB, founding co-PI Hu was a vital part of proposal development and early center organization. The CNS-UCSB Executive Committee has always included one or more leading NSE researchers (Hawker, Hu, Holden). Research and education collaborations have abounded, the most notable between Harthorn and her group in the CNS, and the UC CEIN, where she has led a societal research group, served as an executive committee member, and coauthored publications. Collaborative research projects have also resulted from this.

Another important form of interaction has been expert workshops as a form of both research and engagement. This includes the structured decision making (SDM) work at the University of British Columbia on codeveloping expert decision risk tools in the case of EHS risk uncertainty [44.102], Delphi experiments with nanoscientists by nanoresearcher *Berube* et al. [44.103], expert meetings at the Project on Emerging Nanotechnologies [44.70], and others.

Both CNSs have leveraged center-level funding and infrastructure to implement highly successful programs to increase the societal understanding of the next generation of NSE-in-training and the NSE understanding of the next generation of societal researchers-in-training [44.104, 105]. At UCSB this has taken place primarily by the integration of NSE graduate fellows into the social science and science policy work of the CNS [44.104]; at ASU this has been primarily through the embedding of social scientists in the bench science lab settings [44.105]. Both approaches have used new modes to develop new capacities that can be models for programs elsewhere. With funder-driven incentives, NNI NSECs have also successfully integrated societal information into NSE courses and NSE content into social science courses over the past decade, as well as jointly providing public engagement activities.

44.7 Concluding Remarks

Because of its proactive, anticipatory societal work as well as its complexity as a class of technologies, nanotechnology is in many respects a new hybrid form of science in society, much though our history of science colleagues point to its deep historical antecedents. The NNI has led the world in funding a robust and diverse societal implications research, education, and engagement enterprise. The latter has been highly productive in terms of both scholarly merit and broader impacts of its work. Even more impressive are the many intersecting lines of connection forged between the societal research community and the nanoscience R&D community.

This science in society *testbed* has developed the intellectual infrastructure, broader partnerships, and deeper ties *across the aisle* with NSE for future societal work even as the institutional and programmatic funding for it is diminishing. The work of the two CNSs at UCSB and ASU has demonstrated again and again the importance of center-level funding for ramping up this new and important national nanoELSI project.

What the future holds is not clear for the next decade of societal work. We have learned a great deal about the upstream as a research site. We've created two national clearinghouses to share information, approaches, toolkits and data, and we have developed and specifically tailored a methodological toolkit for research in the upstream of new technology development and societal emergence. We have developed an international network of scholars that is already expanding its footprint around the globe, in the Global North and the Global South, and this community stands ready to tackle the rising issues ahead with converging technologies, and the unanticipated consequences that will doubtless accompany their development and deployment in society.

This is also the moment when we can see the path forward to socially sustainable technology development with unusual clarity. Responsible innovation frameworks provide a normative guideline, but practices on the ground will inevitably require new collaborative approaches and effort and inspiration from societal and technological experts. As we expand the scope of societal research and the responsible innovation framework to other emerging technologies, it is likely the relevance and impacts of our work on nanotechnologies in society will become more and more evident.

References

44.1 National Nanotechnology Coordinating Office: National Nanotechnology Initiative Strategic Plan (NNCO, Washington DC 2014), available online a: http://www.nano.gov/about-nni/what/vision-goals http://www.nano.gov/nanotech-101/what/definition

44.2 Royal Society & the Royal Academy of Engineering: *Nanoscience and Nanotechnologies: Opportunities and Uncertainties* (Royal Society, London 2004) Available for download at: https://royalsociety.org/~/media/Royal_Society_Content/policy/publications/2004/9693.pdf

44.3 J. McEwen, J. Boyer, K. Sun, K. Rothenberg, N. Lockhart, M. Guyer: The ethical, legal, and social implications program of the national human genome research institute: Reflections on an ongoing experiment, Annu. Rev. Genom. Hum. Genet. **15**, 481–505 (2014)

44.4 J. Schummer: Societal and ethical implications of nanotechnology: Meanings, interest groups, and social dynamics. In: *Nanotechnology Challenges: Implications for Philosophy, Ethics and Society*, ed. by J. Schummer, D. Baird (World Scientific Publishing, Hackensack 2006) pp. 413–449

44. ELSA|Norway: Network of ELSA-researchers in Norway, available online at https://www.ntnu.edu/elsa/what-is-elsa-research, last accessed August 17, 2016

44.6 European Commission: *HORIZON 2020 – Work Programme 2016 – 2017: Nanotechnologies, Advanced Materials, Biotechnology and Advanced Manufacturing and Processing* (European Commission, Brussels 2016), Available for download at: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/nanotechnologies

44.7 National Nanotechnology Coordinating Office: *Ethical-Legal Issues*, (NNCO, Washington DC 2016) Available for download at: http://www.nano.gov/you/ethical-legal-issues

44.8 D. Baird, T. Vogt: Societal and ethical interactions with nanotechology ("SEIN") – An introduction, Nanotech. Law Bus. J. **1**(4), 101–107 (2004)

44.9 F. Alloff, P. Lin, D. Moore: *What is Nanotechnology and Why Does it Matter?: From Science to Ethics* (Wiley-Blackwell, West Sussex 2010)

44.10 J. Schummer, D. Baird (Eds.): *Nanotechnology Challenges: Implications for Philosophy, Ethics and Society* (World Scientific Publishing, Hackensack 2006)

44.11 A. Nordmann: Noumenal technology: Reflections on the incredible tininess of nano. In: *Nanotechnology Challenges: Implications for Philosophy, Ethics and Society*, ed. by J. Schummer, D. Baird (World Scientific Publishing, Hackensack 2006) pp. 49–72

44.12 B. Lewenstein: What counts as a 'social and ethical issue' in nanotechnology? In: *Nanotechnology Challenges: Implications for Philosophy, Ethics and Society*, ed. by J. Schummer, D. Baird (World Scientific Publishing, Hackensack 2006) pp. 201–216

44.13 R. Berne: *Nanotalk: Conversations with Scientists and Engineers about Ethics, Meaning and Belief in the Development of Nanotechnology* (Lawrence Erlbaum, Mahwah 2006)

44.14 R. Sandler: *Nanotechnology: The Social and Ethical Issues Project on Emerging Nanotechnologies Report* (Woodrow Wilson Center, Washington DC 2009)

44.15 E. Michelson, R. Sandler, D. Rejeski: Nanotechnology. In: *From Birth to Death and Bench to Clinic: The Hastings Center Bioethicsm Briefing Book for Journalists, Policymakers, and Campaigns*, ed. by M. Crowley (The Hastings Center, Garrison 2008) pp. 111–116

44.16 M. Johansson: Working for next to nothing: Labor in the global nanoscientific community. In: *The Social Life of Nanotechnology*, ed. by B.H. Harthorn, J. Mohr (Routledge, New York 2012) pp. 88–110

44.17 B.H. Harthorn, J. Mohr (Eds.): *The Social Life of Nanotechnology* (Routledge, New York 2012)

44.18 B. Harthorn, J. Mohr: Introduction: The social scientific view of nanotechnologies. In: *The Social Life of Nanotechnology*, ed. by B.H. Harthorn, J. Mohr (Routledge, New York 2012) pp. 1–15

44.19 R. Appelbaum, M. Gebbie, S. Han, G. Stocking: Will China's quest for indigenous innovation succeed? Some lessons from nanotechnology, Technol. Soc. **46**, 149–163 (2016)

44.20 R. Appelbaum, R. Parker: China's bid to be the global nanotech leader: Advancing nanotechnology through state-led programs and international collaborations, Sci. Public Policy **35**(5), 319–334 (2008)

44.21 C. Cao, R. Appelbaum, R. Parker: Research is high and the market is far away: Commercialization of nanotechnology in China, Technol. Soc. **35**, 55–64 (2013)

44.22 X. Han, G. Stocking, M. Gebbie, R. Appelbaum: Will they stay or will they go? International graduate students and their decisions to stay or leave the U.S. upon graduation, PLoS ONE **10**(3), e0118183 (2015)

44.23 N. Invernizzi: Implications of nanotechnology for labor and employment: Assessing nanotechnology products in Brazil. In: *Can Emerging Technologies Make a Difference in Development?*, ed. by R. Parker, R. Appelbaum (Routledge, New York 2012) pp. 140–152

44.24 S. Cozzens, J. Wetmore (Eds.): *Nanotechnology and the Challenges of Equity, Equality and Development* (Springer, Dordrecht 2011)

44.25 C. Engeman, J. Rogers-Brown, B. Harthorn: Organized but uninvited: Civil society organization in the nanotechnology sphere. In preparation, CNS-UCSB, Santa Barbara, CA. Paper presented at the Society for the Study of Social Problems, Montreal, Canada, Aug 11–13, 2017.

44.26 D. Kahan, D. Braman, P. Slovic, J. Gastil, G. Cohen: Cultural cognition of the risks and benefits of nanotechnology, Nat. Nanotechnol. **4**, 87–90 (2009)

44.27 B. Bimber, L. Copeland: Digital media and traditional political participation over time in the US, J. Info. Technol. Politics **10**(2), 125–137 (2013)

44.28 C. Bosso (Ed.): *Governing Uncertainty: Environmental Regulation in the Age of Nanotechnology* (Routledge, New York 2010)

44.29 D. Scheufele, E. Corley, T.-J. Shih, K. Dalrymple, S. Ho: Religious beliefs and public attitudes toward nanotechnology in Europe and the United States, Nat. Nanotechnol. **4**(2), 91–94 (2009)

44.30 E. Corley, Y. Kim, D. Scheufele: Leading U.S. nanoscientists' perceptions about nano media coverage and the public communication of scientific research findings, J. Nanopart. Res. **13**(12), 7041–7055 (2011)

44.31 D. Brossard, D. Scheufele, E. Kim, B. Lewenstein: Religiosity as a perceptual filter: Examining processes of opinion formation about nanotechnology, Public Underst. Sci. **18**(5), 546–558 (2008)

44.32 National Academies of Sciences, Sackler Colloquia. Info. available at: http://www.nasonline.org/programs/sackler-colloquia/completed_colloquia/ (2012, 2013)

44.33 National Academies of Sciences: The Science of Science Communication II, Proc. Natl. Acad. Sci.

USA **111**(Supplement 4), 13583–13671 (2014)
44.34 W.P. McCray: From lab to iPod: A story of discovery and commercialization in the post-Cold War era, Technol. Cult. **50**(1), 58–81 (2009)
44.35 C. Mody: *Instrumental Community: Probe Microscopy and the Path to Nanotechnology* (MIT Press, Cambridge 2011)
44.36 W.P. McCray: *The Visioneers: How a Group of Elite Scientists Pursued Space Colonies, Nanotechnologies, and a Limitless Future* (Princeton University Press, Princeton 2013)
44.37 P. Shapira, J. Youtie, Y. Li: Social science contributions Compared in synthetic biology and nanotechnology, J. Responsible Innov. **2**(1), 143–148 (2015)
44.38 J. Youtie, J. Rogers, T. Heinze, P. Shapira, L. Tang: Career-based influences on scientific recognition in the United States and Europe: Longitudinal evidence from curriculum vitae data, Res. Policy **42**(8), 1341–1355 (2013)
44.39 J. Youtie, L. Kay: Acquiring nanotechnology capabilities: Role of mergers and acquisitions, Technol. Anal. Strateg. Manag. **26**(5), 547–563 (2014)
44.40 L. Kay, J. Youtie, P. Shapira: Signs of things to come? What patent submissions by small and medium-sized enterprises say about corporate strategies in emerging technologies, Technol. Forecast. Soc. Change **85**, 17–25 (2014)
44.41 S. Cozzens: Building equity and equality into nanotechnology. In: *The Yearbook of Nanotechnology in Society, Vol. II: Nanotechnology and the Challenges of Equity, Equality and Development*, ed. by S. Cozzens, J. Wetmore (Springer, New York 2011) pp. 433–457
44.42 P. Slovic (Ed.): *The Perception of Risk* (Earthscan, London 2000)
44.43 N. Pidgeon, R. Kasperson, P. Slovic (Eds.): *The Social Amplification of Risk* (Earthscan, London 2003)
44.44 N. Pidgeon, B.H. Harthorn, T. Satterfield: Nanotechnology risk perceptions and communication: Emerging technologies, emerging challenges, Risk Anal. (special issue) **31**(11), 1694–1700 (2011)
44.45 N. Pidgeon, B.H. Harthorn, K. Bryant, T. Rogers-Hayden: Deliberating the risks of nanotechnologies for energy and health applications in the United States and United Kingdom, Nat. Nanotechnol. **4**(2), 95–98 (2009)
44.46 T. Satterfield, M. Kandlikar, C. Beaudrie, J. Conti, B.H. Harthorn: Anticipating the perceived risk of nanotechnologies: will they be like other controversial technologies, Nat. Nanotechnol. **4**, 752–758 (2009)
44.47 B. Harthorn, N. Pidgeon, T. Satterfield: CNS-UCSB IRG 3 Synthesis Report: Understanding Nanotechnologies' Risks and Benefits: Emergence, Expertise and Upstream Participation (CNS-UCSB, Santa Barbara 2016), available online at http://www.cns.ucsb.edu.
44.48 J. Schot, A. Rip: The past and future of Constructive Technology Assessment, Technol. Forecast. Soc. Change **54**, 251–268 (1996)
44.49 D. Guston: Understanding 'anticipatory governance, Soc. Stud. Sci. **44**(2), 218–242 (2014)
44.50 R. Owen, J. Bessant, M. Heintz (Eds.): *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (Wiley, Chichester, West Sussex 2013)
44.51 B.H. Harthorn: Methodological challenges posed by emergent nanotechnologies and cultural values. In: *The Handbook of Emergent Technologies and Social Research*, ed. by S. Hesse-Biber (Oxford University Press, Oxford 2011) pp. 65–88
44.52 W. Poortinga, N.F. Pidgeon: *Public Perceptions of Genetically Modified Food and Crops, and the GM Nation? Public Debate on the Commercialisation of Agricultural Biotechnology in the UK*, Understanding Risk Working Paper 04-01 (Centre for Environmental Risk, Norwich 2004)
44.53 J. Willsdon, R. Willis: *See Through Science: Why Public Engagement Needs to Move Upstream* (Demos, London 2004)
44.54 D. Kahan, J. Gastil, J. Reedy, D. Braman: Deliberation across the cultural divide: Assessing the potential for reconciling conflicting cultural orientations to reproductive technology, Paper 4688, Yale University Faculty Scholarship Series, (Yale University, New Haven 2008), available online at http://digitalcommons.law.yale.edu/fss_papers/4688
44.55 P. Devine-Wright: Reconsidering public attitudes and public acceptance of renewable energy technologies: a critical review, Working Paper 1.4 (2007) School of Environment and Development, University of Manchester, Oxford Road, Manchester M13 9PL, UK, available online at http://www.sed.manchester.ac.uk/research/beyond_nimbyism/
44.56 T. Dietz, P. Stern (Eds.): *Public Participation in Environmental Assessment and Decision Making* (National Academies Press, Washington DC 2008)
44.57 D. Fiorino: Citizen participation and environmental risk: A survey of institutional mechanisms, Sci. Technol. Hum. Values **15**(2), 226–243 (1990)
44.58 B.H. Harthorn, C. Shearer, J. Rogers: Exploring ambivalence: techno-enthusiasm and skepticism in US nanotech deliberations. In: *Quantum Engagements: Social Reflections of Nanoscience and Emerging Technologies*, ed. by T. Zuelsdorf (IOS, Amsterdam 2011) pp. 75–89
44.59 T. Satterfield, J. Conti, B.H. Harthorn, N. Pidgeon, A. Pitts: Understanding shifting perceptions of nanotechnologies and their implications for policy dialogues about emerging technologies, Sci. Public Policy **40**(2), 247–260 (2013)
44.60 M. Cobb: Creating informed public opinion: citizen deliberation about nanotechnologies for human enhancements, J. Nanoparticle Res. **13**(4), 1533–1548 (2011)
44.61 S. Davies, P. Macnaghten: Narratives of mastery and resistance: Lay ethics of nanotechnology, NanoEthics **4**(2), 141–151 (2010)
44.62 P. Strandbakken, G. Scholl, E. Stø: *Consumers and Nanotechnology: Deliberative Processes and*

Methodologies (CRC, Boca Raton 2012)
44.63 J. Rogers-Brown, C. Shearer, B.H. Harthorn: From biotech to nanotech: public debates about technological modification of food, Env. Soc.: Adv. Res. **2**(1), 149–169 (2011)
44.64 B.H. Harthorn, J. Rogers, C. Shearer, T. Martin: Debating Nanoethics: U.S. public perceptions of nanotechnology applications for energy and the environment. In: *Debating Science: Deliberation, Values, and the Common Good*, 2nd edn., ed. by D. Scott, B. Francis (Prometheus Books, New York 2012) pp. 227–249
44.65 N. Pidgeon, W. Poortinga, G. Rowe, T. Horlick-Jones, J. Walls, T. O'Riordan: Using surveys in public participation processes for risk decision making: The case of the 2003 British GM Nation? public debate, Risk Anal. **25**(2), 467–479 (2005)
44.66 A. Corner, N. Pidgeon: Nanotechnologies and upstream public engagement: Dilemmas, debates and prospects. In: *The Social Life of Nanotechnology*, ed. by B.H. Harthorn, J. Mohr (Routledge, New York 2012) pp. 247–283
44.67 B. Wynne: Public participation in science and technology: Performing and obscuring a political – Conceptual category mistake, East Asian Sci. Technol. Soc. Int. J. **1**, 99–110 (2007)
44.68 I. Welsh, B. Wynne: Science, scientism and imaginaries of publics in the UK: passive objects, incipient threats, Sci. Culture **22**(4), 540–566 (2013)
44.69 C. Bosso: Settling into the midstream? Lessons for governance from the decade of nanotechnology, J. Nanopart. Res. **18**, 163 (2016)
44.70 E. Michelson: *Assessing the Societal Implications of Emerging Technologies: Anticipatory Governance in Practice* (Earthscan/Routledge, New York 2016)
44.71 Nanotechnology Informal Science Education Network, Boston Museum of Science: http://nisenet.org/
44.72 C. Toumey: Science and democracy, Nat. Nanotechnol. **1**, 6–7 (2006)
44.73 A. Jones, A. Anderson, S. Yeo, A. Greenberg, D. Brossard, J. Moore: Using a deliberative exercise to foster public engagement in nanotechnology, J. Chem. Educ. **91**, 179–187 (2014)
44.74 J. Schummer: Cultural diversity in nanotechnology ethics, Interdiscip. Sci. Rev. **31**(3), 217–230 (2006)
44.75 F. Allhoff, P. Lin, J. Moor, J. Weckert (Eds.): *Nanoethics: The Ethical and Social Implications of Nanotechnology* (Wiley, Hoboken 2007)
44.76 K. McComas, J. Besley: Fairness and nanotechnology concern, Risk Anal. **31**(11), 1749–1761 (2011)
44.77 D. Evensen: Ethics and 'fracking': A review of (the limited) moral thought on shale gas development, WIREs Water **3**, 575–586 (2016)
44.78 J. Weckert, J. Moor: The precautionary principle in nanotechnology. In: *Nanoethics: the ethical and social implications of nanotechnology*, ed. by F. Allhoff, P. Lin, J. Moor, J. Weckert (Wiley, Hoboken 2007) pp. 133–146
44.79 F. Allhoff, P. Lin, D. Moore: *What is Nanotechnology and Why Does It Matter?: From Science to Ethics* (Wiley-Blackwell, Hoboken 2010)
44.80 NIOSH: Approaches to Safe Nanotechnology: Managing the Health and Safety Concerns Associated with Engineered Nanomaterials, DHHS (NIOSH) Publication # 2009-125 (NIOSH, Atlanta 2009), available online at http://www.cdc.gov/niosh/docs/2009-125/
44.81 J. Sass: Nanotechnologies and precautionary principle. Presentation (NRDC, Washington DC 2006), available online at https://www.nrdc.org/sites/default/files/hea_06121401a.pdf
44.82 Center for Responsible Nanotechnology: http://www.crnano.org/precautionary.htm
44.83 R. Parker, R. Appelbaum (Eds.): *Can Emerging Technologies Make a Difference in Development?* (Routledge, New York 2012)
44.84 D. Guston, J. Parsi, J. Tosi: Anticipating the ethical and political challenges of human nanotechnologies. In: *Nanoethics: the Ethical and Social Implications of Nanotechnology*, ed. by F. Allhoff, P. Lin, J. Moor, J. Weckert (Wiley, Hoboken 2007) pp. 185–197
44.85 O. Renn, M. Roco: Nanotechnology and the need for risk governance, J. Nanopart. Res. **8**(2), 153–191 (2006)
44.86 O. Renn, M. Roco: Nanotechnology Risk Governance, White paper No. 2 (International Risk Governance Council, Geneva 2006)
44.87 California Department of Toxic Substances Control (2010, 2011) Chemical Information Call-in. https://www.dtsc.ca.gov/PollutionPrevention/Chemical_Call_In.cfm, http://www.dtsc.ca.gov/PollutionPrevention/Round_One.cfm, http://www.dtsc.ca.gov/PollutionPrevention/Round_Two.cfm
44.88 P. Schulte, G. Roth, L. Hodson, V. Murashov, M. Hoover, R. Zumwaalde, E. Kuempel, C. Geraci, A. Stefaniak, V. Castranova, J. Howard: Taking stock of the occupational safety and health challenges of nanotechnology: 2000–2015, J. Nanopart. Res. **18**, 159 (2016)
44.89 C. Beaudrie, M. Kandlikar, T. Satterfield: From cradle-to-grave at the nanoscale: gaps in US regulatory oversight along the nanomaterial life cycle, Environ. Sci. Technol. **47**(11), 5524–5534 (2013)
44.90 C. Beaudrie, T. Satterfield, M. Kandlikar, B.H. Harthorn: Expert views on regulatory preparedness for managing the risks of nanotechnologies, PLOS One **8**(11), e80250 (2013) doi:10.1371/journal.pone.0080250
44.91 C. Engeman, L. Baumgartner, B. Carr, A. Fish, J. Meyerhofer, T. Satterfield, P. Holden, B.H. Harthorn: Governance implications of nanomaterials companies' inconsistent risk perceptions and safety practices, J. Nanopart. Res. **14**(749), 1–12 (2012)
44.92 C. Engeman, L. Baumgartner, B. Carr, A. Fish, J. Meyerhofer, T. Satterfield, P. Holden, B.H. Harthorn: The hierarchy of environmental, health, and safety practices, in the US nanotech-

nology workplace, J. Occup. Environ. Hyg. **10**(9), 487–495 (2013)

44.93 R. Owen, J. Stilgoe, P. Macnaghten, M. Gorman, E. Fisher, D. Guston: A framework for responsible innovation. In: *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, ed. by R. Owen, J. Bessant, M. Heintz (Wiley, Chichester 2013) pp. 27–50

44.94 J. van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops, H. Romijn (Eds.): *Responsible Innovation 1: Innovative Solutions for Global Issues* (Springer, Dordrecht 2014)

44.95 S. Arnaldi, A. Ferrari, P. Magaudda, F. Marin (Eds.): *Responsibility in Nanotechnology Development* (Springer, Dordrecht 2014)

44.96 S. Davies, C. Selin, G. Gano, A. Pereira: Citizen engagement and urban change: Three case studies of material deliberation, Cities **29**(6), 351–357 (2012)

44.97 P. Stern, H. Fineberg: *Understanding Risk: Informing Decisions in a Democratic Society* (National Academies Press, Washington DC 1996)

44.98 N. Pidgeon, C. Demski, C. Butler, K. Parkhill, A. Spence: Creating a national citizen engagement process for energy policy, Proc. Natl. Acad. Sci. USA **111**, 13606–13613 (2014)

44.99 Center for Nanotechnology in Society at University of California at Santa Barbara: Cumulative presentations 2005-2016, http://www.cns.ucsb.edu/presentations

44.100 H. Collins, R. Evans: The third wave of science studies, Soc. Stud. Sci. **32**(2), 235–296 (2002)

44.101 M. Gorman, J. Groves, R. Catalano: Societal dimensions of nanotechnology, IEEE Technol. Soc. Mag. **29**(4), 55–64 (2004)

44.102 C. Beaudrie, M. Kandlikar, G. Ramachandran: Using expert judgment for risk assessment. In: *Assessing Nanoparticle Risks to Human Health*, ed. by G. Ramachandran (Elsevier, Oxford 2016) pp. 109–138

44.103 D. Berube, C. Cummings, M. Cacciatore, D. Scheufele, J. Kalin: Characteristics and classification of nanoparticles: Expert Delphi survey, Nanotoxicology **5**(2), 236–243 (2011)

44.104 B. Fastman, M. Metzger, B.H. Harthorn: Forging new connections between nanoscience and society in the UCSB Center for Nanotechnology in Society Science & Engineering Fellows Program. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer International, Cham 2016) pp. 375–393

44.105 E. Fisher, M. O'Rourke, E. Kennedy, R. Evans, M. Gorman, T. Seager: Mapping the integrative field: Taking stock of socio-technical collaborations, J. Responsible Innov. **2**(1), 39–61 (2015)

45. Environment, Health and Safety Issues in Nanotechnology

Rui Chen, Chunying Chen

With the rapid development of nanotechnology and its application, lots of nanomaterials or nanorelated products are now used in society. Nanotechnology offers substantial economic and societal benefits, but its impacts on environment, health, and safety (EHS) issues are not clearly understood or defined. Interactions between nanomaterials (sourced from nanotechnology development) and the human body and even with the ecological system have attracted much concern. In this chapter, the impacts of the development of nanotechnology on EHS issues has been surveyed focusing on present knowledge of the most important nanomaterials, dominant physicochemical characteristics which contribute to relevant toxicities, and state-of-the-art techniques and established biomarkers within this research field. This information may not be enough to fill the knowledge gap concerning the impacts of nanotechnology on EHS, but should help scientists and authorities to realize the risks involved and to take steps for sustainable development in the foreseeable future.

45.1 **Impacts of the Development of Nanotechnology on Environment, Health, and Safety Issues** 1559

45.2 **Current Progress of the Most Important Nanomaterials** 1562
45.2.1 Carbon Nanotubes 1562
45.2.2 Silver Nanoparticles 1568
45.2.3 Zinc Oxide Nanoparticles 1568

45.3 **Physicochemical Characteristics of Nanoparticles that Determine the Toxicity Impacts** 1569
45.3.1 Particle Size and Surface Area 1570
45.3.2 Nanostructure and Shape 1571
45.3.3 Agglomeration and Aggregation 1572
45.3.4 Chemical Composition, Purity, and Impurities 1573
45.3.5 Coating, Surface Modification, and Surface Charges 1573

45.4 **Novel Techniques and Biomarker Development in Nanotoxicology** 1574
45.4.1 The Development of Biomarkers for Evaluation of EHS Impacts 1574
45.4.2 Novel Techniques Used for ADME Study in Nanotoxicology 1576

45.5 **Conclusion and Perspectives** 1580

References 1580

45.1 Impacts of the Development of Nanotechnology on Environment, Health, and Safety Issues

As illustrated in this handbook, nanotechnology has undergone broad development in terms of the techniques of characterization, fabrication, and manufacturing to keep pace with actual requirements from fields such as electronics, cosmetics, catalysis, gas sensing, energy storage, structural materials, environmental remediation, medical diagnostics, and drug delivery [45.1]. Nanomaterials (NMs) or nanoparticles (NPs) with novel physicochemical properties are the main products or research subjects in nanotechnology development. Among such properties, the most significant change in the properties of NPs is the increase in the surface-to-volume ratio compared to their large bulk counterparts. Importantly, the emergence of these products and research subjects, along with their applications and extinctions has led to great concern over adverse impacts on the environment, health, and safety (EHS). NPs may be released into the environment with the possibility of human or ecosystem exposure at various stages in their life cycles (Fig. 45.1). It is proposed that NPs may enter the body and accumulate in organs, such as lung, liver, spleen, kidneys, or brain, potentially causing harm or death to humans and animals, and the release of NPs into the environment might

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_45

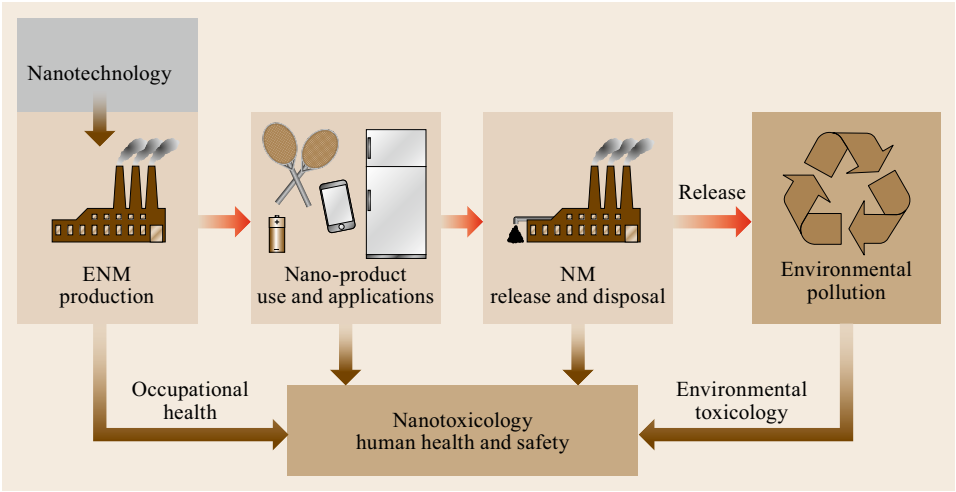


Fig. 45.1 The impacts of the development of nanotechnology on environment, health, and safety (EHS). The *red arrows* show the flows of nanomaterials to the environment: production, use, and disposal of nanoproducts, and release in the environment. The *brown arrows* show the potential health and safety impacts from procedures initiated by the development of nanotechnology

harm the ecological system. Consequently, the field of nanotoxicology emerged around 2004 to 2005 in an attempt to answer critical questions related to these EHS issues [45.2, 3]. *Oberdörster* [45.4] defined nanotoxicology as the study of the adverse effects of engineered nanomaterials (ENMs) on living organisms and the ecosystem, including the prevention and amelioration of such adverse effects. According to a report [45.5] from the Committee to Develop a Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials (National Research Council, USA), there are many issues in relation to EHS of NMs:

1. Engineered nanomaterial sources and manufacturing
2. Modifications, fate, and transport
3. Bioavailability and dose
4. Effects on organisms and ecosystems.

As particularly important for EHS, topics related to nanotoxicology and ecotoxicology of NMs will be paid the most attention in this chapter.

NMs/NPs are also defined as nanosized materials/particles that have at least one dimension less than 100 nm. Man-made NMs are generally called engineered nanomaterials (ENMs). Ultrafine particles (UFPs) refers to airborne particulates of less than 100 nm in aerodynamic diameter, i. e., particles generated through the use of nanotechnology, diesel exhaust particles, products of cooking, heating, and wood burning in indoor environments [45.9]. Alongside the booming development of nanotechnology, life-cycle assessment of EHS impacts is important yet difficult to carry out because of the lack of enough detail on the ENMs [45.6]:

1. The manufacturing processes of newly developed NMs are usually evolutional, i. e., with constant changes before final standardization. Thus, EHS impacts could vary considerably among different production procedures.
2. The formula of NMs and related manufacturing processes are new and often subject to confidentiality constraints in the nanotechnology industry.
3. NM release and ecotoxicology impacts are ambiguous at both the nanoproduct use and application stage and at the final disposal and recycling process stage.

This means risk estimations for impacts of NMs on EHS vary from one scenario to another. However, we still need to deepen our understanding of such impacts from the whole life-cycle perspective in order to address uncertainties over possible future outcomes. This may help to bridge the knowledge gap and provide an opportunity to improve the health and safety of scientists, workers, and consumers who work with or use NMs/NM products or are exposed to nanoparticles in society.

The NM production process plays an important role in nanotechnology. This is because a high release potential correlates directly with the mass production process of NMs, generally involving pure and uncombined nanomaterials. Thus, workers in this environment are not only at the forefront of risk due to advancements in nanotechnology, but also their working environments have

Table 45.1 Production/utilization quantities of ten nanomaterials in the world and in Europe (in t/yr). Reproduced with permission from [45.8]

ENMs	Worldwide (t/yr) median and 25/75 percentile	Europe (t/yr) median and 25/75 percentile	US (t/yr) [45.6] range	Switzerland (t/yr) [45.7] in brackets values extrapolated to Europe
TiO_2	3000 (550–5500)	550 (55–3000)	7800–38 000	435 (38 000)[b]
ZnO	550 (55–550)	55 (5.5–28 000)		70 (6100)
SiO_2	5500 (55–55 000)	5500 (55–55 000)		75 (6500)
FeO_x	55 (5.5–5500)	550 (30–5500)		365 (32 000)
AlO_x	55 (55–5500)	550 (0.55–500)		0.005 (0.4)
CeO_x	55 (5.5–550)	55 (0.55–2800)	35–700	
CNT	300 (55–550)	550 (180–550)	55–1101	1 (87)
Fullerenes	0.6 (0.6–5.5)	0.6 (0.6–5.5)	2–80	
Ag	55 (5.5–550)	5.5 (0.6–55)	2.8–20	3.1 (270)
Quantum dots (QDs)	0.6 (0.6–5.5)	0.6 (0.6–5.5)		

[a] The median and the 25/75 percentile are given, rounded to two significant numbers.
[b] The values in brackets for Switzerland were extrapolated using the population of Switzerland (6.9 million) to Europe (593 million)

high priority for the establishment and implementation of responsible practices [45.10]. The basis of evaluation of EHS impacts from nanotechnology is the structure of the ENM life cycle. The first step in NM release assessment is to obtain knowledge of the production process which determines the maximum amount of NMs that could potentially be released to the environment. This means that the most accurate estimation of NMs that could potentially be released to the environment may be the one that is based on the production volume of the raw nanomaterials. In 2012, *Piccinno* et al. [45.8] summarized the production quantities of ten engineered nanomaterials based on their own survey and other literature sources [45.6, 7]. NMs like silica, titania, alumina, iron oxides, and zinc oxides dominate the ENM market in terms of mass flow through the global economy (Table 45.1). Within this list, TiO_2 is by far the most significant ENM in terms of exposure based on modeling estimation [45.11, 12]. The modeling of the NM's flow should generally track its entire life cycle which is illustrated in Fig. 45.1: this includes ENM production, incorporation into products, ENM release from products during use, transport and disposal of ENM during and after recycling processes, transfer to air, soil, water and sediments (ecosphere), and transport within environmental compartments of the nature system. The amounts of ENM in said compartments provides the basis for calculating the overall environmental concentrations of ENM [45.13]. Recently, using dynamic probabilistic material flow modeling, *Sun* et al. [45.14] found that the concentrations of all ENMs in all compartments of the nature system are increasing with the increase of production and applications of ENMs, i. e., of nano-TiO_2, nano-ZnO, nano-Ag, and carbon nanotubes (CNTs) for example, in the world. ENMs generally enter the Earth's surface water first and end up in the sediment where the accumulated concentrations of CNTs and nano-TiO_2 were estimated to be 6.7 μg/kg (CNTs) and about 40 mg/kg (nano-TiO_2) in the worst-case scenario [45.14]. The concentration of NMs in the nature system is a crucial factor in the evaluation of EHS impacts. However, one report shows that Ag ENM contributed to a higher ecotoxicity to freshwater algae, daphnia, and fish compared to TiO_2 ENMs, even though a much lower concentration had accumulated in the ecosystem [45.15].

In nanotechnology there are two principle nanoscale manufacturing strategies: the top-down and the bottom-up approach [45.16]. The top-down approach begins at the bulk level and involves miniaturization to nanoscale dimensions via carving or grinding methods, i. e., lithography, etching, or milling. The bottom-up approaches begin at the atomic or molecular level with growth through nucleation and/or growth from liquid, solid, or gas precursors via chemical reactions or physical processes, i. e., through techniques like sol-gel or epitaxy [45.16, 17]. The top-down approach dominates compared to the bottom-up approach, but it is believed to produce more waste. It has been suggested that the bottom-up approach represents should be emphasized in order to minimizing the unwanted waste for sustainable development in nanotechnology [45.17].

The release of NMs during production and manufacturing not only constitute a possible release pathway to the environment, but also has high relevance in regard to occupational exposure. With no release and no exposure there will be no impact on EHS. Actually, from a life-cycle viewpoint there is significant impact from

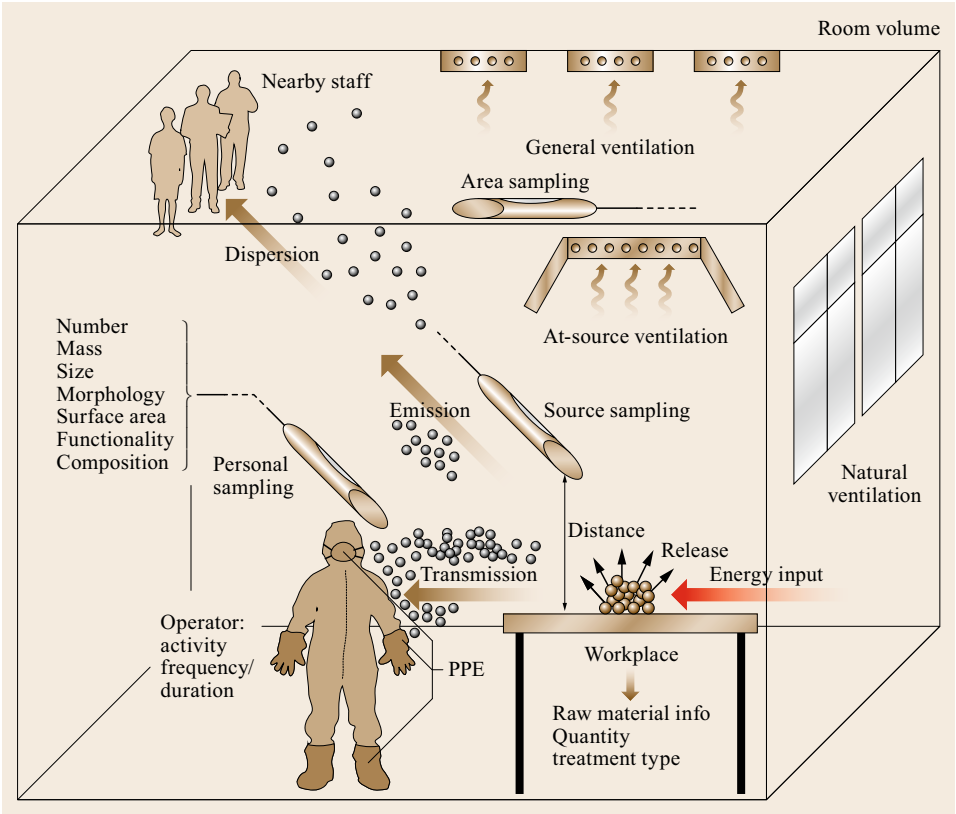


Fig. 45.2 Diagram representing various elements and processes in an occupational exposure scenario. Reproduced with permission from [45.18]

the process of NM production. Figure 45.2 shows a typical occupational NM-handling setting where ENMs could be released, emitted, and transported, resulting in actual exposure to workers [45.18]. Exposure assessment and management are critical in order to minimize impacts on the production environment, although there are too many variants in application situations. Detailed procedures and strategies for occupational exposure assessment and risk evaluation can be found in the literature [45.10, 18]. Further, appropriate risk management practices including engineering control systems in working environments or personal protective equipment (PPE) for individuals are important but beyond the scope of this chapter. In addition, it would be informative if more effort is paid in the future to evaluations of impacts during use and application processes, and the final disposal of nanorelated products.

45.2 Current Progress of the Most Important Nanomaterials

45.2.1 Carbon Nanotubes

Much of the concern associated with nanotechnology comes from the potential risks of CNTs [45.19, 20]. Typically a few nanometers in diameter, CNTs are long and thin cylinders of carbon with a single wall or multiple walls defined as single-walled carbon nanotubes (SWCNTs) and multiwalled carbon nanotubes (MWCNTs), respectively. They have a needle-like shape and are strong and tough in character, similar to asbestos. In vitro evidence showed that exposure to CNTs could reduce viability, induce apoptosis, and cause DNA (deoxyribonucleic acid) damage [45.23, 24]. These adverse effects may result from the interruption of cellular redox status after CNT exposure, e.g., increasing ROS (reactive oxygen species) and malondialdehyde (MDA) levels, as well as altering superoxide dismutase (SOD) activity and glutathione peroxidase (GSH-Px) levels as previously reported [45.25–27]. In vitro studies based on mammalian cell models often show inconsistent results because of the different cell lines, doses, and exposure times used. However, in vivo pulmonary tox-

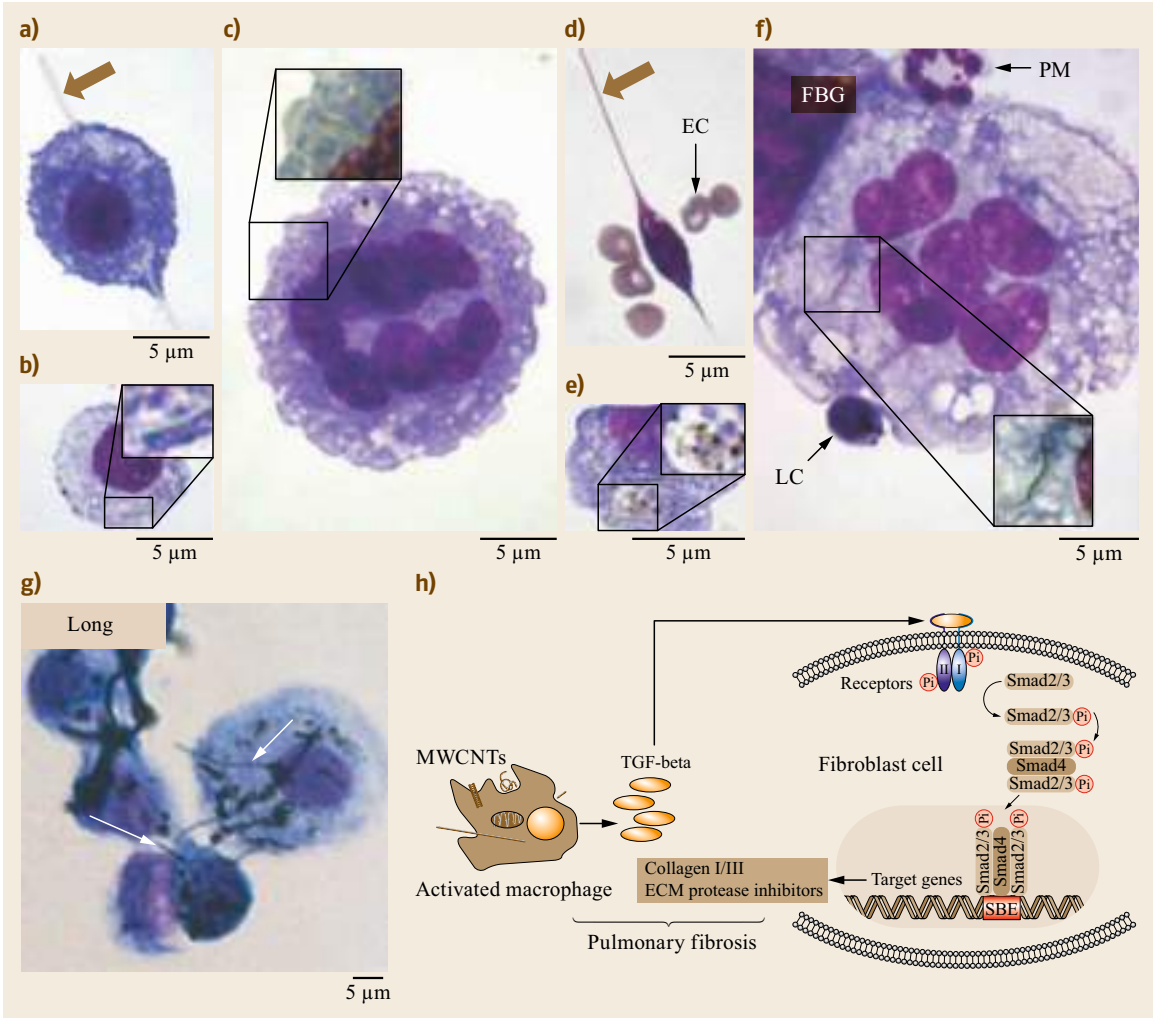


Fig. 45.3a–h The phagocytosis of asbestos and CNTs by macrophages. **(a,b)** Histological sections show incorporation of long-fiber amosite **(a)** (*arrow*) leading to *frustrated phagocytosis*, but short-fiber amosite **(b)** (see *inset*) is successfully phagocytosed by peritoneal macrophages. **(c)** Representative image of a foreign body giant cell after injection of long-fiber amosite containing short fiber fragments. **(d)** Like long-fiber amosite, long CNTs ($> 15\,\mu m$) also lead to frustrated phagocytosis in peritoneal macrophages (EC, erythrocytes). **(e)** In contrast, short CNTs ($< 5\,\mu m$) can be readily phagocytosed (see *inset*). **(f)** Foreign body giant cells (FBGC) are also present after injection of long CNTs in peritoneal macrophages (see *inset* for internalized fibers) (PMN, polymorphonuclear leukocyte; LC, lymphocyte). **(g)** Representative optical microscope image to show the in vivo cellular uptake of MWCNTs by alveolar macrophages. **(h)** MWCNTs promote pulmonary fibrosis through the TGF-β/small mothers against decapentaplegic (Smad) signaling pathway. All images are shown at $\times 1000$ magnification with a $5\,\mu m$ scale bar (after [45.21], **(g,h)** from [45.22] reprinted with permission)

icity studies usually show CNT causing clear adverse toxicity effects. It has been found in animal experiments that exposure to long CNTs (longer than $15\,\mu m$) at the mesothelial lining of the body cavity of mice will result in asbestos-like, length-dependent, pathogenic behaviors including inflammation and granulomas [45.21]. The macrophages at the mesothelial lining may resort to *frustrated phagocytosis* and giant cell formation caused by macrophage fusion in inflammatory conditions when attempting to phagocytose or engulf long CNTs (Fig. 45.3a–f). These results suggest a potential link between inhalation exposure to long CNTs and mesothelioma which is generally regarded as the specific type of cancer caused by inhalation exposure to

asbestos, although it remains unknown whether there will be sufficient exposure in the environment or workplace to actually cause it.

Granuloma formation and/or fibrotic responses in the lungs are the dominant injuries after inhalation exposure to CNTs. It is worth noting that long CNTs have much more significant effects [45.22, 28]. The underlying mechanisms were systematically investigated in our lab. The alveolar macrophages show the same *frustrated phagocytosis* phenomenon with peritoneal macrophages when phagocytosing long MWCNTs (Fig. 45.3g). It was found that TGF (transforming growth factor)-β1 could be induced in alveolar macrophages or released from the bronchiolar epithelium by MWCNT exposure, which plays an important role in MWCNT-caused pulmonary fibrosis (Fig. 45.3h). At the same time, the fibroblast-specific protein-1, α-smooth muscle actin, and collagen III were significantly upregulated by the induction of TGF-β1 [45.22]. Further research showed TGF-β1 secreted from macrophages serves as a critical paracrine stimuli to fibroblasts by directly activating TGF-β/small mothers against decapentaplegic homolog 2 (Smad2) signaling in fibroblasts or epithelial cells. This stimulation then induces the fibroblast-to-myofibroblast transition and epithelial–mesenchymal transition (EMT), respectively, both of which contribute to CNT-induced fibrotic responses [45.28].

The inhalation of carbon nanotubes has a high possibility to induce pulmonary fibrosis but also can directly cause acute and serious injuries to the pulmonary system [45.30, 31]. In our research, we tested the pulmonary toxicity of SWCNTs by intranasally instilling exposure using the spontaneously hypertensive rat model [45.32]. The inflammation, oxidative stress, and cell damage levels in the bronchoalveolar lavage fluid (BALF) were significantly increased at 24 h postexposure. The increased endothelin-1 in BALF and plasma and angiotensin I-converting enzyme in plasma suggested endothelial dysfunction in the pulmonary circulation and peripheral vascular thrombosis. Histological data on lung tissues showed that SWCNT-laden macrophages move to centrilobular locations and form multifocal pulmonary granuloma [45.32]. The formation of pulmonary granuloma lesions were the result of the lungs' immune response to the foreign substances that can not easily be eliminated [45.33, 34]. Further, the subchronic toxic effects and cardiovascular effects of CNTs were disclosed at 7 and 30 days postexposure to the inhalation system in rats [45.35]. These results suggest that respiratory exposure to CNTs not only induces acute pulmonary toxicity but also poses long-term adverse effects on the whole cardiovascular system. Humans with existing cardiovascular diseases are also susceptible to CNT exposure.

Knowledge of biodistribution and biopersistence in the body can help better understand the potential effects of their exposure to humans. Using the Raman spectroscopy mapping method it has been found that inhaled CNTs have the ability to penetrate deep into the lungs, cross the pulmonary epithelium, and enter the bloodstream [45.36]. Then, the extrapulmonary transport of CNTs was disclosed after inhalation exposure in C57BL/6 J mice by testing at 5 h/day for 12 days with a concentration of 5 mg/m^3 in a whole-body inhalation system [45.29]. It has been shown that inhaled MWCNTs, which were deposited in the lung, were transported to the parietal pleura, the respiratory musculature, liver, kidney, and heart in a singlet form and accumulated even until day 336 postexposure (Fig. 45.4). Similar biodistribution evidence for CNTs in liver, kidneys, and spleen has been reported after exposure to MWCNTs in rats using the inhalation system [45.35, 37]. After entering the blood circulation, CNTs will interact with blood proteins, i. e., fibrinogen (BFg), immunoglobulin, albumin, transferrin (Tf), and ferritin. Indications showed that there was a competitive binding of different blood proteins onto the surface of SWCNTs, which was governed by each protein's unique structure and the amount of hydrophobic residues of each kind of protein (Fig. 45.5). Atomic force microscopy (AFM) images indicated that the adsorption of transferrin and BSA (bovine serum albumin) quickly reached thermodynamic equilibrium at about 10 min, while fibrinogen and gamma-globulin (γ-Ig) gradually packed onto the SWCNT surface in a much longer period of about 5 h. Both fluorescence spectroscopy and SDS-PAGE have shown surprising competitive adsorptions among all the blood proteins examined, with the competitive order: BFg > γ-Ig > Tf > BSA. The far-UV circular dichroism (CD) spectra observation shows that the protein secondary structure has greatly changed for BFg and γ-Ig, with a decrease in the α-helical content and an increase in the β-sheet structure (Fig. 45.5g). In addition, our molecular dynamics simulations of SWCNTs binding with BFg, BSA, γ-Ig, and Tf complexes showed that both the contact residue numbers and binding surface areas exhibited the same sequence order: BFg > γ-Ig > Tf > BSA, in agreement with the experimental findings. Further analysis showed that the π–π stacking interactions between SWCNTs and aromatic residues (Trp, Phe, and Tyr) play a critical role in their binding capabilities. These different protein-coated SWCNTs may have different cytotoxicity by influencing the subsequent cellular responses [45.38].

The knowledge on the ADME (absorption, distribution, metabolism and excretion) of NMs could give help on the understanding of their adverse impacts to EHS. Much more research efforts should be taken

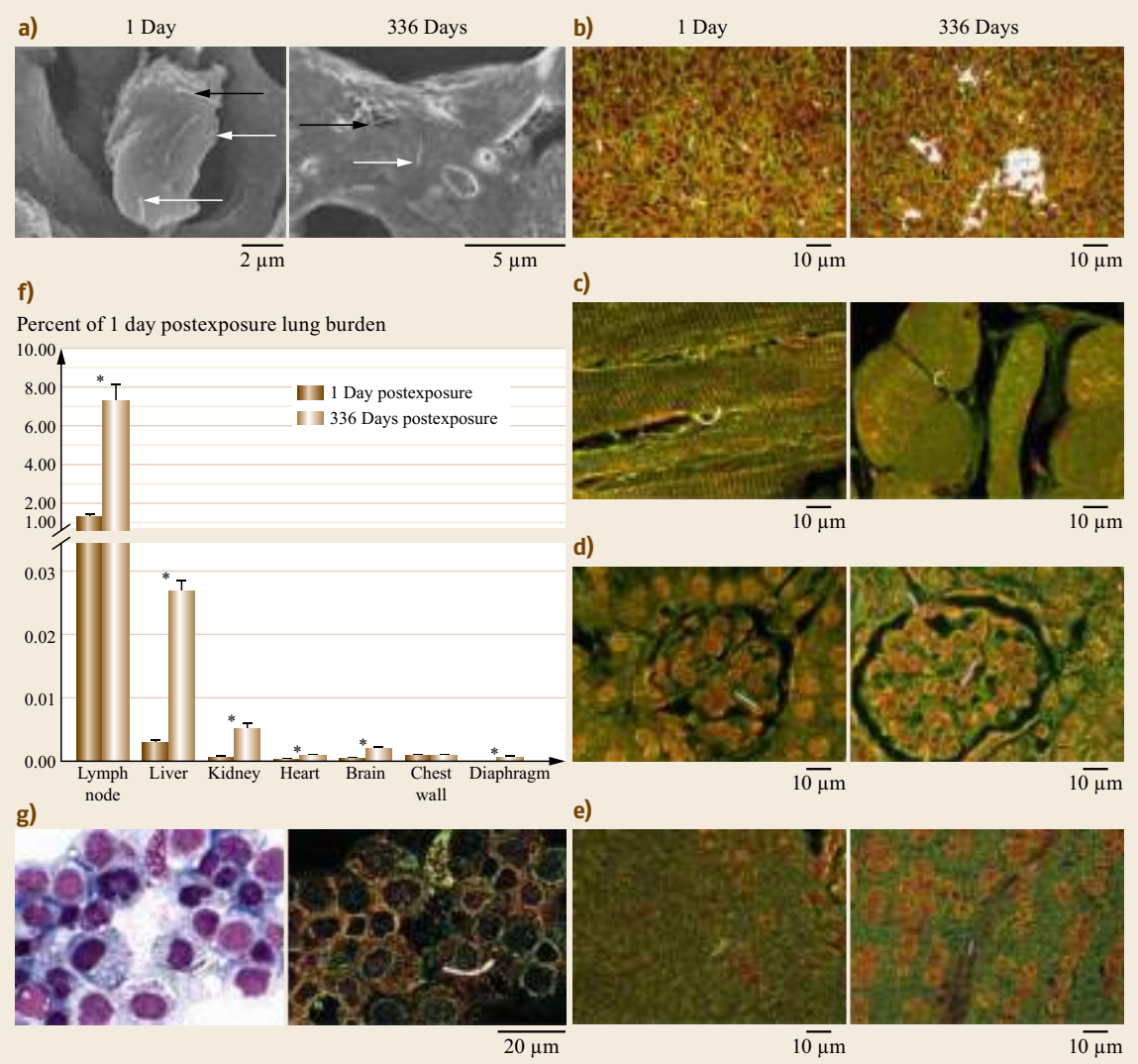


Fig. 45.4a–g Extrapulmonary biodistribution of CNTs after inhalation exposure. **(a)** Field-emission scanning electron microscopy (FESEM) images of alveolar macrophages 1 and 336 days postexposure. *Arrows* show the typical examples of MWCNT fibers protruding from the surface of alveolar macrophages (*black arrow* multiple fibers, *white arrows* singlets). **(b–e)** Enhanced dark-field images of tissue sections 1 and 336 days after CNT inhalation exposure. Tissues are from tracheobronchial lymph nodes, diaphragm, kidney and brain, respectively. **(f)** Percentage of 1 day postexposure lung burden detected in tracheobronchial lymph nodes, extrapulmonary organs, diaphragm, and chest wall. By 336 days postexposure there was a 7-fold increase in MWCNT in extrapulmonary organs and diaphragm over that measured at 1 day postexposure. Categories of extrapulmonary organs are ordered relative to MWCNT concentration in the respective tissue. *Asterisks* indicate significant difference between day 1 and 336 days postexposure values, $P < 0.05$. **(g)** Light and enhanced dark-field micrographs of MWCNTs detected in lavage of pleural space. The figure shows a comparison of the light and enhanced dark-field image of a singlet MWCNT in lavage of the pleural space in mice at 336 days postexposure (after [45.29])

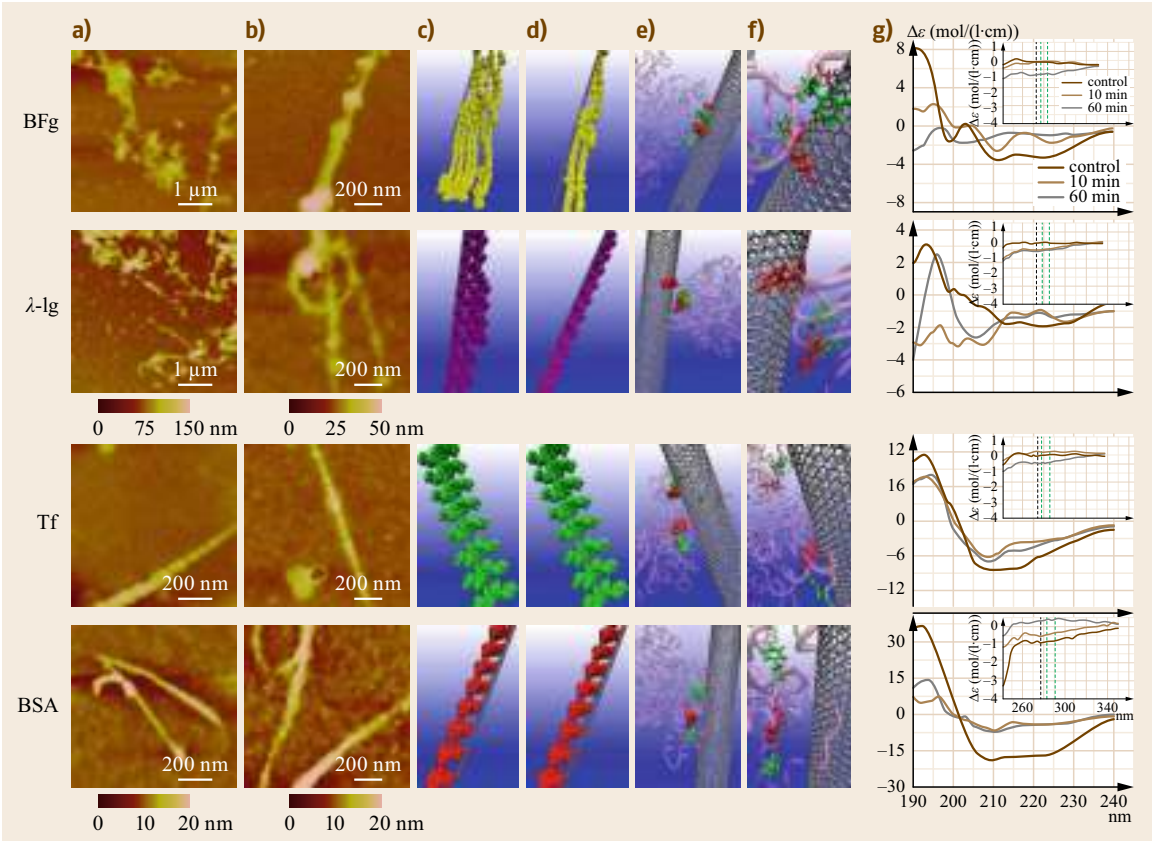
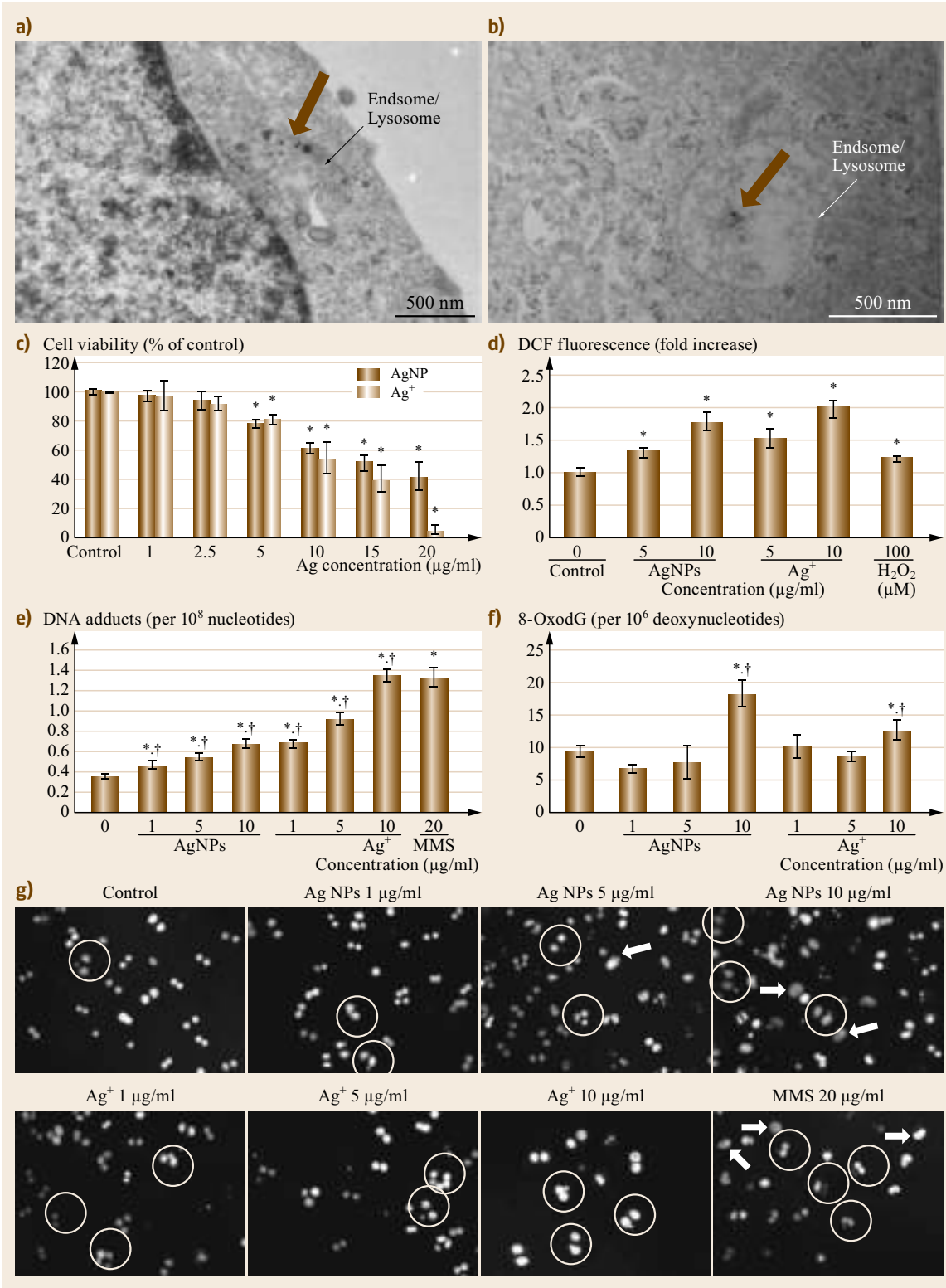


Fig. 45.5a–g The interactions between SWCNTs and human blood proteins (BFg, γ-Ig, Tf, BSA). AFM images of proteins after incubation with SWCNTs for 10 min **(a)** and 5 h **(b)**. Molecular modeling illustrations for proteins (bead representations) binding to SWCNTs after incubation for 10 min **(c)** and 5 h **(d)**. **(e)** Locations of the most preferred binding sites on proteins for SWCNTs. Residues highlighted in van der Waals representation corresponding to tyrosine in *red* and phenylalanine in *green*. Other protein parts are presented in *transparent pink* with the new cartoon drawing method. **(f)** The detailed orientations of aromatic rings of tyrosine and phenylalanine residues interacted to six-member rings of SWCNTs colored in *silver*. The tyrosine residues are rendered as Licorice representations and shown in *red*, with phenylalanine residues are in *green*. **(g)** The far-UV circular dichroism (CD) spectra of proteins after incubation with SWCNTs and the insets are near-UV CD spectra of proteins incubated with SWCNTs (after [45.38])

on this topic for CNTs. Other carbon NMs including fullerene and derivatives, graphene, and graphene oxide have a lower production quantity compared with CNTs (Table 45.1). Even so, a cautious attitude should be taken with all these carbon nanomaterials [45.40]. More information and data on the impact of these carbon NMs on EHS are urgently needed to support future effective assessment.

Fig. 45.6a–g Cytotoxicity and genotoxicity analysis of BSA-coated AgNPs in the cell line CHO-K1. TEM (transmission electron microscope) images of CHO-K1 cells when treated with 10 µg/ml of AgNPs for 6 h **(a)** and 24 h **(b)**. For cells treated with AgNPs and Ag^+, shown in the figure are changes in mitochondrial activity **(c)** by CCK-8 assay and intracellular ROS levels **(d)** that were labeled by DCFH-DA fluorescence density and determined by flow cytometry. DNA adducts induced by AgNPs and Ag^+ after 24 h exposure **(e)**. Methylmethanesulfonate (MMS) was used as a positive control. DNA oxidative adduct, 8-hydroxydeoxyguanosine (8-OxodG), which can be induced by AgNPs and Ag^+ after 24 h exposure **(f)**. Fluorescence microscopic images of micronuclei induced by AgNPs and Ag^+ after 24 h exposure **(g)**. Statistically significant difference from control is expressed as $*(P < 0.05)$ and significant difference between AgNPs and Ag^+ for the same amount of silver is expressed as $\dagger(P < 0.05)$. Reproduced with permission from Elsevier [45.39] ►


a)
Endsome/
Lysosome
500 nm
b)
Endsome/
Lysosome
500 nm
c) Cell viability (% of control)
AgNP
Ag+
Control
Ag concentration (µg/ml)
d) DCF fluorescence (fold increase)
Control
AgNPs
Ag+
H2O2
(µM)
Concentration (µg/ml)
e) DNA adducts (per 10^8 nucleotides)
AgNPs
Ag+
MMS
Concentration (µg/ml)
f) 8-OxodG (per 10^6 deoxynucleotides)
AgNPs
Ag+
Concentration (µg/ml)
g)
Control
Ag NPs 1 µg/ml
Ag NPs 5 µg/ml
Ag NPs 10 µg/ml
Ag+ 1 µg/ml
Ag+ 5 µg/ml
Ag+ 10 µg/ml
MMS 20 µg/ml

45.2.2 Silver Nanoparticles

With excellent antimicrobial activity, silver nanoparticles (AgNPs) are widely used as a major ingredient in industrial, consumer, and medical products [45.41, 42]. For example, AgNP-related products have been used in consumer products such as surface cleaners, room sprays, toys, antimicrobial paints, home appliances, food storage containers, textiles, and medical products including wound dressings, surgical instruments, urinary catheters, and bone prostheses. Several investigations have shown that silver-containing textiles could release significant amounts of Ag into the washing liquid when laundering fabrics [45.43–46]. It is inevitable that these discharged silver ions or silver nanoparticles are released into the ecosystem through sewer systems and constitute a high impact on EHS [45.47]. It has been discerned that Ag mostly transforms and precipitates to silver-containing particles like Ag_2S or AgCl during handling or disposal processes [45.46, 48]. Further, recent studies revealed that the reduction of Ag^+ in natural waters can form AgNPs naturally via the mediation of dissolved organic matter or sunlight [45.49–51]. This means that silver has complex transformation mechanisms in the environment [45.52].

As release from source products is inevitable, there is a high possibility that AgNPs will enter the human body and distribute within the circulation system or the organs after exposure. It has been shown that AgNPs are captured by the reticuloendothelial system and quickly distributed to other tissues after entering the blood [45.53–56]. In detail, AgNPs were mostly distributed to the liver, spleen, lung, and kidneys, and much less to the brain, heart, and testes after *i.v.* exposure [45.57, 58]. Interestingly, the concentrations of Ag in the blood remains steady with a slightly decreasing trend till at least 14 days for a single i.v. injection of AgNP, which means that AgNPs may be degraded and removed in a time-dependent manner [45.59]. The kinetic distributions of AgNPs in the body are important because they determine relevant toxicity. However, the mechanism of AgNP toxicity has not been clearly defined. Currently, there are three hypotheses to explain the toxicity of AgNPs after exposure. First, the ions released from NPs contribute to their toxicity. Second, the toxicity is caused through a silver ion-independent mechanism. Third, the Ag ions released from the endocytosed AgNPs in the cell dominate the toxicities, and this generally happens at endosome or lysosome locations. In reality, a combination of these may lead to the AgNP toxicity mechanism. *Jiang* et al. [45.39] showed that AgNPs were mostly localized in the endosome/lysosome in Chinese hamster ovary (CHO-K1) cells after exposure (Fig. 45.6a,b). Then, the cytotoxicity and genotoxicity of BSA-coated AgNPs on CHO-K1 were investigated. This showed that both BSA-AgNPs alone and Ag ions can up-regulate the cellular ROS level with the production of oxidative adducts, 8-hydroxydeoxyguanosine (8-OxodG), and induce the formation of micronuclei (Fig. 45.6c–f). The genotoxicity of AgNPs is prominent within the induced adverse outcomes, and merit special attention in future [45.60, 61]. Further research in our lab showed that there is a high possibility that AgNPs will cause adverse effects by the third mechanism listed above. Evidence concerning the intracellular fate of AgNPs is derived using x-ray absorption near-edge structure (XANES) spectroscopy, which is used to study the degradation and biointeraction of nanoparticles in biological systems [45.62, 63]. About 14.2% of the internalized silver was oxidized to Ag−O− after 12 h incubation of AgNPs in the CHO-K1 cell line. While 61.5% of the silver was present as Ag−S− after 24 h, suggesting that binding of Ag to sulfide groups of amino acids and proteins may be involved in the dissolution process of AgNPs. This intracellular binding of silver ions to proteins disrupts protein structure and function, thereby resulting in toxic effects [45.64].

The surface charge and various surface coatings of AgNPs play an important role in in vivo and in vitro toxicity, biodistribution, and pharmacokinetics [45.65]. Branched polyethyleneimine (BPEI)-coated AgNPs led to the highest toxicity in cells and mice, while the nearly neutral polyethylene glycol (PEG)-coated AgNPs showed low toxicity and long half-life in blood, indicating a high potential for application in medical products [45.65]. This means that the safety of silver NMs can be improved by adoption of particular surface coating strategies in their application [45.66–68]. Improving our understanding of the toxicity mechanisms of AgNPs could help with their safety evaluation and potential risks to EHS, and from different aspects such understanding may contribute to facilitating the design of safer and more effective antimicrobial products in the future.

45.2.3 Zinc Oxide Nanoparticles

Zinc oxide NPs have been used in personal care products including cosmetics and sunscreens as they efficiently absorb ultraviolet radiation and are also highly transparent to visible light. Attracting great concern are potential impacts through human exposure and health impacts at each stage of their manufacture and use [45.69, 70]. Dermal penetration is the first step in evaluating the toxic impacts of ZnO NPs on humans in regard to their long-term skin retention character.

Gulson et al. [45.71] tested the penetration efficiency of ZnO NPs by focusing on ^{68}Zn from enriched ^{68}Zn particles in sunscreens. Small amounts of ^{68}Zn tracer were tested in the blood of 20 human volunteers who received twice daily applications over 5 days. The highest ^{68}Zn concentrations were found in females after exposure to the ^{68}Zn nanoparticles [45.71]. This means that long-term retention of NMs on the skin may lead to dermal penetration by outer particles, although usually at the trace amounts level. For damaged skin, exposure of epidermal and dermal cells to high concentrations of ZnO NPs is sure to cause cytotoxicity and undesirable immune responses, which will instigate the penetration of even more particles to even deeper sites or even into the lymphatic or blood circulation [45.70, 72].

Inhalation exposure to ZnO fumes in the working environment can affect normal bodily functions, for example by causing metal fume fever, which is characterized in humans by short-term pulmonary and systemic alterations and is caused by inhalation of metal oxide aerosol [45.73]. It has been reported that humans exposed to ZnO particle fumes (MMD 300 nm) at 2.5 mg/m^3 for 2 h developed symptoms of metal fume fever with a significant increase of cytokine interleukin (IL)-6 in peripheral blood [45.74]. Human clinical experimental study showed that inhalation exposure of ZnO fumes for 10 to 30 min at a high concentration of 20–42 mg/m^3 increased the number of proinflammatory cytokines and neutrophils in BALF [45.75]. Animal experiments illustrated that inhalation of relatively lower concentrations of ZnO NPs, 1.1 to 4.9 mg/m^3 with occupationally relevancy for 2 week, can generate similar inflammation outcomes in the BALF in rats [45.76]. The National Institute for Occupational Safety and Health (NIOSH) recommended the permissible exposure limit of 5 mg/m^3 averaged over a work shift of up to 10 h/day, 40 h/week for zinc oxide fume inhalation exposure (NIOSH, publication number 0675). Whereas no exposure limit is available for nano-ZnO aerosol to date.

Both in vitro and in vivo systems illustrated that mechanisms underlying ZnO NP toxicity similarly include generation of ROS and subsequent DNA damage and apoptosis [45.77–79]. Oxidative stress produces *free radicals* and has been implicated in various adverse outcomes. For example, *Sharma* et al. [45.80] showed that significant DNA damage in a human epidermal cell line (A431) occurred upon ZnO NP exposure for 6 h at relatively higher concentrations (0.8 mg/ml to 5 mg/ml), while oxidative stress was induced at lower concentrations (0.008–0.8 mg/ml). Zinc ions released from the NPs serve as another key determinant for the toxicity of ZnO NPs [45.81]. Previous research showed that surface coating can considerably mitigate the cytotoxic potential of ZnO NPs by inhibiting ion release and/or minimizing ROS production, i. e., silica shell [45.82] and PEGylation [45.83]. In our own recent research, we found preincubation of ZnO NPs in a supplemented cell culture medium for 24 h could form a hard protein corona on the NP surface. These precoated stable protein layers on the NP surface facilitated further protein adsorption during the cell culture process. The amount of proteins adsorbed on precoated NPs could inhibit both ROS generation and ZnO dissolution and result in lower cytotoxicity [45.84]. This finding demonstrates an effective and convenient way to achieve safe biomedical and environmental applications of ZnO NPs. Besides surface coating, particle size, aspect ratio, doping type, and doping concentration all contribute to their toxicities [45.85]. Quantitative structure–activity relationships (QSAR) modeling is a powerful tool to predict how these physicochemical properties affect the toxicities of ZnO NPs [45.86]. Thus, there is a high chance of developing applicable models to guide the design and optimization of ZnO NPs with safe characteristics in terms of EHS issues.

45.3 Physicochemical Characteristics of Nanoparticles that Determine the Toxicity Impacts

Conventional toxicology places emphasis on chemical composition, dose, and exposure route as the primary variants in toxicity assessment of a chemical or a bulk material. However, many other factors, especially the physicochemical characteristics of nanoparticles, should be considered when we perform the same work in nanotoxicology (Fig. 45.7) [45.87]. This makes nanotoxicity evaluation much more complex and difficult than conventional toxicology [45.88]. There are two types of toxicity-related variants: intrinsic factors such as nanostructure, size, aggregation state etc. as listed in the right column of Table 45.2, and extrinsic factors, such as dose (presented by the metrics of mass/surface area/particle number), cell-/organ-specific responsiveness, exposure routes (e.g., pulmonary exposure, dermal exposure, oral gavage, ocular exposure, intravenous/intradermal/intramuscular/subcutaneous/intraperitoneal administrations), and biological species (e.g., rat, mouse, *Caenorhabditis elegans*, zebrafish, etc.) [45.87]. Understanding the func-

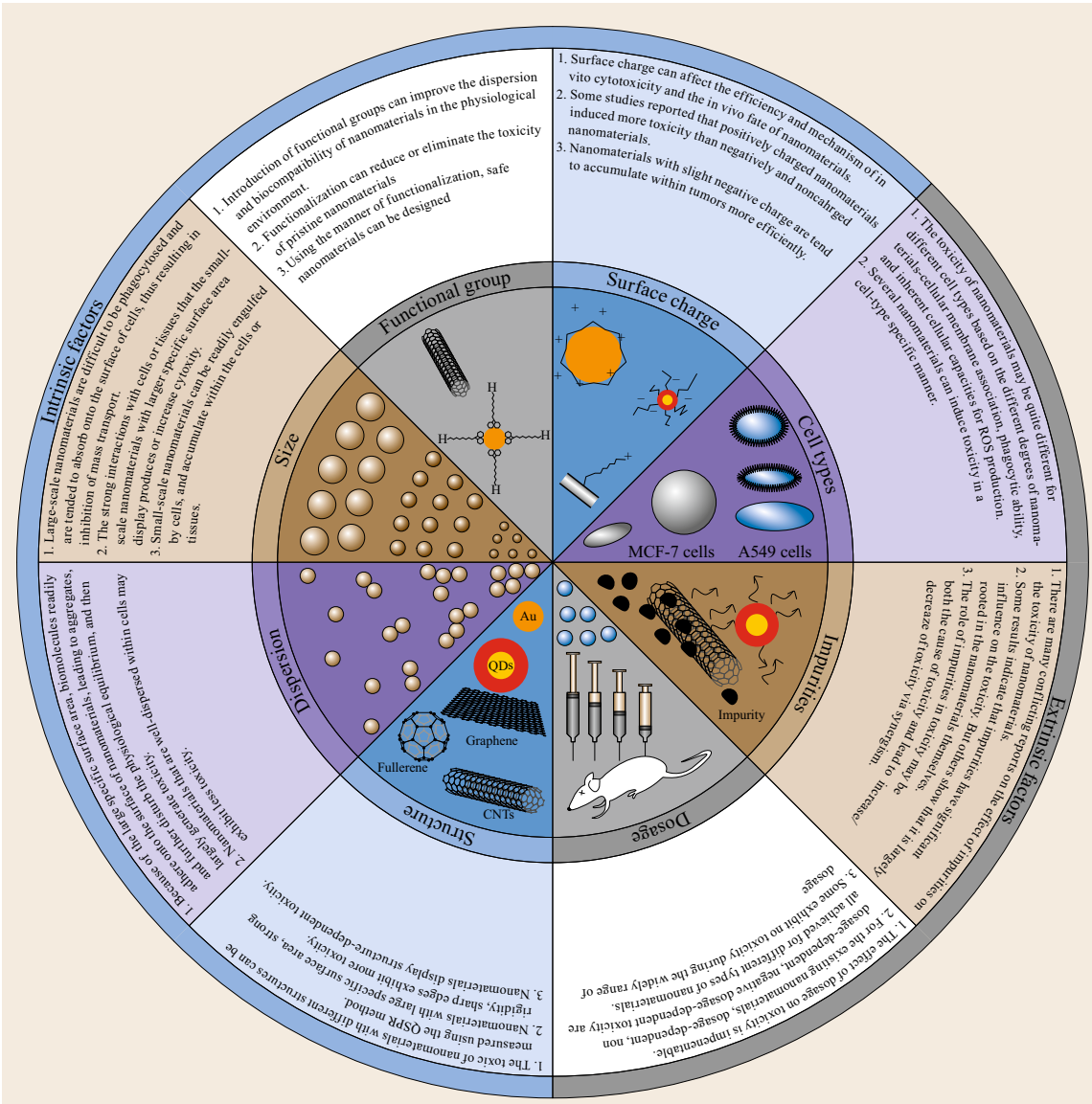


Fig. 45.7 Dominant factors influencing the toxicities of nanomaterials. They are classified into two categories: the intrinsic factors that mostly arise from the physicochemical properties of engineered nanoparticles (ENPs), and the extrinsic factors that include all other factors that can influence the toxic responses of NMs in experimental study or nanosafety assessment during real exposure scenarios (after [45.87])

tions of various physicochemical characteristics of NPs would facilitate the intended design of products with minimized EHS impacts in the future.

45.3.1 Particle Size and Surface Area

Particle size is a critical parameter in NM EHS impacts, which generally leads to the arising of other intrinsic factors. Particle size usually refers to the individual particle, not the agglomeration or aggregated ones which form during production, application, release, and disposal processes. As NM size decreases, the surface area/mass ratio exponentially increases and more atoms are displayed on the particle surface compared with the bulk material. The decreased particle size and increased surface areas make the NM surfaces more reactive to the surrounding biological components. Therefore, chemical reactions including catalytic

Table 45.2 Comparison of the physicochemical characteristics that dominate the toxic responses of nanomaterials with the bulk materials

Bulk material or nanomaterials	Nanomaterials
Chemical composition	Nanostructure (shape, crystal forms)
Particle mass concentration	Particle size/distribution
Reactivity	Particle number
Conductivity	Aggregation/agglomeration
Morphology (crystalline, amorphous, shape, etc.)	Degree and state of dispersion Zeta potential
Physical form (solid, aerosol, suspension, etc.)	Quantum effects Surface area
Purity/Impurities	Surface charge
Surface modification	Surface coating and chemistry
Solubility	Surface structure, crystal face
	Surface adsorbability
	Self-assembly
	...

reactions and redox reactions may occur at the surface of NPs causing the formation of ROS (such as superoxide anions or hydrogen peroxide) and subsequently oxidizing other interacting biomolecules. Many studies have indicated that a small size can increase toxic effects compared to the bulk counterpart [45.89–92]. The particle size influences cellular uptake in vitro, i. e., large latex spheres (>200 nm) had a slower uptake and processing ability compared to smaller ones (50 and 100 nm) [45.93]. In a well-known in vivo study example, inhalation of ultrafine TiO_2 particles (20 nm diameter, 12 week) showed significantly prolonged retention and more inflammation in lung than the inhalation of fine TiO_2 particles (250 nm diameter, 12 week) [45.94]. This means that the physicochemical characteristics of size and surface area have dominant roles in determining toxic potential when NPs interact with organisms.

45.3.2 Nanostructure and Shape

Engineered nanoparticles (ENPs) may exhibit various shapes and structures, e.g., spheres, needles, tubes, rods, etc. For example, CNTs exist as either SWCNTs which are graphite sheets rolled to form a seamless cylinder, or MWCNTs that have many cylinders stacked one inside the other. Similar to CNTs, fullerenes are carbon allotropes composed entirely of carbon, but they are different in form, e.g., hollow sphere, ellipsoid. One representative fullerene is C_{60}. CNTs and fullerenes have various uses based on their novel physicochemical characteristics, e.g., in the biomedical field of bioimaging and therapy [45.95]. However, typically they have poor solubility and are difficult to manipulate in solvents or physiological solutions like serum, which influences their toxicity, metabolism kinetics, and biomedical functions [45.23]. One in vitro study showed that SWCNTs, MWCNTs, and C_{60} exhibit quite different cytotoxicity to alveolar macrophages, although they all represent carbon materials (Fig. 45.8). The comparative toxicities illustrated the following sequence on a mass basis: SWCNTs > MWCNT (with diameters ranging from 10 to 20 nm) > C_{60} [45.96]. CNTs with different lengths contribute to obviously different lung fibrosis potential after inhalation exposure [45.28]. Further evidence showed that short-length MWCNTs not only had lower toxicity but also higher cellular uptake and more obvious removal ability compared to longer versions, which can result in the higher promotion of PC12 cell differentiation [45.98]. This means that the cytotoxicity and even the bioeffects of different carbon nanomaterials in vitro are highly dependent on their nanostructures. In more research from our lab, it was shown that anatase TiO_2 particles had a much higher toxicity than rutile ones on the central nervous system [45.99]. The shape of nanoparticles is another important factor that directly affects cellular uptake. For example, it was shown that longer Au nanorods had much less cellular uptake when compared to shorter

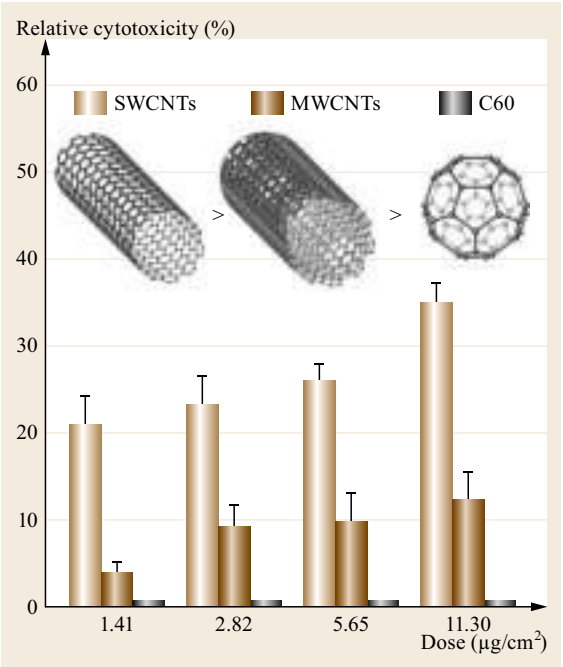


Fig. 45.8 The ranking of cytotoxicity of carbon NMs on a mass basis is: SWCNTs > MWCNTs > C_{60}. Reproduced with permission from reference [45.96]

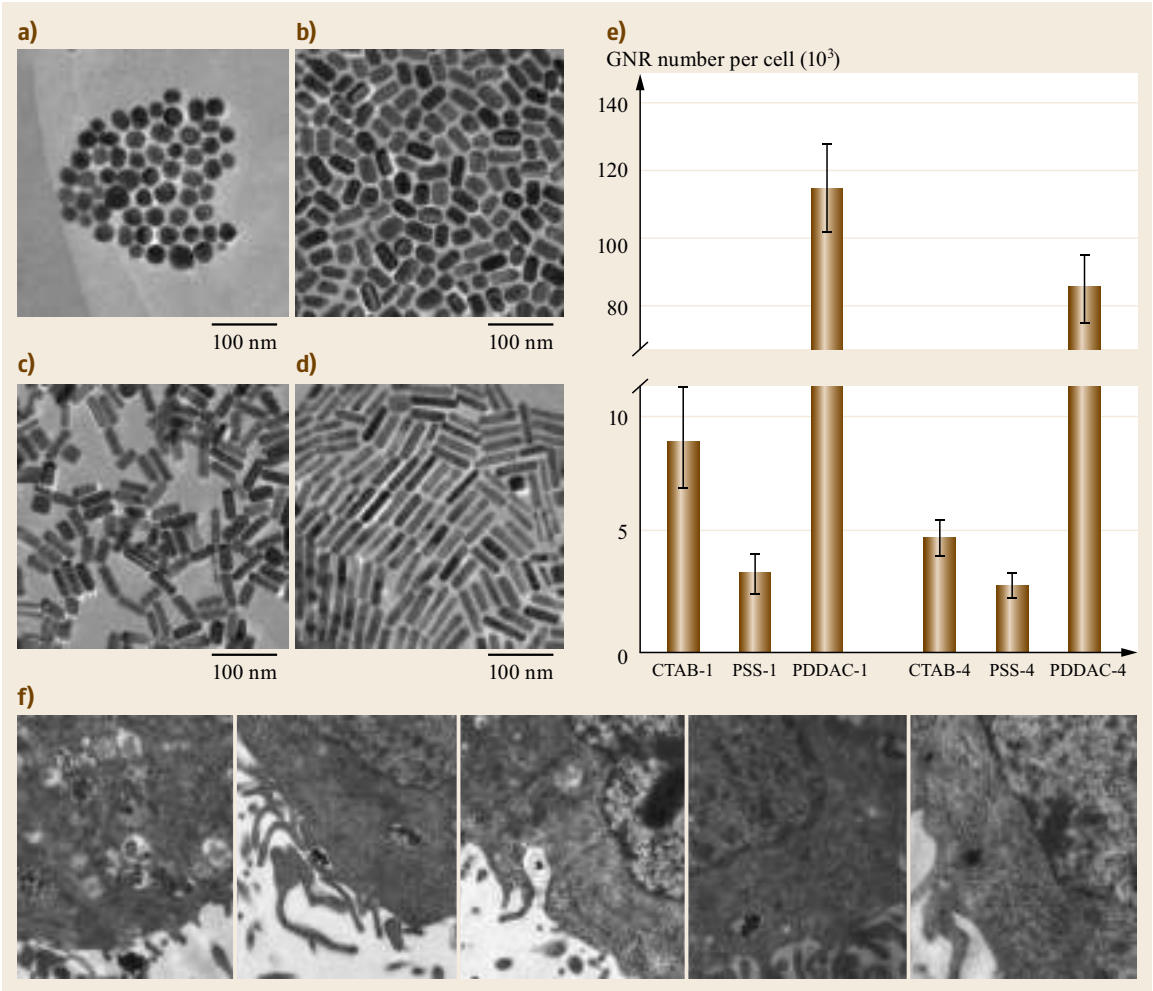


Fig. 45.9a–f Shape and surface-coating determine cellular uptake of Au nanoparticles in Michigan Cancer Foundation-7 (MCF-7) cells. Transmission electron microscope images of CTAB-coated Au nanorods with different aspect ratios: **(a)** CTAB-1, **(b)** CTAB-2, **(c)** CTAB-3, and **(d)** CTAB-4. **(e)** shows the shape- and surface-coating-dependent cellular uptake of Au nanorods coated by CTAB. **(f)** TEM images show the processes of cellular uptake. The Au nanorods form aggregates, enter into vesicles in this form, and move into lysosome. Reproduced from [45.97], with permission from Elsevier

ones, with similar particle surface charges used in the test (Fig. 45.9) [45.24, 97]. Spherical Au nanoparticles entered cells more easily than rod-shaped nanoparticles which may be because a longer membrane wrapping time is required for the rod-shaped particles [45.97].

45.3.3 Agglomeration and Aggregation

NPs have higher surface activity compared to larger sized versions of the same chemical composition. On this basis, the increased surface size also has a strong tendency for agglomeration/aggregation. An agglomeration is a collection of particles that are loosely bound together by relatively weak forces, including van der Waals forces, electrostatic forces, simple physical entanglement, and surface tension, with a resulting external surface area similar to the sum of the surface area of the individual components. Aggregated particles are a cohesive mass consisting of particulate subunits tightly bound by covalent or metallic bonds due to a surface reconstruction, often through melting or annealing of the surface, and often having an external surface area significantly smaller than the sum of the surface areas of the individual components. Agglomerates may be reversible under certain chemical/biological conditions whereas an aggregate

will not easily release primary particles under normal conditions [45.100]. Particulate matter in the air may come from release during the ENM production process. It usually represents a highly complex mixture of particles with different sizes. Field research showed that airborne NMs in occupational settings are mostly 200 to 400 and 2000 to 3000 nm in diameter [45.101, 102]. However, in some special environments such as metal processing, burning, photocopying or printing, and nanomaterial handling processes, the nanopeak of the aerosol is still extremely prominent [45.103–106]. The NMs in the air easily form agglomerates, which influences absorption in the organism after deposition in the inhalation system. The exact deposition dosages of differently sized particles could be estimated by the multiple path particle dosimetry (MPPD) model which was developed for estimating dose in the lower respiratory tract based on particle sizes [45.107, 108]. Toxic effects within the body are usually dominated by the NM agglomeration situation. It has been reported in mice that temporary organ injury occurs in lung and liver due to delayed clearance of agglomerated MWCNTs, while well-dispersed forms were easily eliminated from these organs [45.109]. Similarly, highly agglomerated SWCNTs cause more serious toxic effects on glial cells in both the peripheral nervous system and central nervous system compared with better dispersed SWCNTs [45.110]. In the estimation of the EHS impacts of NMs, the extent of agglomeration/aggregation should be taken into account and thoroughly investigated.

45.3.4 Chemical Composition, Purity, and Impurities

The chemical composition, also defined as elemental composition and chemical structure, is an intrinsic property of all materials. In the development of nanotechnology we see tremendously different NM chemical compositions. Roughly, there are four categories of NMs based on their chemical composition:

1. Carbon nanomaterials (e.g., carbon nanotubes, nanowires, nanocantilevers, graphene, and fullerenes)
2. Metallic and metallic oxide materials (e.g., gold or silver nanoparticles, magnetic nanoparticles, QDs, titanium dioxide, and iron oxide)
3. Silicon nanomaterials (silicon or silica nanoparticles)
4. Organic nanomaterials (e.g., DNA, polymers, polymeric micelles, liposomes, or nanoparticles prepared from polymers or lipids).

Some NMs are a hybrid of different compositions, with the various characteristics used to advantage in the final product. Recently, high content screening (HCS), a cell-based multiparametric image analysis technique, was adopted in the evaluation of eight different compositions of NMs [45.77]. It was found that Ag and ZnO NMs possessed much higher cytotoxicity to 16HBE cells than SiO_2, TiO_2, and CeO_2 based on the same mass level [45.77]. Impurities are inevitable in ENMs, and this needs to be considered when implementing the risk assessment. Raw nanotubes especially those in commercial products usually contain significant impurities, such as inert synthesis supports (silica, alumina) and metal catalysts (iron, cobalt, nickel), derived from the large-scale production procedures, postfabrication and postpurification treatments. Much evidence has shown that the impurities in NMs contribute greatly to increased toxicity through induction of oxidative stress [45.111–113].

45.3.5 Coating, Surface Modification, and Surface Charges

Coating and surface modifications of nanoparticles play an essential role in control of the physicochemical and surface properties of NMs. By employing these techniques, scientists can reduce material toxicity, increase solubility, enhance biocompatibility, and prevent aggregation of NMs in solution or in the air [45.114]. It was indicated that conventional hydrophobic fullerenes and the surface-modified hydrophilic ones have significant differences in terms of toxicity and biological functions [45.115, 116]. We investigated the influence of surface coatings on the cellular uptake and the cytotoxicity of gold nanorods, and found that the cellular uptake of Au NRs is highly dependent on the surface chemistry, as evidenced by poly (diallyldimethyl ammonium chloride) (PPDDAC)-coated Au nanorods exhibiting a much greater ability to be internalized by the cells compared to poly-sodium-*p*-styrenesulfate (PSS) and Cetyltrimethylammonium bromide (CTAB)-coated ones (Fig. 45.9e) [45.97, 117]. In nanotechnology, various nanosynthesis and nanofabrication methods have been applied to the NMs in order to obtain particular characteristics. However, the positive intention of NM modification may lead to unintended side effects, which in turn leads to alarm bells concerning possibly damaging health effects including poor biocompatibility, and acute and chronic toxicity. Surface charge, which can be manipulated by surface modification, is one of the most important factors in relation to the biological functions of nanoparticles. In contrast to neutral and negatively charged nanoparticles, positively charged particles can lead to the most efficient cell membrane penetration and cellular internalization because of their effective binding to negatively charged groups on the cell surface [45.118]. Fluorescent polystyrene (PS) particles

are good research objects to study the intercellular dynamics of allogenic materials in cell models. We investigated the effects of negatively charged carboxylated polystyrene (COOH-PS) and positively charged amino-modified polystyrene (NH_2-PS) nanoparticles with three different diameters (50, 100, and 500 nm) on cancer HeLa cells and normal NIH 3T3 cells during cell cycles. It was disclosed that the PS nanoparticles maintain a distance from the spindle and the chromosomes during the whole mitosis. They were never found to be associated with any components of the mitotic apparatus and no abnormal cell division was detected after their internalization. In addition, no abnormal daughter cells were detected. The data showed that the PS nanoparticles do not influence mitosis which is highly conserved in mammal cells in both normal and cancer cells [45.119].

Besides the aforementioned physicochemical characteristics, there are other factors, as listed in Table 45.2, that may have an influence on toxic impacts on EHS. Because of this situation the toxicity assessment of NMs is usually highly complex with sometimes even fragmentary and contradictory conclusions obtained from various sources [45.120]. Scientists have to face this difficulty and consider characterizing NMs in detail using state-of-the-art nanotechnologies.

45.4 Novel Techniques and Biomarker Development in Nanotoxicology

45.4.1 The Development of Biomarkers for Evaluation of EHS Impacts

The discovery of NM toxicity mechanisms could contribute a series of biomarkers for assessment of EHS impacts. Herein, biomarker is defined as key molecular or cellular events that have a direct link to health outcomes after exposure to NMs. It is highly feasible to fit developed biomarkers with in vitro cell model testing during risk assessment of EHP impacts due to the low costs, lack of an ethical animal abuse issue, and the high-efficiency of testing procedures [45.121]. This is in accordance with the strategy suggested by the US Environmental Protection Agency (EPA) and the National Research Council (NRC): cellular response or adverse outcome pathway (AOP)-based toxicity testing for adverse health effect evaluation should be the preferred toxicity testing strategy in the 21st century [45.122, 123]. Further, the in vitro testing techniques based on the biomakers of AOP generally provide more sensitive hazard identification than those detection methods using the endpoints of toxic effects when considering either the time or dose factors.

For the ENMs most investigated in nanotoxicology research, it is generally accepted that high NM surface area represents high ROS-generating capability which causes oxidative stress in biological systems after nano–bio interactions [45.124]. The high ROS level in the organism leads to the negative effects of the depletion of cellular antioxidants and accumulation of peroxidized lipids [45.25, 125]. In relation to these earlier inducing events, various markers at different tiers can be used in the assessment of EHS impacts of NMs along with the development of toxicity outcomes (Fig. 45.10).

Wang et al. [45.28] showed that MWCNTs could directly activate TGF-β/Smad2 signaling in fibroblasts or epithelial cells with the consequences of inducing the fibroblast-to-myofibroblast transition and EMT, both of which contribute to CNT-induced fibrotic responses. Therefore, it is suggested that TGF-β and the TGF-β receptor signaling system could be considered as potential biomarkers in the risk evaluation of CNTs. Attributed to the same carbon nanomaterial, graphene seems to induce toxicity through a similar mechanism. *Li* et al. [45.126] showed that the signaling pathways of both mitogen-activated protein kinase (MAPK) and TGF-β were activated and resulted in the downstream Bcl-2 protein family initiating mitochondrial-related apoptosis after incubation of the graphene in the RAW 264.7 macrophage cell line. Furthermore, p38 mitogen-activated protein kinase (p38 MAPK), extracellular signal-regulated kinase (ERK), and c-Jun N-terminal kinase (JNK) were significantly upregulated. Similar to the CNT exposure, the stimulated transcription activity of the Smad proteins verified that TGF-β was also found to be activated by graphene. As a result, Bim and Bax were significantly upregulated with the consequences of causing permeabilization of the mitochondrial outer membrane and the release of mitochondrial proapoptotic factors into the cytosol. These results indicate that graphene and CNTs could induce ROS production and trigger the mitochondrial apoptotic pathway, thus leading to apoptosis.

The endoplasmic reticulum (ER) stress and interruption of the aforementioned mitochondria metabolic signaling pathways has been fully investigated in the evaluation of EHS impacts [45.56, 81, 127–131]. ER stress, also known as unfolded protein response (UPR), is an important cellular self-protection mechanism that could be used as an early biomarker for evaluating the toxicity of NMs [45.81]. Toxicity initiates as the adaptation response and ends with toxicity formed after an

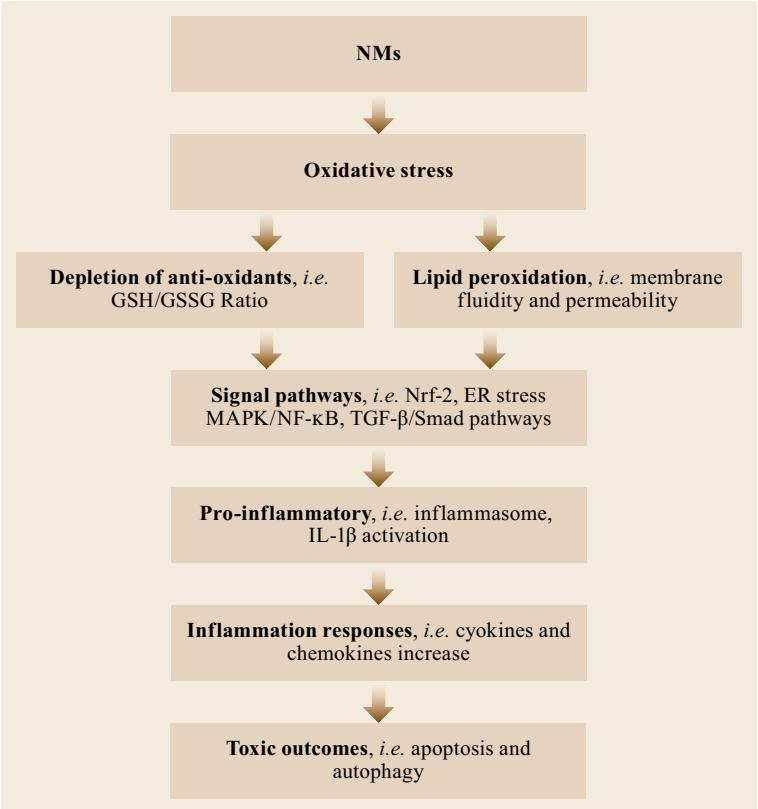


Fig. 45.10 Toxicity mechanisms of NMs and proposed biomarkers in assessment of EHS impacts

exposure dosage higher than the physiological threshold of the cellular self-protection capacity. The ER stress responses contain three dominant stress sensors that include inositol-requiring protein 1 (IRE1), PKR-like endoplasmic reticulum kinase (PERK), and activating transcription factor-6 (ATF-6). The ER-chaperone protein 78 kDa glucose-regulated protein (GRP78), or called binding immunoglobulin protein (BiP), binds these three stress sensors in the normal state, and in its active state releases them when stress occurs. IRE1, which functions as an endoribonuclease, initiates the splicing of x-box binding protein 1 (xbp-1) mRNA after activation. Then, the newly produced spliced form of *xbp-1* (*xbp-1s*) induces the transcription of chaperone protein-encoding genes for self-protection functions (Fig. 45.11). The *xbp-1* splicing and induction of BiP and C/EBP homologous protein (CHOP) have been identified as the specific biomarkers of ER stress responses [45.81, 132]. It is worth noting that ER stress, as a sensitive earlier biomarker, could be used to define the lowest observable effect concentrations (LOEC) in cell and animal models for predesigned safe applications of ENMs [45.77].

High-throughput screening (HTS) techniques and modeling based on QSAR should be developed for analysis of potential EHS hazards from NMs especially when handling NMs with exceedingly variable physicochemical characteristics [45.86, 133]. HTS techniques hold great potential, enabling faster and more reproducible toxicity testing for basic NMs as well as other nanorelated EHS samples. QSAR generally has a close relation to the application of various HTS methods. Therefore, researchers can predict a wide range of properties, including biological activities and toxicities, for new materials using QSAR models; this process depends entirely on the initial input data generated from HTS detection methods. Thus, this predictive toxicological approach will finally aid the design or optimization of NPs with commercially useful properties that are also safe to humans and the environment. Recently, *Le* et al. [45.86] developed QSAR models linking nanoparticle properties to cell viability, membrane integrity, and oxidative stress based on the testing data of 45 types of ZnO nanoparticles with varying particle sizes, aspect ratios, doping types, doping concentrations, and surface coatings. It was found that influencing factors such as the concentrations of NPs the cells are exposed to, the type of surface coating, the nature and extent of doping, and the aspect ratio of the particles provide significant contributions to the cellular toxicity of the NPs. These

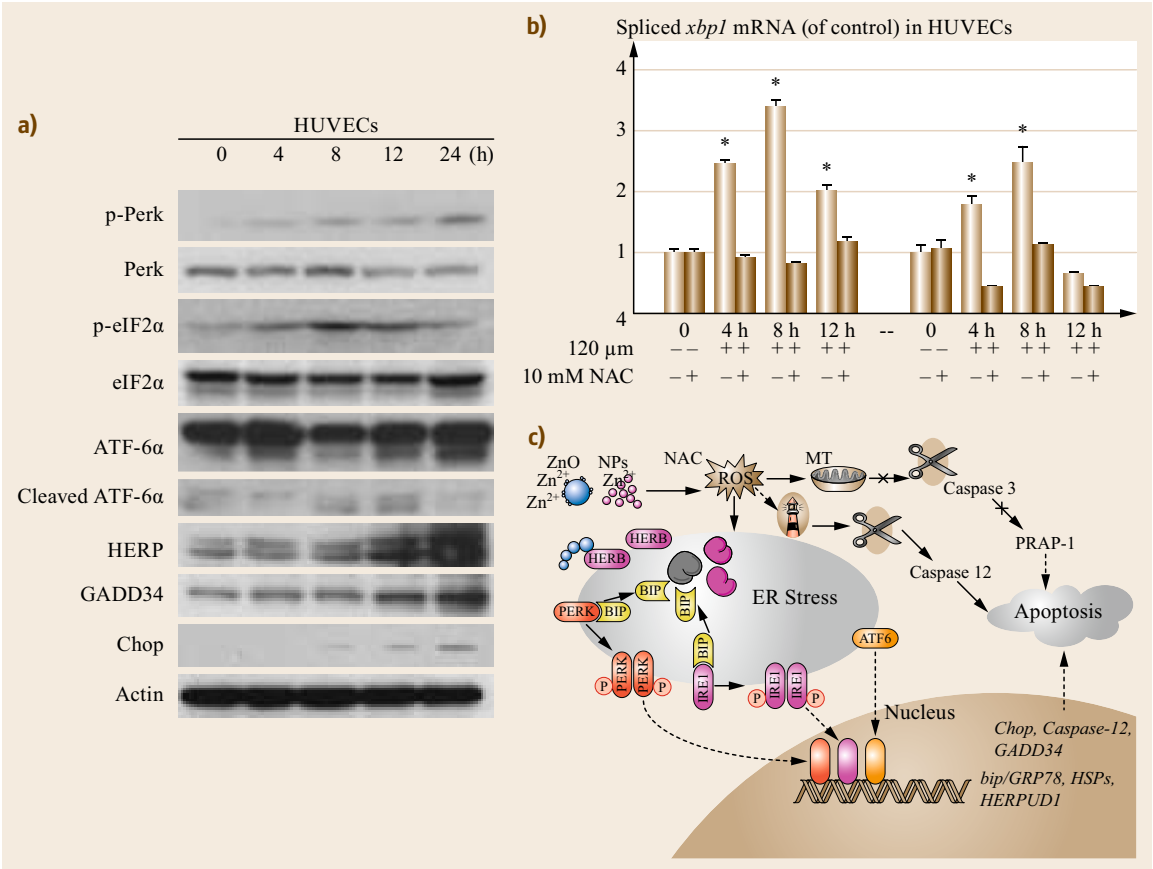


Fig. 45.11 Endoplasmic reticulum (ER) stress is an earlier biomarker for assessment of EHS impacts. **(a)** Western blotting analysis of ER stress-related proteins after exposure in human umbilical vein endothelial cells (HUVEC) at 240 μM. Reprinted with permission from [45.81] copyright 2014 American Chemical Society. **(b)** The PCR detection of spliced *xbp-1* was used for the detection of ER stress response in HUVECs after exposures to ZnO NP and zinc ions. The antioxidant (NAC, *N*-acetyl-cysteine) was used as antioxidant to abolish the ROS accumulation in cells. **(c)** The proposed mechanism for the activation of ER stress pathway after exposure to ZnO NPs

valuable strategies and techniques will promote the development of novel pathway mechanism-based in vitro approaches for assessment of EHS impact of NMs and risk assessment of influences on humans or contamination of ecosystems.

45.4.2 Novel Techniques Used for ADME Study in Nanotoxicology

After they are released and distributed, NMs generally appear as trace elements in biological and environmental science samples. The knowledge about ADME of NMs has high relations to the following evaluations of their EHS impacts. Nuclear analytical techniques (NATs) play important roles in facing the challenges in nanotoxicology [45.134–136]. NATs are a collection of techniques that deal with nuclear excitations, nuclear reactions, electron inner shell excitations, and/or radioactive decay. Typically, they have the advantages of absolute quantification, high sensitivity, excellent accuracy and precision, low matrix effects, and non-destructiveness in analytical applications [45.137, 138]. NATs mostly focus on the analysis of parameters in the samples e.g., nuclear mass, spin and magnetic moment, and excited states. Their application in nanotoxicology have been well summarized in the literature [45.134]. At the molecular level, synchrotron radiation circular dichroism spectroscopy (SRCD), x-ray absorption fine structure (XAFS), circular dichroism spectroscopy (CS), and AFM can be applied to characterize nanomaterials and analyze bio–nano interactions (Fig. 45.12). The cellular visualization of nanoparticles and the study of their transformation within the cell can be achieved through confocal laser scanning mi-

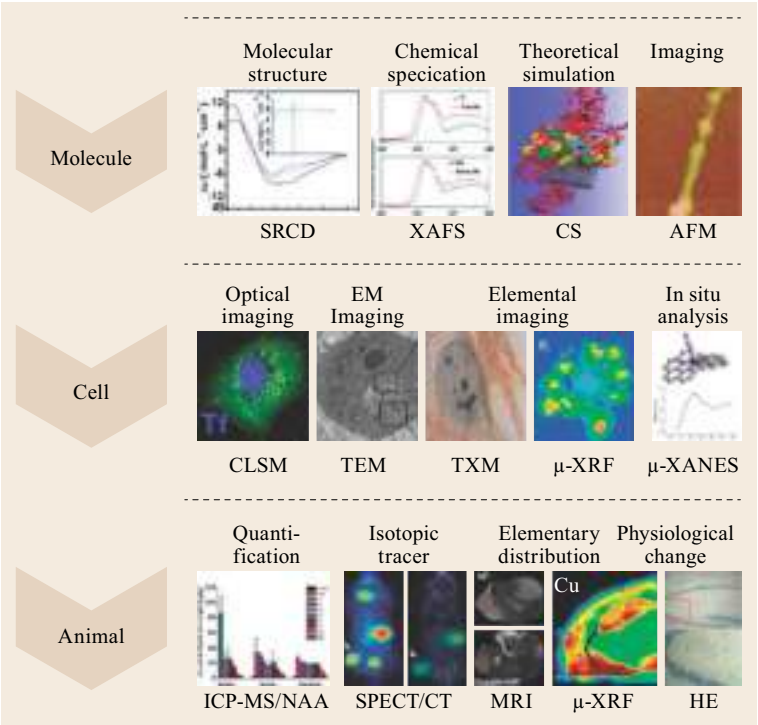


Fig. 45.12 Systematic platform using integrated advanced nuclear techniques for nanotoxicology. Adapted from [45.134] with permission from the Royal Society of Chemistry

croscopy (CLSM), TEM, scanning transmission x-ray microscopy (TXM), x-ray fluorescence (μ-XRF), and x-ray absorption near-edge structure (μ-XANES). Several approaches can execute in vivo ADME analysis: neutron activation analysis (NAA), ICP-MS (inductively coupled plasma mass spectrometry), inductively coupled plasma optical emission spectroscopy (ICP-OES), isotopic tracing (IT), single-photon emission computed tomography (SPECT), positron emission tomography (PET), x-ray computed tomography (CT), magnetic resonance imaging (MRI), fluorescence imaging, and photoacoustic imaging. The NAA, ICP-MS, ICP-OES, and IT techniques require the sacrificing of animals to obtain quantitative information on the NMs. SPECT, PET, x-ray CT, and MRI can show the whole body distribution of nanoparticles directly, while their quantification capabilities are limited. Fluorescence and photoacoustic-based methods have their own unique advantages in some scenarios of in vivo imaging, but the penetration depths are limited due to the limited penetration capability of light and sound compared to x-rays or γ-rays [45.139]. In recent research using the XANES technique, it was disclosed that disulfide bonds in BSA change into Au–S bonds quantitatively on the binding sites of BSA and Au nanorods. At the same time, a stable plane was found for gold–sulfur interaction supported by molecular dynamic simulation [45.140]. Further, the XANES technique has the ability to dissect the chemical transformation of metal NMs in biological systems. For example, it is known that silver NPs can be internalized by cells and degraded inside lysosomes, but how the cytotoxicity developed during the degradation process is unknown. Silver k-edge XANES clearly indicated that intracellular silver can be transformed from elemental Ag to a silver-sulfur (Ag−S−) form which played the dominant role in cytotoxicity [45.141]. XRF imaging is a powerful tool for elemental imaging in biological samples of cells and tissues. It is feasible to combine XRF imaging with XANES for analysis of the same sample. For example, this combined strategy has been used to study the distribution of Se when the worm *C. elegans* engulfs CdSe QDs. XRF imaging showed the locations of accumulation and distribution of Se, while Se k-edge XANES indicated the change in Se chemical forms in different body regions. This combined analysis can provide more accurate information on NMs than the more generally used fluorescence-based optical imaging [45.63]. There are two kinds of TXM, soft x-ray TXM and hard x-ray TXM (Nano-CT). The former can distinguish the NMs and organelles and is suitable for imaging a local region inside a cell due to the small x-ray penetration depth ($\approx 5\,\mu m$) [45.142]. While the latter, Nano-CT, can provide a much wider imaging field in a single cell ($\approx 2-50\,\mu m$) which could be used for three-dimensional (3-D) imaging of the con-

Table 45.3 Nuclear analytical and related techniques for physicochemical characterization, chemical element imaging, quantification, and speciation in nanotoxicological studies. Reproduced with permission from reference [45.134]

Techniques	Exciting radiation	Particle emission	Depth information	Spatial resolution	Detection limit	Chemical structure	Quantitative analysis	Imaging	Materials	Availability	Matrix effects
NAA	Thermal neutrons	Alpha, electrons, neutrons, and x-ray photons	None	None	0.001–0.1 ppm	No	Yes, 2–10%	No	Solids and liquids	Very rare	Less
ICP-MS	Plasma	Ions	None	None	0.00001–10 ppm	Little	Yes, high precision with isotope dilution	No, imaging by LA	Solution	Common	Yes, severe
SIMS	Ions, atoms for FABMS	Secondary ions	3–10 nm	0.02 μm best, 1–5 μm typical	0.1–10 ppm	Some chemical bonds	Yes, to 25% relative error	Yes	Any solid, some liquids	Relatively uncommon	Yes
XRD	x-ray tube or synchrotron x-rays	Diffracted x-rays	None	50 nm–10 μm synchrotron x-rays	1–5% mixtures	Yes, species identity and structure	Yes, to 5–10%	No	Crystalline solids and polymers	Common	Yes
XAS	Synchrotron and some rotating anode x-ray sources	Transmitted or characteristic x-rays	Yes	50 nm synchrotron x-rays	100 ppm	Yes	No	No	Solids and nonvolatile liquids	Rare, at synchrotron facilities	Yes

Table 45.3 (continued)

Techniques	Exciting radiation	Particle emission	Depth information	Spatial resolution	Detection limit	Chemical structure	Quantitative analysis	Imaging	Materials	Availability	Matrix effects
XRF	x-ray tube, synchrotron x-rays	Characteristic x-rays	None	50 nm–10 μm synchrotron x-rays	0.1—10 ppm (≈ synchrotron mg kg^{-1}) x-ray	No	Yes, high precision	Yes	Solids and nonvolatile liquids	XRF common, XRF Mapping only at synchrotron facilities	Yes
EDX	Electron	Characteristic x-ray photons	0.5—5 μm, depends on matrix	2 nm–5 μm, depends on matrix	100–1000 ppm	No	Yes, relative error	No	Any solid	Moderately common	Yes
EELS	Electron	Characteristic electron energy loss		1 nm	1000 ppm	Yes	Yes	No	Any solid	Moderately common	Yes
PIXE	Protons	Characteristic x-ray photons	None	2 μm–20 mm	0.1—10 ppm	No	Yes, 5%	Yes	Any solid	Rare	Yes
Isotopic tracing		Radioactive: α, β, γ-rays and stable isotopes	No		10^{-4}–10^{-8} g	No	Yes	Yes	Any solid	Common	No
PET	Positron	γ-ray	Yes	1 mm	100 ppt	No	Yes	Yes	Any solid	Common	No
SPECT	Single-photon	γ-ray	Yes	1 mm	100 ppt	No	Yes	Yes	Any solid	Common	No

struction of metal NMs [45.141]. Soft x-ray TXM also has its own powerful applications for the analysis of specific elements like Ce in CeO_2 NPs or Gd in Gd-metallofullerenol [45.143, 144]. The characteristics of the whole list of NATs and related techniques are provided in Table 45.3.

45.5 Conclusion and Perspectives

EHS impacts related to nanotechnology were investigated in this chapter. It is obvious that the presence and release of nanoparticles into the environment has important implications for human health and the environment. Scientists should not only focus on the beneficial effects of NPs, but also need to be careful in regard to possible adverse impacts on EHS. Particle attributes such as size, agglomeration or aggregation, composition, structure, coating, and purity are important for the determination of potential impacts on EHS, although they are generally quite complicated. State-of-the-art analysis methods and techniques should be used and developed for the gradually emerging challenges. Such techniques and biomarkers will have meaningful utilization in the protection of human populations, especially in the case of occupational exposure, and in regard to the total ecosystem. As the mechanisms of nanotoxicity and ecotoxicity are still poorly understood, the study of such EHS impacts through NMs definitely needs the input of greater resources. Both researchers and authorities should strive for a sound balance between the development of nanotechnology and the scientific evaluation of EHS impacts in order to achieve a sustainable and harmonious development of nanotechnology.

Acknowledgments. This work was supported by the Ministry of Science and Technology of China (2016YFA0201600), the National Natural Science Foundation of China (21477029, 21320102003, 21277037, 21403043, and 21277080), the Chinese Academy of Sciences (XDA09040400), the Beijing Natural Science Foundation (No. 2152037), Major Project of the National Social Science Fund (Grant No. 12&ZD117) "Ethical issues of high-tech", and the National Science Fund for Distinguished Young Scholars (11425520).

References

45.1 M.H. Hassan: Nanotechnology. Small things and big changes in the developing world, Science **309**, 65–66 (2005)

45.2 K. Donaldson, V. Stone, C.L. Tran, W. Kreyling, P.J. Borm: Nanotoxicology, Occup. Environ. Med. **61**, 727–728 (2004)

45.3 G. Oberdörster, E. Oberdörster, J. Oberdörster: Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles, Environ. Health Perspect. **113**, 823–839 (2005)

45.4 G. Oberdörster: Safety assessment for nanotechnology and nanomedicine: Concepts of nanotoxicology, J. Intern. Med. **267**, 89–105 (2010)

45.5 National Research Council: Health and safety aspects of engineered nanomaterials committee to develop a research strategy for environmental. In: *A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials*, ed. by the Committee to Develop a Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials, chaired by J.M. Samet (National Academies, Washington DC 2012)

45.6 C.O. Hendren, X. Mesnard, J. Droge, M.R. Wiesner: Estimating production data for five engineered nanomaterials as a basis for exposure assessment, Environ. Sci. Technol. **45**, 2562–2569 (2011)

45.7 K. Schmid, M. Riediker: Use of nanoparticles in Swiss industry: A targeted survey, Environ. Sci. Technol. **42**, 2253–2260 (2008)

45.8 F. Piccinno, F. Gottschalk, S. Seeger, B. Nowack: Industrial production quantities and uses of ten engineered nanomaterials in Europe and the world, J. Nanopart. Res. **14**, 1–11 (2012)

45.9 N. Li, S. Georas, N. Alexis, P. Fritz, T. Xia, M.A. Williams, E. Horner, A. Nel: A work group report on ultrafine particles (American Academy of Allergy, Asthma & Immunology): Why ambient ultrafine and engineered nanoparticles should receive special attention for possible adverse health outcomes in human subjects, J. Allergy Clin. Immunol. **138**, 386–396 (2016)

45.10 R. Chen, C. Chen: Scenarios in the workplace and risk assessment of carbon nanomaterials. In: *Biomedical Applications and Toxicology of Carbon Nanomaterials*, ed. by C. Chen, H. Wang (Wiley, Weinheim 2016)

45.11 F. Gottschalk, C. Lassen, J. Kjoelholt, F. Christensen, B. Nowack: Modeling flows and concentrations of nine engineered nanomaterials in the Danish environment, Int. J. Env. Res. Public Health **12**, 5581–5602 (2015)

45.12 T.Y. Sun, F. Gottschalk, K. Hungerbühler, B. Nowack: Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials, Environ. Pollut. **185**, 69–76 (2014)

45.13 Y.S. Tian, F. Gottschalk, K. Hungerbühler, B. Nowack: Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials, Environ. Pollut. **185**, 69–76 (2014)

45.14 T.Y. Sun, N.A. Bornhoft, K. Hungerbühler, B. Nowack: Dynamic probabilistic modeling of environmental emissions of engineered nanomaterials, Environ. Sci. Technol. **50**, 4701–4711 (2016)

45.15 M. Miseljic, S.I. Olsen: Life-cycle assessment of engineered nanomaterials: A literature review of assessment status, J. Nanopart. Res. **16**, 33 (2014)

45.16 H. Sengul, T.L. Theis, S. Ghosh: Toward sustainable nanoproducts: An overview of nanomanufacturing methods, J. Ind. Ecol. **12**, 329–359 (2008)

45.17 R. Dhingra, S. Naidu, G. Upreti, R. Sawhney: Sustainable nanotechnology: Through green methods and life-cycle thinking, Sustainability **2**, 3323–3338 (2010)

45.18 Y. Ding, T.A. Kuhlbusch, M. Van Tongeren, A.S. Jimenez, I. Tuinman, R. Chen, I.L. Alvarez, U. Mikolajczyk, C. Nickel, J. Meyer, H. Kaminski, W. Wohlleben, B. Stahlmecke, S. Clavaguera, M. Riediker: Airborne engineered nanomaterials in the workplace – A review of release and worker exposure during nanomaterial production and handling processes, J. Hazard. Mater. **322**, 17–28 (2017)

45.19 Y. Zhao, G. Xing, Z. Chai: Nanotoxicology: Are carbon nanotubes safe?, Nat. Nanotechnol. **3**, 191–192 (2008)

45.20 J.C. Bonner: Carbon nanotubes as delivery systems for respiratory disease: Do the dangers outweigh the potential benefits?, Expert. Rev. Respir. Med. **5**, 779–787 (2011)

45.21 C.A. Poland, R. Duffin, I. Kinloch, A. Maynard, W.A. Wallace, A. Seaton, V. Stone, S. Brown, W. Macnee, K. Donaldson: Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study, Nat. Nanotechnol. **3**, 423–428 (2008)

45.22 P. Wang, X. Nie, Y. Wang, Y. Li, C. Ge, L. Zhang, L. Wang, R. Bai, Z. Chen, Y. Zhao, C. Chen: Multiwall carbon nanotubes mediate macrophage activation and promote pulmonary fibrosis through TGF-beta/Smad signaling pathway, Small **9**, 3799–3811 (2013)

45.23 Y. Liu, Y. Zhao, B. Sun, C. Chen: Understanding the toxicity of carbon nanotubes, Acc. Chem. Res. **46**, 702–713 (2013)

45.24 F. Zhao, Y. Zhao, Y. Liu, X. Chang, C. Chen, Y. Zhao: Cellular uptake, intracellular trafficking and cytotoxicity of nanomaterials, Small **7**, 1322–1337 (2011)

45.25 P. Moller, D.V. Christophersen, D.M. Jensen, A. Kermanizadeh, M. Roursgaard, N.R. Jacobsen, J.G. Hemmingsen, P.H. Danielsen, Y. Cao, K. Jantzen, H. Klingberg, L.G. Hersoug, S. Loft: Role of oxidative stress in carbon nanotube-generated health effects, Arch. Toxicol. **88**, 1939–1964 (2014)

45.26 R. Alshehri, A.M. Ilyas, A. Hasan, A. Arnaout, F. Ahmed, A. Memic: Carbon nanotubes in biomedical applications: Factors, mechanisms and remedies of toxicity, J. Med. Chem. **59**, 8149–8167 (2016)

45.27 Y.Y. Guo, J. Zhang, Y.F. Zheng, J. Yang, X.Q. Zhu: Cytotoxic and genotoxic effects of multi-wall carbon nanotubes on human umbilical vein endothelial cells in vitro, Mutat. Res. **721**, 184–191 (2011)

45.28 P. Wang, Y. Wang, X. Nie, C. Braini, R. Bai, C. Chen: Multiwall carbon nanotubes directly promote fibroblast-myofibroblast and epithelial-mesenchymal transitions through the activation of the TGF-beta/Smad signaling pathway, Small **11**, 446–455 (2015)

45.29 R.R. Mercer, J.F. Scabilloni, A.F. Hubbs, L. Wang, L.A. Battelli, W. McKinney, V. Castranova, D.W. Porter: Extrapulmonary transport of MWCNT following inhalation exposure, Part. Fibre. Toxicol. **10**, 38 (2013)

45.30 K. Donaldson, C.A. Poland, F.A. Murphy, M. MacFarlane, T. Chernova, A. Schinwald: Pulmonary toxicity of carbon nanotubes and asbestos-similarities and differences, Adv. Drug Deliv. Rev. **65**, 2078–2086 (2013)

45.31 M. Sharma, J. Nikota, S. Halappanavar, V. Castranova, B. Rothen-Rutishauser, A.J. Clippinger: Predicting pulmonary fibrosis in humans after exposure to multi-walled carbon nanotubes (MWCNTs), Arch. Toxicol. **90**, 1605–1622 (2016)

45.32 C. Ge, L. Meng, L. Xu, R. Bai, J. Du, L. Zhang, Y. Li, Y. Chang, Y. Zhao, C. Chen: Acute pulmonary and moderate cardiovascular responses of spontaneously hypertensive rats after exposure to single-wall carbon nanotubes, Nanotoxicology **6**, 526–542 (2012)

45.33 I. Huizar, A. Malur, Y.A. Midgette, C. Kukoly, P. Chen, P.C. Ke, R. Podila, A.M. Rao, C.J. Wingard, L. Dobbs, B.P. Barna, M.S. Kavuru, M.J. Thomassen: Novel murine model of chronic granulomatous lung inflammation elicited by carbon nanotubes, Am. J. Respir. Cell Mol. Biol. **45**, 858–866 (2011)

45.34 C.C. Chou, H.Y. Hsiao, Q.S. Hong, C.H. Chen, Y.W. Peng, H.W. Chen, P.C. Yang: Single-walled carbon nanotubes can induce pulmonary injury in mouse model, Nano Lett. **8**, 437–445 (2008)

45.35 R. Chen, L. Zhang, C. Ge, M.T. Tseng, R. Bai, Y. Qu, C. Beer, H. Autrup, C. Chen: Subchronic toxicity and cardiovascular responses in spontaneously hypertensive rats after exposure to multiwalled carbon nanotubes by intratracheal instillation, Chem. Res. Toxicol. **28**, 440–450 (2015)

45.36 T. Ingle, E. Dervishi, A.R. Biris, T. Mustafa, R.A. Buchanan, A.S. Biris: Raman spectroscopy analysis and mapping the biodistribution of inhaled carbon nanotubes in the lungs and blood of mice, J. Appl. Toxicol. **33**, 1044–1052 (2013)

45.37 A.R. Reddy, D.R. Krishna, Y.N. Reddy, V. Himabindu: Translocation and extra pul-

monary toxicities of multi wall carbon nanotubes in rats, Toxicol. Mech. Methods **20**, 267–272 (2010)

45.38 C. Ge, J. Du, L. Zhao, L. Wang, Y. Liu, D. Li, Y. Yang, R. Zhou, Y. Zhao, Z. Chai, C. Chen: Binding of blood proteins to carbon nanotubes reduces cytotoxicity, Proc. Natl. Acad. Sci. USA **108**, 16968–16973 (2011)

45.39 X. Jiang, R. Foldbjerg, T. Miclaus, L. Wang, R. Singh, Y. Hayashi, D. Sutherland, C. Chen, H. Autrup, C. Beer: Multi-platform genotoxicity analysis of silver nanoparticles in the model cell line CHO-K1, Toxicol. Lett. **222**, 55–63 (2013)

45.40 C. Chen, H. Wang: *Biomedical Applications and Toxicology of Carbon Nanomaterials* (Wiley, Weinheim 2016)

45.41 S. Chernousova, M. Epple: Silver as antibacterial agent: Ion, nanoparticle and metal, Angew. Chem. Int. Ed. Engl. **52**, 1636–1653 (2013)

45.42 K. Chaloupka, Y. Malam, A.M. Seifalian: Nanosilver as a new generation of nanoproduct in biomedical applications, Trends Biotechnol **28**, 580–588 (2010)

45.43 C. Lorenz, L. Windler, N. von Goetz, R.P. Lehmann, M. Schuppler, K. Hungerbuhler, M. Heuberger, B. Nowack: Characterization of silver release from commercially available functional (nano) textiles, Chemosphere **89**, 817–824 (2012)

45.44 L. Geranio, M. Heuberger, B. Nowack: The behavior of silver nanotextiles during washing, Environ. Sci. Technol. **43**, 8113–8118 (2009)

45.45 T.M. Benn, P. Westerhoff: Nanoparticle silver released into water from commercially available sock fabrics, Environ. Sci. Technol. **42**, 4133–4139 (2008)

45.46 D.M. Mitrano, E. Rimmele, A. Wichser, R. Erni, M. Height, B. Nowack: Presence of nanoparticles in wash water from conventional silver and nano-silver textiles, ACS Nano **8**, 7208–7219 (2014)

45.47 R. Kaegi, A. Voegelin, C. Ort, B. Sinnet, B. Thalmann, J. Krismer, H. Hagendorfer, M. Elumelu, E. Mueller: Fate and transformation of silver nanoparticles in urban wastewater systems, Water Res **47**, 3866–3877 (2013)

45.48 R. Kaegi, A. Voegelin, B. Sinnet, S. Zuleeg, H. Hagendorfer, M. Burkhardt, H. Siegrist: Behavior of metallic silver nanoparticles in a pilot wastewater treatment plant, Environ. Sci. Technol. **45**, 3902–3908 (2011)

45.49 Y. Yin, J. Liu, G. Jiang: Sunlight-induced reduction of ionic Ag and Au to metallic nanoparticles by dissolved organic matter, ACS Nano **6**, 7910–7919 (2012)

45.50 N. Akaighe, R.I. Maccuspie, D.A. Navarro, D.S. Aga, S. Banerjee, M. Sohn, V.K. Sharma: Humic acid-induced silver nanoparticle formation under environmentally relevant conditions, Environ. Sci. Technol. **45**, 3895–3901 (2011)

45.51 N.F. Adegboyega, V.K. Sharma, K. Siskova, R. Zboril, M. Sohn, B.J. Schultz, S. Banerjee: Interactions of aqueous Ag^{+} with fulvic acids: Mechanisms of silver nanoparticle formation and investigation of stability, Environ. Sci. Technol. **47**, 757–764 (2013)

45.52 D. Lu, Q. Liu, T. Zhang, Y. Cai, Y. Yin, G. Jiang: Stable silver isotope fractionation in the natural transformation process of silver nanoparticles, Nat. Nanotechnol. **11**, 682–686 (2016)

45.53 Y. Li, J.A. Bhalli, W. Ding, J. Yan, M.G. Pearce, R. Sadiq, C.K. Cunningham, M.Y. Jones, W.A. Monroe, P.C. Howard, T. Zhou, T. Chen: Cytotoxicity and genotoxicity assessment of silver nanoparticles in mouse, Nanotoxicology **8**(Suppl 1), 36–45 (2014)

45.54 Z. Wang, G. Qu, L. Su, L. Wang, Z. Yang, J. Jiang, S. Liu, G. Jiang: Evaluation of the biological fate and the transport through biological barriers of nanosilver in mice, Curr. Pharm. Des. **19**, 6691–6697 (2013)

45.55 Y. Zhang, Y. Zhang, G. Hong, W. He, K. Zhou, K. Yang, F. Li, G. Chen, Z. Liu, H. Dai, Q. Wang: Biodistribution, pharmacokinetics and toxicology of Ag_2S near-infrared quantum dots in mice, Biomaterials **34**, 3639–3646 (2013)

45.56 R. Chen, L. Zhao, R. Bai, Y. Liu, L.P. Han, Z.F. Xu, F. Chen, H. Autrup, D.X. Long, C.Y. Chen: Silver nanoparticles induced oxidative and endoplasmic reticulum stresses in mouse tissues: Implications for the development of acute toxicity after intravenous administration, Toxicol. Res. **5**, 602–608 (2016)

45.57 A. Chrastina, J.E. Schnitzer: Iodine-125 radiolabeling of silver nanoparticles for in vivo SPECT imaging, Int. J. Nanomed. **5**, 653–659 (2010)

45.58 D.P. Lankveld, A.G. Oomen, P. Krystek, A. Neigh, A. Troost-de Jong, C.W. Noorlander, J.C. Van Eijkeren, R.E. Geertsma, W.H. De Jong: The kinetics of the tissue distribution of silver nanoparticles of different sizes, Biomaterials **31**, 8350–8361 (2010)

45.59 Y. Xue, S. Zhang, Y. Huang, T. Zhang, X. Liu, Y. Hu, Z. Zhang, M. Tang: Acute toxic effects and gender-related biokinetics of silver nanoparticles following an intravenous injection in mice, J. Appl. Toxicol. **32**, 890–899 (2012)

45.60 Y. Li, T. Qin, T. Ingle, J. Yan, W. He, J.J. Yin, T. Chen: Differential genotoxicity mechanisms of silver nanoparticles and silver ions, Arch. Toxicol. **91**, 509–519 (2017)

45.61 S. Kim, D.Y. Ryu: Silver nanoparticle-induced oxidative stress, genotoxicity and apoptosis in cultured cells and animal tissues, J. Appl. Toxicol. **33**, 78–89 (2013)

45.62 L. Wang, J. Li, J. Pan, X. Jiang, Y. Ji, Y. Li, Y. Qu, Y. Zhao, X. Wu, C. Chen: Revealing the binding structure of the protein corona on gold nanorods using synchrotron radiation-based techniques: understanding the reduced damage in cell membranes, J. Am. Chem. Soc. **135**, 17359–17368 (2013)

45.63 Y. Qu, W. Li, Y. Zhou, X. Liu, L. Zhang, L. Wang, Y.F. Li, A. Iida, Z. Tang, Y. Zhao, Z. Chai, C. Chen: Full assessment of fate and physiological behavior of quantum dots utilizing Caenorhabditis elegans as a model organism, Nano Lett **11**, 3174–3183 (2011)

45.64 X. Jiang, T. Miclaus, L. Wang, R. Foldbjerg, D.S. Sutherland, H. Autrup, C. Chen, C. Beer: Fast intracellular dissolution and persistent cellular uptake of silver nanoparticles in CHO-K1 cells: Im-

plication for cytotoxicity, Nanotoxicology **9**, 181–189 (2015)
45.65 C. Pang, A. Brunelli, C. Zhu, D. Hristozov, Y. Liu, E. Semenzin, W. Wang, W. Tao, J. Liang, A. Marcomini, C. Chen, B. Zhao: Demonstrating approaches to chemically modify the surface of Ag nanoparticles in order to influence their cytotoxicity and biodistribution after single dose acute intravenous administration, Nanotoxicology **10**, 129–139 (2016)
45.66 M. Ahamed, M. Karns, M. Goodson, J. Rowe, S.M. Hussain, J.J. Schlager, Y. Hong: DNA damage response to different surface chemistry of silver nanoparticles in mammalian cells, Toxicol. Appl. Pharmacol. **233**, 404–410 (2008)
45.67 A.K. Suresh, D.A. Pelletier, W. Wang, J.L. Morrell-falvey, B. Gu, M.J. Doktycz: Cytotoxicity induced by engineered silver nanocrystallites is dependent on surface coatings and cell types, Langmuir **28**, 2727–2735 (2012)
45.68 X. Yang, A.P. Gondikas, S.M. Marinakos, M. Auffan, J. Liu, H. Hsu-Kim, J.N. Meyer: Mechanism of silver nanoparticle toxicity is dependent on dissolved silver and surface coating in Caenorhabditis elegans, Environ. Sci. Technol. **46**, 1119–1127 (2012)
45.69 S.R. Saptarshi, A. Duschl, A.L. Lopata: Biological reactivity of zinc oxide nanoparticles with mammalian test systems: An overview, Nanomedicine **10**, 2075–2092 (2015)
45.70 M.J. Osmond, M.J. McCall: Zinc oxide nanoparticles in modern sunscreens: An analysis of potential exposure and hazard, Nanotoxicology **4**, 15–41 (2010)
45.71 B. Gulson, M. McCall, M. Korsch, L. Gomez, P. Casey, Y. Oytam, A. Taylor, M. McCulloch, J. Trotter, L. Kinsley, G. Greenoak: Small amounts of zinc from zinc oxide particles in sunscreens applied outdoors are absorbed through human skin, Toxicol. Sci. **118**, 140–149 (2010)
45.72 T.G. Smijs, S. Pavel: Titanium dioxide and zinc oxide nanoparticles in sunscreens: Focus on their safety and effectiveness, Nanotechnol. Sci. Appl. **4**, 95–112 (2011)
45.73 P. Kelleher, K. Pacheco, L.S. Newman: Inorganic dust pneumonias: The metal-related parenchymal disorders, Environ. Health Perspect. **108**, 685–696 (2000), Suppl 4
45.74 J.M. Fine, T. Gordon, L.C. Chen, P. Kinney, G. Falcone, W.S. Beckett: Metal fume fever: characterization of clinical and plasma IL-6 responses in controlled human exposures to zinc oxide fume at and below the threshold limit value, J. Occup. Environ. Med. **39**, 722–726 (1997)
45.75 W.G. Kuschner, A. D'Alessandro, H. Wong, P.D. Blanc: Early pulmonary cytokine responses to zinc oxide fume inhalation, Environ. Res. **75**, 7–11 (1997)
45.76 H.C. Chuang, H.T. Juan, C.N. Chang, Y.H. Yan, T.H. Yuan, J.S. Wang, H.C. Chen, Y.H. Hwang, C.H. Lee, T.J. Cheng: Cardiopulmonary toxicity of pulmonary exposure to occupationally relevant zinc oxide nanoparticles, Nanotoxicology **8**, 593–604 (2014)
45.77 L. Huo, R. Chen, X. Shi, R. Bai, P. Wang, Y. Chang, C. Chen: High-content screening for assessing nanomaterial toxicity, J. Nanosci. Nanotechnol. **15**, 1143–1149 (2015)
45.78 V. Sharma, P. Singh, A.K. Pandey, A. Dhawan: Induction of oxidative stress, DNA damage and apoptosis in mouse liver after sub-acute oral exposure to zinc oxide nanoparticles, Mutat. Res. **745**, 84–91 (2012)
45.79 J. Wang, X. Deng, F. Zhang, D. Chen, W. Ding: ZnO nanoparticle-induced oxidative stress triggers apoptosis by activating JNK signaling pathway in cultured primary astrocytes, Nanoscale Res. Lett. **9**, 117 (2014)
45.80 V. Sharma, R.K. Shukla, N. Saxena, D. Parmar, M. Das, A. Dhawan: DNA damaging potential of zinc oxide nanoparticles in human epidermal cells, Toxicol. Lett. **185**, 211–218 (2009)
45.81 R. Chen, L. Huo, X. Shi, R. Bai, Z. Zhang, Y. Zhao, Y. Chang, C. Chen: Endoplasmic reticulum stress induced by zinc oxide nanoparticles is an earlier biomarker for nanotoxicological evaluation, ACS Nano **8**, 2562–2574 (2014)
45.82 M. Ramasamy, M. Das, S.S. An, D.K. Yi: Role of surface modification in zinc oxide nanoparticles and its toxicity assessment toward human dermal fibroblast cells, Int. J. Nanomed. **9**, 3707–3718 (2014)
45.83 M. Luo, C. Shen, B.N. Feltis, L.L. Martin, A.E. Hughes, P.F. Wright, T.W. Turney: Reducing ZnO nanoparticle cytotoxicity by surface modification, Nanoscale **6**, 5791–5798 (2014)
45.84 H. Yin, R. Chen, P.S. Casey, P.C. Ke, T.P. Davis, C. Chen: Reducing the cytotoxicity of ZnO nanoparticles by a pre-formed protein corona in a supplemented cell culture medium, RSC Adv **5**, 73963–73973 (2015)
45.85 R. Chen, C. Chen: Nanotoxicity. In: *The Nanobiotechnology Handbook*, ed. by Y. Xie (Taylor Francis, Abingdon 2012) pp. 599–620
45.86 T.C. Le, H. Yin, R. Chen, Y. Chen, L. Zhao, P.S. Casey, C. Chen, D.A. Winkler: An experimental and computational approach to the development of ZnO nanoparticles that are safe by design, Small **12**, 3568–3577 (2016)
45.87 F. Zhao, H. Meng, L. Yan, B. Wang, Y. Zhao: Nanosurface chemistry and dose govern the bioaccumulation and toxicity of carbon nanotubes, metal nanomaterials and quantum dots in vivo, Sci. Bull. **60**, 3–20 (2015)
45.88 H.S. Nalwa, Y. Zhao: *Nanotoxicology* (American Scientific, Stewenson Ranch 2007)
45.89 Z. Chen, H. Meng, G. Xing, C. Chen, Y. Zhao, G. Jia, T. Wang, H. Yuan, C. Ye, F. Zhao, Z. Chai, C. Zhu, X. Fang, B. Ma, L. Wan: Acute toxicological effects of copper nanoparticles in vivo, Toxicol. Lett. **163**, 109–120 (2006)
45.90 Y. Liu, Y. Gao, L. Zhang, T. Wang, J. Wang, F. Jiao, W. Li, Y. Liu, Y. Li, B. Li, Z. Chai, G. Wu, C. Chen: Potential health impact on mice after nasal instillation of nano-sized copper particles and their

translocation in mice, J. Nanosci. Nanotechnol. **9**, 6335–6343 (2009)

45.91 L. Zhang, R. Bai, B. Li, C. Ge, J. Du, Y. Liu, L. Le Guyader, Y. Zhao, Y. Wu, S. He, Y. Ma, C. Chen: Rutile TiO(2) particles exert size and surface coating dependent retention and lesions on the murine brain, Toxicol. Lett. **207**, 73–81 (2011)

45.92 C. Carlson, S.M. Hussain, A.M. Schrand, L.K. Braydich-Stolle, K.L. Hess, R.L. Jones, J.J. Schlager: Unique cellular interaction of silver nanoparticles: size-dependent generation of reactive oxygen species, J. Phys. Chem. B **112**, 13608–13619 (2008)

45.93 J. Rejman, V. Oberle, I. Zuhorn, D. Hoekstra: Size-dependent internalization of particles via the pathways of clathrinand caveolae-mediated endocytosis, Biochem. J. **377**, 159–169 (2004)

45.94 G. Oberdörster, J. Ferin, B.E. Lehnert: Correlation between particle size, in vivo particle persistence and lung injury, Environ. Health Perspect. **102**, 173–179 (1994), Suppl 5

45.95 S. Beg, M. Rizwan, A.M. Sheikh, M.S. Hasnain, K. Anwer, K. Kohli: Advancement in carbon nanotubes: Basics, biomedical applications and toxicity, J. Pharm. Pharmacol. **63**, 141–163 (2011)

45.96 G. Jia, H. Wang, L. Yan, X. Wang, R. Pei, T. Yan, Y. Zhao, X. Guo: Cytotoxicity of carbon nanomaterials: single-wall nanotube, multi-wall nanotube, and fullerene, Environ. Sci. Technol. **39**, 1378–1383 (2005)

45.97 Y. Qiu, Y. Liu, L. Wang, L. Xu, R. Bai, Y. Ji, X. Wu, Y. Zhao, Y. Li, C. Chen: Surface chemistry and aspect ratio mediated cellular uptake of Au nanorods, Biomaterials **31**, 7606–7619 (2010)

45.98 L. Meng, R. Chen, A. Jiang, L. Wang, P. Wang, C.Z. Li, R. Bai, Y. Zhao, H. Autrup, C. Chen: Short multiwall carbon nanotubes promote neuronal differentiation of PC12 cells via up-regulation of the neurotrophin signaling pathway, Small **9**, 1786–1798 (2013)

45.99 J. Wang, Y. Liu, F. Jiao, F. Lao, W. Li, Y. Gu, Y. Li, C. Ge, G. Zhou, B. Li, Y. Zhao, Z. Chai, C. Chen: Time-dependent translocation and potential impairment on central nervous system by intranasally instilled TiO(2) nanoparticles, Toxicology **254**, 82–90 (2008)

45.100 R.A. Yokel, R.C. Macphail: Engineered nanomaterials: Exposures, hazards and risk prevention, J. Occup. Med. Toxicol. **6**, 7 (2011)

45.101 M. Seipenbusch, A. Binder, G. Kasper: Temporal evolution of nanoparticle aerosols in workplace exposure, Ann. Occup. Hyg. **52**, 707–716 (2008)

45.102 D. Brouwer: Exposure to manufactured nanoparticles in different workplaces, Toxicology **269**, 120–127 (2010)

45.103 R. Chen, B. Hu, Y. Liu, J. Xu, G. Yang, D. Xu, C. Chen: Beyond PM2.5: The role of ultrafine particles on adverse health effects of air pollution, Biochim. Biophys. Acta - Gen. Subjects **1860**, 2844–2855 (2016)

45.104 R. Chen, X. Shi, R. Bai, W. Rang, L. Huo, L. Zhao, D. Long, D. Pui, C. Chen: Airborne nanoparticle pollution in a wire electrical discharge machining workshop and potential health risks, Aerosol Air Qual. Res. **15**, 284–294 (2015)

45.105 X. Shi, R. Chen, L. Huo, L. Zhao, R. Bai, D. Long, D.Y. Pui, W. Rang, C. Chen: Evaluation of nanoparticles emitted from printers in a clean chamber, a copy center and office rooms: Health risks of indoor air quality, J. Nanosci. Nanotechnol. **15**, 9554–9564 (2015)

45.106 L.E. Murr, K.F. Soto, K.M. Garza, P.A. Guerrero, F. Martinez, E.V. Esquivel, D.A. Ramirez, Y. Shi, J.J. Bang, J. Venzor III: Combustion-generated nanoparticulates in the El Paso, TX, USA/Juarez, Mexico Metroplex: Their comparative characterization and potential for adverse health effects, Int. J. Environ. Res. Public Health **3**, 48–66 (2006)

45.107 F.R. Cassee, H. Muijser, E. Duistermaat, J.J. Freijer, K.B. Geerse, J.C. Marijnissen, J.H. Arts: Particle size-dependent total mass deposition in lungs determines inhalation toxicity of cadmium chloride aerosols in rats. Application of a multiple path dosimetry model, Arch. Toxicol. **76**, 277–286 (2002)

45.108 S. Anjilvel, B. Asgharian: A multiple-path model of particle deposition in the rat lung, Fundam. Appl. Toxicol. **28**, 41–50 (1995)

45.109 M.M. Methner, M.E. Birch, D.E. Evans, B.K. Ku, K. Crouch, M.D. Hoover: Identification and characterization of potential sources of worker exposure to carbon nanofibers during polymer composite laboratory operations, J. Occup. Environ. Hyg. **4**, D125–130 (2007)

45.110 L. Belyanskaya, S. Weigel, C. Hirsch, U. Tobler, H.F. Krug, P. Wick: Effects of carbon nanotubes on primary neurons and glial cells, Neurotoxicology **30**, 702–711 (2009)

45.111 K. Pulskamp, S. Diabate, H.F. Krug: Carbon nanotubes show no sign of acute toxicity but induce intracellular reactive oxygen species in dependence on contaminants, Toxicol. Lett. **168**, 58–74 (2007)

45.112 C. Ge, W. Li, Y. Li, B. Li, J. Du, Y. Qiu, Y. Liu, Y. Gao, Z. Chai, C. Chen: Significance and systematic analysis of metallic impurities of carbon nanotubes produced by different manufacturers, J. Nanosci. Nanotechnol. **11**, 2389–2397 (2011)

45.113 L. Meng, A. Jiang, R. Chen, C.Z. Li, L. Wang, Y. Qu, P. Wang, Y. Zhao, C. Chen: Inhibitory effects of multiwall carbon nanotubes with high iron impurity on viability and neuronal differentiation in cultured PC12 cells, Toxicology **313**, 49–58 (2013)

45.114 G. Zeng, G. Wang, F. Guan, K. Chang, H. Jiao, W. Gao, S. Xi, B. Yang: Human amniotic membrane-derived mesenchymal stem cells labeled with superparamagnetic iron oxide nanoparticles: the effect on neuron-like differentiation in vitro, Mol. Cell Biochem. **357**, 331–341 (2011)

45.115 J.J. Yin, F. Lao, P.P. Fu, W.G. Wamer, Y. Zhao, P.C. Wang, Y. Qiu, B. Sun, G. Xing, J. Dong, X.J. Liang, C. Chen: The scavenging of reactive oxygen species and the potential for cell protection by functionalized fullerene materials, Bio-

materials **30**, 611–621 (2009)
45.116 Y. Liu, F. Jiao, Y. Qiu, W. Li, F. Lao, G. Zhou, B. Sun, G. Xing, J. Dong, Y. Zhao, Z. Chai, C. Chen: The effect of $Gd@C_{82}(OH)_{22}$ nanoparticles on the release of Th1/Th2 cytokines and induction of TNF-alpha mediated cellular immunity, Biomaterials **30**, 3934–3945 (2009)
45.117 L. Wang, Y.F. Li, L. Zhou, Y. Liu, L. Meng, K. Zhang, X. Wu, L. Zhang, B. Li, C. Chen: Characterization of gold nanorods in vivo by integrated analytical techniques: Their uptake, retention, and chemical forms, Anal. Bioanal. Chem. **396**, 1105–1114 (2010)
45.118 R.R. Arvizo, O.R. Miranda, M.A. Thompson, C.M. Pabelick, R. Bhattacharya, J.D. Robertson, V.M. Rotello, Y.S. Prakash, P. Mukherjee: Effect of nanoparticle surface charge at the plasma membrane and beyond, Nano Lett **10**, 2543–2548 (2010)
45.119 Y. Liu, W. Li, F. Lao, Y. Liu, L. Wang, R. Bai, Y. Zhao, C. Chen: Intracellular dynamics of cationic and anionic polystyrene nanoparticles without direct interaction with mitotic spindle and chromosomes, Biomaterials **32**, 8291–8303 (2011)
45.120 H.F. Krug: Nanosafety research–are we on the right track?, Angew. Chem. Int. Ed. Engl. **53**, 12304–12319 (2014)
45.121 H. Thomas: Toxicology for the twenty-first century, Nature **460**, 208–212 (2009)
45.122 M. Firestone, R. Kavlock, H. Zenick, M. Kramer, U.E.W.G.F. Toxicity: The US environmental protection agency strategic plan for evaluating the toxicity of chemicals, J. Toxicol. Environ. Health B Crit. Rev. **13**, 139–162 (2010)
45.123 D. Krewski, D. Acosta Jr., M. Andersen, H. Anderson, J.C. Bailar III, K. Boekelheide, R. Brent, G. Charnley, V.G. Cheung, S. Green Jr., K.T. Kelsey, N.I. Kerkvliet, A.A. Li, L. McCray, O. Meyer, R.D. Patterson, W. Pennie, R.A. Scala, G.M. Solomon, M. Stephens, J. Yager, L. Zeise: Toxicity testing in the 21st century: A vision and a strategy, J. Toxicol. Environ. Health B Crit. Rev. **13**, 51–138 (2010)
45.124 A. Nel, T. Xia, L. Madler, N. Li: Toxic potential of materials at the nanolevel, Science **311**, 622–627 (2006)
45.125 L. Wang, Y. Min, D. Xu, F. Yu, W. Zhou, A. Cuschieri: Membrane lipid peroxidation by the peroxidase-like activity of magnetite nanoparticles, Chem. Commun. **50**, 11147–11150 (2014)
45.126 Y. Li, Y. Liu, Y. Fu, T. Wei, L.L. Guyader, G. Gao, R.S. Liu, Y.Z. Chang, C. Chen: The triggering of apoptosis in macrophages by pristine graphene through the MAPK and TGF-beta signaling pathways, Biomaterials **33**, 402–411 (2012)
45.127 L. Huo, R. Chen, L. Zhao, X. Shi, R. Bai, D. Long, F. Chen, Y. Zhao, Y.Z. Chang, C. Chen: Silver nanoparticles activate endoplasmic reticulum stress signaling pathway in cell and mouse models: The role in toxicity evaluation, Biomaterials **61**, 307–315 (2015)
45.128 R. Chen, D. Ling, L. Zhao, S. Wang, Y. Liu, R. Bai, S. Baik, Y. Zhao, C. Chen, T. Hyeon: Parallel comparative studies on mouse toxicity of oxide nanoparticle- and gadolinium-based T1 MRI contrast agents, ACS Nano **9**, 12425–12435 (2015)
45.129 H. Yuan, Q. Zhang, J. Guo, T. Zhang, J. Zhao, J. Li, A. White, P.L. Carmichael, C. Westmoreland, S. Peng: A PGC-1alpha-mediated transcriptional network maintains mitochondrial redox and bioenergetic homeostasis against doxorubicin-induced toxicity in human cardiomyocytes: Implementation of TT21C, Toxicol. Sci. **150**, 400–417 (2016)
45.130 H. James, N. Sashi, S. Rachel, C. Conn, K. Sinead, W. Yvonne: A high-throughput dual parameter assay for assessing drug-induced mitochondrial dysfunction provides additional predictivity over two established mitochondrial toxicity assays, Toxicol. Vitro **27**, 560–569 (2012)
45.131 J. Xu, M. Han, Y. Ren, J. Li: The principle of compromise in competition: Exploring stability condition of protein folding, Sci. Bull. **60**, 76–85 (2015)
45.132 A. van Schadewijk: E.F. van't Wout, J. Stolk, P.S. Hiemstra: A quantitative method for detection of spliced x-box binding protein-1 (XBP1) mRNA as a measure of endoplasmic reticulum (ER) stress, Cell Stress Chaperones **17**, 275–279 (2012)
45.133 A. Nel, T. Xia, H. Meng, X. Wang, S. Lin, Z. Ji, H. Zhang: Nanomaterial toxicity testing in the 21st century: Use of a predictive toxicological approach and high-throughput screening, Acc. Chem. Res. **46**, 607–621 (2013)
45.134 C. Chen, Y.F. Li, Y. Qu, Z. Chai, Y. Zhao: Advanced nuclear analytical and related techniques for the growing challenges in nanotoxicology, Chem. Soc. Rev. **42**, 8266–8303 (2013)
45.135 Z. Chai, X. Mao, Z. Hu, Z. Zhang, C. Chen, W. Feng, S. Hu, H. Ouyang: Overview of the methodology of nuclear analytical techniques for speciation studies of trace elements in the biological and environmental sciences, Anal. Bioanal. Chem. **372**, 407–411 (2002)
45.136 L. Wang, C. Chen: Pathophysiologic mechanisms of biomedical nanomaterials, Toxicol. Appl. Pharm. **299**, 30–40 (2016)
45.137 Y.F. Li, C. Chen, A. Bai, S. Li, Q. Wang, J. Wang, Y. Gao, Y. Zhao, Z. Chai: Simultaneous speciation of selenium and mercury in human urine samples from long-term mercury-exposed populations with supplementation of selenium-enriched yeast by HPLC-ICP-MS, J. Anal. At. Spectrom. **22**, 925–930 (2007)
45.138 C. Chen, Z. Chai, Y. Gao: *Nuclear Analytical Techniques for Metallomics and Metalloproteomics* (RSC, Cambridge 2010)
45.139 D. Chen, W. Chao, N. Xin, S. Li, R. Li, M. Guan, L. Zhuang, C. Chen, C. Wang, C. Shu: Photoacoustic imaging guided near-infrared photothermal therapy using highly water-dispersible single-walled carbon nanohorns as theranostic agents, Adv. Funct. Mater. **24**, 6621–6628 (2014)
45.140 L. Wang, J. Li, J. Pan, X. Jiang, Y. Ji, Y. Li, Y. Qu, Y. Zhao, X. Wu, C. Chen: Revealing the binding structure of the protein corona on gold nanorods

using synchrotron radiation-based techniques: Understanding the reduced damage in cell membranes, J. Am. Chem. Soc. **135**, 17359–17368 (2013)

45.141 L. Wang, T. Zhang, P. Li, W. Huang, J. Tang, P. Wang, J. Liu, Q. Yuan, R. Bai, B. Li, K. Zhang, Y. Zhao, C. Chen: Use of synchrotron radiation-analytical techniques to reveal chemical origin of silver-nanoparticle cytotoxicity, ACS Nano **9**, 6532–6547 (2015)

45.142 D. Drescher, P. Guttmann, T. Büchner, S. Werner, G. Laube, A. Hornemann, B. Tarek, G. Schneider, J. Kneipp: Specific biomolecule corona is associated with ring-shaped organization of silver nanoparticles in cells, Nanoscale **5**, 9193–9198 (2013)

45.143 P. Zhang, Y. Ma, Z. Zhang, X. He, J. Zhang, Z. Guo, R. Tai, Y. Zhao, Z. Chai: Biotransformation of ceria nanoparticles in cucumber plants, ACS Nano **6**, 9943–9950 (2012)

45.144 Z. Chen, L. Ying, B. Sun, L. Han, J. Dong, L. Zhang, L. Wang, W. Peng, Y. Zhao, C. Chen: Polyhydroxylated metallofullerenols stimulate IL-1β secretion of macrophage through TLRs/MyD88/NF-κB pathway and NLRP3 inflammasome activation, Small **10**, 2362–2372 (2014)

46. Nanoscience and Nanotechnology Convergence

William S. Bainbridge

A series of major conferences that emerged from the National Nanotechnology Initiative explored connections to other fields of science and engineering, resulting after 15 years in a comprehensive vision of how science, technology, and society could converge. One of two initial orientations focused on the human and ethical implications of nanotechnology, thus connecting nanoscience to the social sciences. The other considered the current partnership and possible future unification of four NBIC domains: Nanotechnology, Biotechnology, Information technology, and new technologies based in Cognitive science. Most recently, the extensive series of book-length conference reports has led to a comprehensive overview, *Handbook of Science and Technology Convergence*, that outlines concepts, research methods, and educational approaches across four *platforms* that comprehensively define the world in which humans live:

1. Foundational technologies
2. Human-scale phenomena
3. The Earth-scale system
4. Societal-scale implications.

46.1 **Background** 1587
46.1.1 The Convergence Social Movement 1587
46.1.2 The Convergence Conferences 1588
46.1.3 Parallel Activities in Europe 1591

46.2 **Convergence with Society** 1591
46.2.1 Convergence Platforms 1592
46.2.2 Handbook Overview of Convergence 1594

46.3 **Concepts and Methods** 1595
46.3.1 Five General Principles 1596
46.3.2 The Information Technology Infrastructure 1597

46.4 **Convergence in Education** 1598
46.4.1 Innovations in Formal Education 1598
46.4.2 Integration of Learning with Life 1599

46.5 **Conclusions** 1600

References 1600

46.1 Background

It should be obvious to anyone that nanoscience and nanotechnology can and should be integrated with several other fields of science and engineering, especially to develop new and better solutions for human problems. Atoms are smaller than the nanoscale, and living cells are larger, yet both physics and microbiology concern phenomenon that operate across the nanoscale. Several components of computer electronics are small enough to qualify as nanotechnology, and the flows of electrons inside semiconductors cover a range of distances. Nanoscale particles and layers on manufactured objects contribute to technologies on a wide variety of scales. More broadly, every human-significant application of nanotechnology requires some kind of collaboration with at least one other technical discipline. As obvious as these points are, it is not clear what general principles can guide collaboration across disciplinary boundaries, and what concerted efforts may be needed to maximize the contributions of nanoscience and nanotechnology to human society.

46.1.1 The Convergence Social Movement

At the beginning of the current century, a key leader of the nanotechnology movement, Mihail C. Roco, envisioned a vast intellectual and social transformation he called *convergence*. He was the first chairman of the Nanoscale Science, Engineering, and Technology (NSET) subcommittee of the National Science and Technology Council, and currently serves as senior

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_46

advisor for Science and Engineering at the National Science Foundation (NSF), two very influential positions in the United States government. With support from several specific NSF programs and other government agencies, he launched a pair of conference sequences that eventually merged together, to chart the goals and means for transformation of science, technology, and society through convergence, building outward from the nanoscale.

The first effort focused on *societal implications*, a theme that was built into the National Nanotechnology Initiative (NNI) from its very beginning [46.1, p. 2]:

> *The NNI is balanced across five broad activities: fundamental research; grand challenges; centers and networks of excellence; research infrastructure; and societal/workforce implications. Under this last activity, nanotechnology's effect on society – legal, ethical, social, economic, and workforce preparation – will be studied to help identify potential concerns and ways to address them. As the NNI is commencing, there is* a rare opportunity to integrate the societal studies and dialogues from the very beginning and to include societal studies as a core part of the NNI investment strategy.

Perhaps because of the substantive connections to biomedical research, or the naive and often sinister early representations of nanotechnology by science fiction writers, from the very beginning it was clear that researchers in this field needed to set high ethical standards. That could require consultations with philosophers of ethics, as well as intensive discussions within the field. But the NNI went further, concluding that the social sciences also needed to be involved. Every new technology has costs as well as benefits, both direct and indirect, so clearly economics has a role to play. As the goal is to achieve benefits for society as a whole, other social sciences also have their distinctive roles to play, most obviously sociology but also cultural anthropology and political science. This need not retard progress through excessive regulation, although on occasion government may need to draw upon social science to set priorities, favoring one technical approach over another. Hopefully, there will be many occasions when the social sciences document how nanotechnology can contribute to societal progress, in ways that encourage greater public investment.

The second effort concerned *converging technologies*, the mutually supportive integration of multiple technical fields. Since the starting point was nanotechnology, the obvious initial partners were biotechnology and information technology, but Roco was convinced that a fourth component was equally necessary, new technologies based on the discoveries of cognitive science. Given that nanoscientists are human beings, and their research instruments typically involve electronic data collection and analysis, all three of the other fields support nanoscale research. Put simply, their brains are biological versions of computers, and both their brains and their nonbiological instrumentation must follow sound cognitive principles. Thus, Roco's initial conception of converging technologies began with these four *NBIC* fields: Nanotechnology, Biotechnology, Information technology, and new technologies based on Cognitive science. But unification was really provided by a very fundamental conception of human discovery and invention [46.2, p. ix]:

> *Convergence of diverse technologies is based on material unity at the nanoscale and on technology integration from that scale. The building blocks of matter that are fundamental to all sciences originate at the nanoscale. Revolutionary advances at the interfaces between previously separate fields of science and technology are ready to create key transforming tools for NBIC technologies. Developments in systems approaches, mathematics, and computation in conjunction with NBIC allow us for the first time to understand the natural world, human society, and scientific research as closely coupled complex, hierarchical systems. At this moment in the evolution of technical achievement, improvement of human performance through integration of technologies becomes possible.*
>
> *Examples of payoffs may include improving work efficiency and learning, enhancing individual sensory and cognitive capabilities, revolutionary changes in healthcare, improving both individual and group creativity, highly effective communication techniques including brain-to-brain interaction, perfecting human–machine interfaces including neuromorphic engineering, sustainable and intelligent environments including neuroergonomics, enhancing human capabilities for defense purposes, reaching sustainable development using NBIC tools, and ameliorating the physical and cognitive decline that is common to the aging mind.*

46.1.2 The Convergence Conferences

For each of these two parallel efforts, a number of planning meetings were held at the National Science Foundation in Arlington, Virginia. The first societal implications conference was held there September 28–29, 2000, and the first converging technologies conference, December 3–4, 2001. Table 46.1 outlines the sequence of conferences, along with the numbers of scientists,

Table 46.1 Societal implications and converging technologies conferences

Conference	Location	Year	Contributors
Societal Implications of Nanoscience and Nanotechnology	Arlington, Virginia	2000	64
Converging Technologies for Improving Human Performance	Arlington, Virginia	2001	79
Nanotechnology: Societal Implications – Maximizing Benefits for Humanity	Arlington, Virginia	2003	130
The Coevolution of Human Potential and Converging Technologies	Los Angeles, California	2003	28
Managing Nano-Bio-Info-Cogno Innovations: Converging Technologies in Society	New York City	2004	34
Progress in Convergence: Technologies for Human Wellbeing	Kailua-Kona, Hawaii	2005	31
Convergence of Knowledge, Technology and Society: Beyond Nano-Bio-Info-Cognitive Technologies	São Paulo, Brazil; Arlington, Virginia; Leuven, Belgium; Seoul, Korea; Beijing, China; again Arlington, Virginia	2011–2012	312
Handbook of Science and Technology Convergence	online	2013–2015	76

engineers, and government leaders who contributed to each. These influential meetings occurred in two waves. The first ran from 2000–2005, with the last book-length reports being published in 2006, keeping the societal implications and converging technologies discussion largely separate from each other. The second wave ran 2011–2015 with the last book length report published in 2016, combining both major themes and providing a solid basis for considering the wider convergence of science, technology, and society.

The 2003 societal implications conference triggered an interesting debate, and is a good example of how the more complex conferences were designed. To what extent should a government agency give the impression that it agrees with especially visionary projections of future technologies and societal conditions? The obvious answer might be that government should always be cautious about associating itself in the public mind with any particular ideology, and some of the more stimulating products of futurology seem more like ideology than like objective science. Yet the future could be very different from the past, so unusual ideas should not be summarily dismissed, especially if they are proposed by brilliant and knowledgeable people. The government needs access to a wide range of perspectives, which it can then assess with care, and the general public deserves to be informed about these debates that may shape the future as well as imagine it.

At the same time, the individual contributors to all of the meetings not only represented a great diversity of intellectual traditions and forms of technical expertise, but also tended to have a diversity of personal styles. It is a mistake to imagine that rigorous science drains personality out of the equation. Thus, it was important to accord equal respect to all of the contributors, all of whom were remarkable people, even though the texts they wrote for the reports differed on every level, from vocabulary and grammar, up to optimism versus pessimism, and conservative versus radical. The practical solution to these challenges was to structure the published reports in different ways, most explicitly for the publications from the 2003 societal implications conference, but the others as well.

In 2005, the National Science Foundation and National Nanotechnology Coordination Office published a 106 page report, that summarized the deliberations of the conference, and served as a record of the events and group discussions. The first day of meetings began at noon, and served as an orientation, with three sets of presentations that followed a welcoming:

1. Considering nanotechnology as a national endeavor to achieve technological and economic goals, guided by a science and education vision
2. Identifying goals for the nanotechnology initiative, to benefit industry and society, and employing social science for assessing nanotechnology
3. Recognizing broader implications for the quality of life, aware of ethical issues, and intelligently navigating nanotechnology through society.

Prior to the gathering, each invited participant had selected or been assigned to two very different discussion groups, from a list of ten, for which they prepared brief contributions:

1. Productivity and equity
2. Future economic scenarios
3. The quality of life
4. Future social scenarios
5. Converging technologies

6. National security and space exploration
7. Ethics, governance, risk, and uncertainty
8. Public policy, legal, and international aspects
9. Interaction with the public
10. Education and human resource development.

On the morning of the second day, the participants divided into panels to discuss the first five topics from this list, and prepared outlines to share with all participants in a plenary session. In the afternoon, the same was done for the second five topics, with individuals working in very different combinations on the two halves of the day. The third and final day of the meeting was a combination of individual work and group sessions, to assemble a rough draft of a formal report. Thus, the very organization of this societal implications conference was an example of convergence.

After the meeting, on the basis of the collective report, individual participants and teams were invited to write up more detailed chapters, encouraged to be visionary as well as rigorous, and given considerable freedom to express their own perspectives. The final result was a two-volume publication, the group workshop report [46.3] and the collection of individual contributions [46.4]. This second volume assembled the chapters into a somewhat different set of seven aspects of societal implications of nanotechnology:

1. Socioeconomic impacts and commercialization of nanotechnology
2. Social scenarios
3. Converging technologies
4. Ethics and law
5. Governance
6. Public perceptions
7. Education.

While converging technologies was among the ten topics of the first report volume, and among the seven of the second, it really interwove across all the other topics. Thus, many analytical concepts relevant to convergence naturally were used extensively in the discussions, including these that were given formal definitions:

- *Converging technologies*: Unification of nanotechnology, biotechnology, information technology and cognitive technologies, based largely on the unity of nature at the nanoscale and on transforming tools for research and production.
- *Consilience*: Convergence of the natural sciences into one, unified body of knowledge.
- *Boundary work*: Effort invested in defining fields of activity, delineating the differences between concepts, and locating things in a socio-cultural context.
- *Diffusion*: The spread of a technical idea or other cultural element from one group, place, or application to another.
- *Interactional expertise*: The skill and experience required to pass information from one scientist to another, especially the ability to act as a cultural broker between significantly different fields.
- *Trading zone*: A metaphoric place, consisting of shared assumptions and terminology, where specialists from different fields can come together, communicate, and cooperate.
- *Co-evolution*: The simultaneous development of two different phenomena in interaction with each other, for example the emergence of nanotechnology coupled with the emergence of research on its societal implications.
- *Transilience*: In technology, a revolutionary leap that renders older technology obsolete, often involving convergence of technologies and a new relationship between producers and customers.
- *Disruptive technology*: New technologies that have the effect of changing the relations between existing organizations and institutions in society, often providing a net benefit even as some people may be harmed at least temporarily.
- *Cultural lag*: A maladjustment in the culture that comes about when rapid change in one area (such as technological innovation) is not immediately matched by concomitant change in other areas that are closely related to it.
- *Gap analysis*: A method of policy analysis that begins with a clear picture of a desired future state, compares it with the current state, and identifies actions that need to be taken to bridge the gap between them.

The three converging technologies conferences held in California, New York, and Hawaii were structured much more simply, as opportunities for leaders across the converging disciplines to present papers of great diversity but significant intellectual depth. After the wave of book publications in 2006, the NBIC community continued to be active, but chiefly through individual and informal efforts. Around the world, a number of scientific organizations and governments showed great interest, often conducting their own local conferences and research activities.

46.1.3 Parallel Activities in Europe

Already in 2004 a report for the European Commission applied this definition: *Converging technologies are enabling technologies and knowledge systems that enable each other in the pursuit of a common goal* and observed [46.5, p. 14]:

> *Conceptually, nanotechnology enables other technologies by providing a common framework for all hardware-level engineering problems. Everything that consists of molecules can, in principle, be integrated with each other. An understanding of properties at the nanoscale allows for the realization of desirable architectures at the micro- and macroscale. Instrumentally, nanotechnology enables biotechnology by developing new imaging techniques, probes and sensors. It contributes to the miniaturization demands of information technology. Also, nanochips and nanosensors are set to enable advances in the new world of bioinformatics.*

Over a period of years, European academic and government agencies developed a perspective complementary to, but somewhat different from the NBIC conceptualization, which seemed distinctively American to them, as expressed in a 2008 report [46.6, pp. II–IV]:

> *... the European concept of Converging Technologies for the European Knowledge Society (CTEKS) adopts a demand-driven approach in which CT respond to societal needs and demands. A major theme in both the US and European approaches is cooperation between the disciplines: the European approach makes a special issue of interdisciplinary cooperation and its specific conditions and problems, whereas the US approach argues for a new unity of science underpinned by reductionism enabled by the possibility of tracking virtually everything down to the nano-level.*
>
> *... both the NBIC and the CTEKS concepts devote much attention to issues of cooperation between scientific disciplines to realize applications of CT. While the NBIC initiative argues for underpinning cooperation through a base of common knowledge at the nano-level, the CTEKS concept argues for strong interdisciplinary cooperation including both hard sciences and social sciences.*

For a time there seemed to be a real difference of emphasis between the American and European approaches, with the Europeans raising a greater variety of ethical concerns, while at the same time suggesting that governments should set priorities to invest heavily in those areas of converging technologies that might be most beneficial for human beings. However, there may not have really been much trans-Atlantic difference of opinion among the scientists and engineers who were involved, rather in the political cultures of their different nations. While the Europeans suggested that converging technologies needed more collaboration with social scientists, in fact social science had been centrally involved in both the societal implications and converging technologies strands of the American effort. For decades, socialist parties had been influential in European politics, with a natural inclination to encourage social science as a tool of government policy, while over the decades a powerful trend in American politics had attempted to end government funding of social science research altogether [46.7]. Near the beginning of its second decade, the converging technologies movement sought to achieve greater convergence within itself and with society.

46.2 Convergence with Society

Even within the American wing of the Converging Technologies movement, social science had played an important role from the very beginning, although not officially becoming one of the government priorities [46.8]. When it came time for an assessment of progress after the first decade, however, it was clear not only that the social sciences should be central to many future efforts, but that convergence should expand to encompass a union of science, technology, and society. Indeed, when Mihail Roco launched a new series of international conferences, he called it *Convergence of Knowledge, Technology, and Society*, including science within the definition of *knowledge* but recognizing that humanities such as history and many less formal fields of human knowledge must also be included. The subtitle of the rather large resulting report was explicitly *Beyond Convergence of Nano-Bio-Info-Cognitive Technologies*, and the introductory paragraph set out an ambitious agenda [46.9, p. xiii]:

> *Convergence of Knowledge and Technology for the Benefit of Society (CKTS) is the core opportunity for progress in the 21st century. It is defined in this study as the escalating and transformative*

interaction among seemingly distinct scientific disciplines, technologies, communities, and domains of human activity to achieve mutual compatibility, synergism, and integration, and through this process to create added value and branch out into emerging areas to meet shared goals. Convergence is as essential to our future knowledge society as engines were to the industrial revolution. CKTS allows society to answer questions and resolve problems that isolated capabilities cannot, as well as to create new competencies, technologies, and knowledge on this basis. This concept is centered on the principles presented in this report: of interdependence in nature and society with application to the essential platforms of the human activity system; the enhancement of creativity and innovation within knowledge and technology through convergence–divergence (spin-off) evolutionary processes; a holistic system deduction approach being applied in decision analysis; the value of higher-level cross-domain languages to generate new solutions and support transfer of new knowledge; and the value of vision-inspired basic research endeavors.

The report was structured in ten substantive sections, each of which was developed through communications among a group of participants, both physically present at one or more of the conferences, and communicating afterward online. The expression *essential platforms of the human activity system* referred to a conceptual classification of domains in which human action has consequences on various scales, usually described in four categories:

1. Foundational science and technology tools
2. The individual human scale
3. The scale of planet Earth
4. The societal scale.

46.2.1 Convergence Platforms

The first platform, appropriately enough, is the original NBIC concept based on nanotechnology and its connections to the three other fields of technology. The contributors to discussion of the NBIC platform began by noting that the costs of research had been reduced considerably over the previous decade in such nano-related areas as gene sequencing, measurement and manipulation of nonbiological objects at the nanoscale, and the capacity of information systems to store and search ever greater and more complex sets of data. In considering the likely developments in the near-term future, the contributors considered the likelihood that the distinction between nanotechnology and the other technologies might fade, as the fields of science and engineering merged, yet what we currently call *nanotechnology* would remain significant if less distinct for many future technologies. They identified two illustrative goals for the coming decade:

1. Developing new tools and science for advanced electronic manufacturing
2. Greatly improve tools for studying the human brain
3. Cybersecurity risk management in development of innovative open source software systems, including those used to manipulate matter at the nanoscale.

The second platform concerns human-scale convergence and the quality of life, the realm in which human beings interact directly with technology, gain the benefits of innovation, and in some cases suffer indirect and unintended negative consequences. The team that considered this set of issues concluded that the really significant developments of the previous decade had occurred in public use of information technology: (1) democratization of Internet, and (2) reconceptualization of computing as a service, rather than the sale of hardware and software products. The challenge for the near future was viewed chiefly as the need for convergence to combat disorganization, notably through collaboratories and virtual organizations, thus emphasizing information technology and giving supportive roles to the three other NBIC fields. In a sense, this analysis treated as problematic the distinction between the human-scale and societal-scale platforms, because the end result of unlimited social disorganization would be a collection of isolated individual human beings, while total organization would unite the entire world. Disorganization would seem to alienate nanotechnology from the other fields, because only a very few individuals would possess a professional understanding of its technical details, yet organization might also wall off nanotechnology from other fields, unless it was inspired by the values of convergence.

The third platform covers the broadly defined natural environment of planet Earth, in the context of the solar system. Biodiversity and climate change were among the central topics, along with resource distribution including energy consumption by human civilization. This platform was not limited to the often rather contentious debates about sustainability, but sought to develop a context for evaluating such issues dispassionately and objectively, while also supporting the general advance of human knowledge about the world in which

we live. The Earth-scale platform was considered from five viewpoints:

1. Knowledge systems
2. Monitoring systems
3. Communication systems
4. Management systems
5. Earth-scale technologies and other contributing technologies.

Summarizing development over the previous decade, the working group considered that the capabilities of science and technology had increased significantly, but the problems had become worse, and the world's political systems were struggling to achieve anything like a consensus as these two trends were going in opposite directions. Seeking a solution to this difficult conflict, four government-supported initiatives were proposed, all requiring world-wide cooperation:

1. Global data and information infrastructure: This entails not merely the development of a unified, searchable, comprehensive database of all valid information about conditions on Earth, but also deployment of vast numbers of sensors for environmental monitoring, including many that would be worn by volunteers among the general publics of all nations.
2. Global catastrophic and existential risk assessment: This requires both significant advances in assessment methodologies, such as computer simulations based on *Big Data* archives, but also establishment of a new organization to conduct the assessments in the most reliable and authoritative manner, similar to the Intergovernmental Panel on Climate Change.
3. Global megacities initiative: To improve the living conditions of residents of large metropolitan areas, several types of systems must be integrated intelligently, including transportation, energy grids, smart homes, and decentralized manufacturing that may in many cases integrate nanotechnology with computer-controlled tools such as 3-D printers.
4. Global materials initiative: The components of megacity and industrial ecosystems must be maximally recyclable, reusable, and renewable, when feasible also serving other environmentally friendly purposes, such as sequestering carbon from the atmosphere in construction materials.

The societal-scale platform was considered primarily in integration with the other three, and in five implication areas that were reported separately:

1. Human health and physical potential
2. Human cognition and communication and the emergence of the cognitive society
3. Societal collective outcomes including manufacturing
4. People and physical infrastructure
5. Convergence of knowledge and technology for a sustainable society.

The concluding chapter of the book-length multiconference report identified six governance goals for the next decade, each of which would require convergence of multiple fields of science and technology for human and societal progress:

1. Advance innovation for economic productivity: New models for innovation must be developed, as well as common technical languages to facilitate collaboration across fields, sophisticated methods for assessing ethical and legal implications, and most probably a shift in emphasis from investment in traditional fields to emerging areas such as nanotechnology and genetics.
2. Support converging technologies to advance human potential and quality of life: Many prominent visionaries of technological progress have advocated development of methods not merely to give humans new tools, but to enhance human minds and bodies directly, for example reducing the effects of aging or improving the ability of children to learn; new methods and structures of socio-technical assessment are required to advise us which of these ideas are both practical and ethical.
3. Set grand challenges in the vision-inspired quadrant: This refers to a four-fold analysis of research that highlight projects that are between pure and applied, called *use-inspired basic research* [46.10], and addressed difficult challenges such as the future of space exploration, global sustainability, rigorous incorporation of imagination in the processes of discovery and invention, and reconceptualizing healthcare in terms of emerging approaches like nano-bio sensing for diagnosis.
4. Design integrated socio-technical systems: One organizational approach would be to found methodology and policy centers at universities and perhaps with industry connections, while others are to support a diversity of research projects developing knowledge and technology platforms, and establish major societal convergence databases to support development of new logical, mathematical, engineering, informatic, and philosophical paradigms.

5. Address deficits in risk governance for next-generation CKTS products: Risk governance can require balanced consideration at several levels of complexity, from simple statistical risk analysis done by the staff of the agency panning the action up to a combination of approaches, adding probabilistic risk modeling, assessment of risk balancing, and risk trade-off and deliberation including external experts and stakeholders in industry, directly affected groups in society, and the general public [46.11].
6. Improve the cultural balance between collaboration/harmony and confrontation, as informed by the global context and by CKTS: This sets idealistic goals for human cooperation, including conflict resolution between people from different nations or groups, maintaining existing international and intergroup harmonies, facilitating convergence between the real and virtual worlds such as Internet-based communities, and enabling better work and personal satisfaction for individuals.

46.2.2 Handbook Overview of Convergence

The 2016 *Handbook of Science and Technology Convergence* had some of the qualities of the collections of individual essays written after the conferences by participants, but was much larger, included additional authors, and was organized much more rigorously. It serves as a capstone to the series of conferences, and is a reference work establishing the conceptual basis of convergence as a mature orientation toward science, technology, and society. Each of the four convergence platforms was examined comprehensively, and will be summarized here. Two other sections of the handbook concerned concepts and methods, and convergence in education, which will be the next two sections of this chapter.

In summarizing the contributions about the Foundational Technologies Platform, Mihail Roco identified similarities across the four NBIC fields, and the ways that each could be central to some forms of convergence with the others. He argued that each of the four fields concerns a domain of nature that has its own fundamental building block: atom, gene, information bit, and neuronal synapse. He also saw four stages in the historical development of the movement:

1. The preliminary interdisciplinary phase of convergence that involved coincidental confluence of fields
2. The nanotechnology stage of convergence, in which concerted research and engineering efforts at the nanoscale called attention to affinities with adjacent fields
3. Convergence of Nanotechnology, Biotechnology, Information technology, and new technologies based on Cognitive science through the organization of NBIC conferences and other collective activities
4. Convergence of Knowledge, Technology, and Society (CKTS) beyond NBIC.

In discussing nano-centered convergence, Roco predicted that nanotechnology is currently transitioning from primarily an area of pure research to increasingly valuable applications, most obvious to the general public perhaps in medicine, but contributing to growth in many sectors of the world economy [46.12, p. 213]:

> *Using the same physical, chemical, and biological properties of nanoscale elements has proven valuable in application areas ranging from advanced materials such as metamaterials and nanophotonic structures, nanoelectronics such as* 10 nm *CMOS, and spintronic memory devices to nanobiotechnology and neurotechnology. Ability to probe single atom, single electron, single photon, single phonon, single charge, single spin, single-molecule vibration, and other parameters at the nanoscale has opened new capabilities for measuring, simulating, and changing continuum properties at the macroscale in all S&T areas.*

The Handbook's section on the Human-Scale Platform emphasized the many ways in which information technology is taking on ever more roles mediating between people and their environment, and between individuals and social groups. Specific topics included online collaboratories, virtual meetings, transformative communication media, robots serving as partners with humans, personalized and interactive arts and literature, open source technology development, and distributed manufacturing in which nanotechnology pays significant roles. An overview of this complex and rapidly changing convergence between human beings and information technologies identifies many fundamental principles, each of which is a domain of research that can progress best if combined with others [46.13, p. 363]:

> *The defining feature of the human species has always been that we employ technology to interact with nature. The modern case of human–computer interaction provides five distinct research and development approaches that could apply to any field of engineering: (1) ergonomics, (2) cognitive modeling, (3) user-centered design, (4) value-sensitive design, (5) technical culture. The felicitous symphony orchestra metaphor for social-technical sys-*

tems also identifies widely applicable principles: (1) division of labor, (2) a harmony of scientific concepts, (3) the need for significant human expertise, (4) social cohesion of teams, (5) human guidance, (6) properly defined general principles for decision making. Humanity may have reached a technological watershed at which the basis of much of the economy shifts from physical objects to information, but in any case innovation will require new collaborations between fields of expertise, in service of human wellbeing.

While the Earth-scale platform is the domain for much research and policy development concerning preservation of the environment, it also includes economic globalization, linguistic convergence, worldwide demographic trends, a sustainable food supply, and the potential human significance of the exploration of outer space. Many current trends seem quite negative, notably environmental degradation and social problems associated with globalization. However, there is also good reason for optimism, if enlightened science-based policies can be implemented [46.14, pp. 534–535]:

It also must be understood that anthropogenic change can also help protect our planetary life-support systems. The emergence and convergence of nanotechnology, biotechnology, information technology, cognitive (NBIC) science promise enormous societal and earth-scale sustainability benefits. Optimists can point to advanced technologies that reduce energy demand, emissions of greenhouse gases, air and water pollution, and industrial wastes. Water supplies and quality are being improved, food is being grown in more environmentally friendly manners, and applications of synthetic biology to reduce species extinction are gaining adherents. Humanity and our life support system can benefit from having these types of technologies sooner than later.

The societal-scale platform combines the institutions of society that have very wide-ranging impacts, including some that seem unscientific, notably government, culture, religion, and ethics. Even as we struggle with many narrow but important policy issues concerning the impacts of new technologies, a huge question lurks in the background: Has science reached its natural limits [46.15]? Well over a century ago, the frontier of the *Wild West* officially closed, and *Frederick Jackson Turner* raised concerns that a frontier is a necessary condition for democracy and cultural progress [46.16]. Without some dynamic, progressive domain for creative human action, society may decline into the equivalent of a global penitentiary, filled with poverty, suffering, and the consequences of injustice. Famously, *Vannevar Bush* called science *the endless frontier*, and *Star Trek* called outer space *the final frontier* [46.17, 18]:

Yet the final frontier *could actually be the nanoscale, and nanotechnology could overcome many economic and technical limitations by permitting efficient human design of structures across all size ranges. Or the* final frontier *could be the human mind, pioneered either by cognitive neuroscience or by cultural sciences, or by a convergence of both in the realization that mentality proceeds both at the individual and the social levels.*

46.3 Concepts and Methods

In the Handbook and the earlier conference books, a vast collection of analytic concepts and research methods was identified that could contribute to convergence of at least some areas of science and technology. Many of them are either already employed or worth considering for nanoscience and nanotechnology. In addition, several conference participants contributed case studies in which nanotechnology collaborated with an adjacent field, from which lessons could be learned.

Fairly often, when two fields interact very intensively, they may create a new field that is a hybrid of both, and an example considered seriously was *nanobiotechnology* [46.19]. As is the case with any new field, a formal definition has not fully congealed as yet, but nanobiotechnology is not merely the application of knowledge from nanoscience to biotechnology. It can also be development of biomimetic but nonbiological processes of use for specialized nanotechnology applications. From biology, we may learn much about separable chemical processes and structures of complex molecules, for example. They function within biology as features of complex evolutionary systems embedded in the biosphere, but can be applied and combined in different ways outside living organisms. This is not to say that success will be easy, even after a fundamental understanding has been achieved. After pilot studies have indicated how a valuable application could be de-

signed, perfecting and deploying it can still require very great investments.

A related example is convergence of *nanotechnology with neuroscience*, which not only gives each field new possibilities derived from the other, but also allows their partnership to contribute to computer and information science [46.20]. A major goal would be neuromorphic or neuromimetic computing, the creation of successors to computer chips in which the elements were structured and function like the neurons and synapses of the human brain. While research on the human brain is advancing rapidly, it is exceedingly difficult to study phenomena in living brains on the level of local nets of neurons, let alone individual neurons, in part because functional magnetic resonance imaging cannot operate at the fine resolution required. It may, however, be possible to deduce the structures and functions of small-scale neuron networks, with the assistance of insights from nanoscience. The results might require development of new programming methods in computer science, and many other innovations that cannot yet be specified.

Convergence, however, is not inexorable, unidirectional, or even the only creative process in a dynamic socio-technical system. Indeed, at times divergence becomes dominant, and a field effectively splits in two, each half develops somewhat independently, and then the pair become ready for a new phase of progress to be achieved by re-converging. An example of almost constant fission and recombination is social psychology, which is itself a convergence between psychology and sociology [46.21]. Under ideal but fairly common circumstances, a general convergence–divergence cycle can become established, in four phases [46.22, p. 79]:

1. Creative assembling of contributions from multiple fields leading to new concepts or ideas
2. System integration leading to a new assembly or invention for known uses
3. Technological innovation outputs leading to new products and applications
4. Spin-off outcomes that lead to solutions not possible before and that produce new competencies, tools, and applications.

46.3.1 Five General Principles

Both in the conference contributions, and in the wider social science literature on science and technology, a number of relatively obvious general principles have coalesced, even as research and visionary theorizing are identifying others. Five principles in particular have frequently been cited [46.23]:

1. *Exploiting interdependence among domains*: As in the example of nanotechnology-neuroscience convergence, the fields begin with very different vocabularies so shared terminology must be developed; databases must be structured and annotated so that both fields may effectively use them, and the scientists and engineers must be both assisted and rewarded in developing an encompassing social network including some interdisciplinary formal organizations. The motivation may begin with a preexisting awareness of how practitioners of each field need help from the other, to accomplish goals they have already defined. But then there must be sufficient vision, perhaps provided by leaders from both specialties, to recognize new opportunities of benefit to both, that can efficiently be exploited at reasonable cost. Thus, connections will begin in some of the most obvious areas, but lead to promising discoveries in other areas, hopefully triggering a cascade of creative cooperation.
2. *Improving the convergence–divergence evolutionary cycle*: A starting point would be to search through the literature on the history and social science of science and technology innovation, both to harvest relevant theories and findings that have already been articulated, but not yet combined, and to look for new questions and empirical relationships. For example, in general does the convergence–divergence cycle work best when it is visible to all major participants and occurring in isolation from potentially disruptive external forces? It is conceivable that sometimes the cycle is more productive when it is not fully understood by participants, because a diversity of opinion may facilitate innovative ideas, or lack of full comprehension may be the result of working at the cutting edge of discovery. Or, it may be that the collision of two initially separate convergence–divergence cycles could either magnify their productivity, or interfere with it. The cycle is itself a new theory, potentially generating new hypotheses, so quite reasonably it should be the focus of many kinds of research, including experiments as well as rigorous observation of naturally occurring cycles, to develop the knowledge required to improve today's rate of scientific discovery and engineering invention.
3. *System-logic deductive decision making and problem solving*: A common error is conceptualizing a complex system exclusively in terms of one of its components, and thus from only one perspective, thereby missing out on opportunities for more successful management of the system. Using system logic can mean integrating multiple analysis, for example considering multiple pathways through

the system before attempting to unify them through formal logic, statistical analysis, or computer simulation. Essentially always, human beings are components of the system, even if only as researchers studying it or the beneficiaries of its successful operation. Many choices exist concerning the tools with which to understand a complex dynamic system, from fractal analysis to network visualizations to knowledge mapping and beyond. At some point in the process it may be worth distinguishing two parallel systems: (1) a portion of the natural world, such as the complex interaction of structures and chemicals inside a living cell, and (2) a portion of the human world, such as the biochemists and nanoscientists studying the cell, and even a human medical patient to whom the cell belongs. Then new insights may come from combining the two systems into one, within a flexible, hierarchical logic system for making decisions.

4. *Creating and applying high-level cross-domain languages to facilitate transfer of knowledge and new solutions* In some cases, as has happened often in medicine, two different fields may have different names for the same phenomenon, so they can abandon one term in favor of the other, as they coordinate their technical dictionaries. In other cases, one field may have a concept of value to the other, which could give the term a more significant but no less precise meaning; an example is the transfer of *vacancy chain* from semiconductor physics, describing a nano-scale path along which electrons may travel, to sociology where it came to describe an almost identical structure in which the movement of employment opportunities or durable goods in economic exchanges may be described by essentially the same mathematics [46.24]. Once two or more fields have combined and harmonized their terminologies, not only can experts in one area function as near-experts in another area, but it becomes easier for highly motivated and intelligent volunteers to join the team in the capacity of citizen scientists [46.25]. Associated with the development of a high-level language, many of the crucial but smaller factual details can move from human minds to computer-based information systems, just as people no longer do long division by hand, thus redefining expertise less as total command of a particular area of science, to fluency with the high-level cross-domain language and skill operating the information system that manages the details.
5. *Using vision-inspired basic research to address long-term challenges*: How to be visionary and rigorous at the same time is very difficult, especially if great investments are required to develop a solution for problems that are still emerging, or for opportunities that we can only hope will become available in the distant future. Yet some significant fraction of our effort today should be invested in preparing for a positive long-term future, and some methods of forecasting, scenario development, and prototype research may have value for the particular domains in question. The dramatic phrase *grand challenge* has been traditionally applied to vast issues like protection of the environment or economic development of impoverished societies. Another term that must be taken seriously despite its grandiose connotation is *anticipatory governance*, which can refer to constant reassessment of the means to an important goal that does not merely respond to changes in the contexts of innovation and application, but also vigorously seeks new routes to the goal even when the current path seems satisfactory. Among the potentially innovative approaches is *reverse mapping*. Rather than starting one's analysis from the current situation and planning a route from here to the distant goal, one can analytically reverse the direction of time, and begin analysis at the goal and work backward from there to the current situation. Many aspects of vision-inspired basic research are risky, and people generally avoid taking risks with big investments. Yet it can be especially risky to follow a supposedly tried-and-true path to a revolutionary goal, without thinking critically about alternatives and exploring them in creative ways.

46.3.2 The Information Technology Infrastructure

Many participants at each of the conferences evaluated the kinds of information and communication technologies that would best facilitate convergence. Several examples were considered that superficially looked like conventional systems, but really were quite revolutionary when described in any depth. For example, it is obvious that researchers can benefit from shared databases, especially if they can be accessed online so that researchers need not be physically on site. Slightly more innovative is the plan to design each one with metadata descriptives and statistical analysis tools that are conducive both for communication across fields and for exploratory analyses incorporating new ideas. But those are somewhat antique ideas, while what is really needed is a *collective intelligence system* [46.26]. It would be the emergent result of synergy among three components:

1. A constantly evolving archive that goes beyond data to information and knowledge
2. Software and hardware tools that are not merely sophisticated but capable of rapid enhancement such as through user-created open source software and perhaps in some cases by nanotechnology-based quick-response 3-D printing of new instrumentation
3. A community composed of both subject-matter experts and diverse nonexpert contributors organized in a dynamic social network of vigorous personal commitment as well as in a computer-supported communication network.

Clearly, creation of a global collective intelligence system, or even a much more modest system for supporting convergence across many fields of science and engineering, will require innovations and investments of many kinds. The traditional methods for managing libraries and curating their contents are obsolete, for example the notion that each publication can be assigned to a single topic area category, when much more sophisticated and multidimensional information search methods are needed [46.27]. Indeed, simply accessing sources of information is of little value if we do not have multiple appropriate means for analyzing it and connecting it with other data, not merely through web links or statistical analysis but also through theory-based computer simulations [46.28]. Assembling the information, even using the most modern methods, is also insufficient, because we also need advanced means for assembling the team of people who will use and give human meaning to the information, for example through online collaboratories [46.29].

At the opposite scale extreme from nanoscience, but connected to it via global convergence, is the need for intensive whole earth monitoring [46.30]. Conceptualizing the Earth as a unified but complex natural system, increasingly affected by the artificial systems humanity builds upon it, the immense challenge is not merely tracking gross trends, but also constantly measuring and understanding local events and processes that may be cyclical, chaotic, or linear but heading in an unfavorable direction for human well-being. The data must be comprehensive but available to scientists, leaders of industry and other institutions important for the economy, and government policy-makers. Availability to journalists, social scientists, and leaders of nonprofit organizations is also important, yet proper concern must be addressed to issues of privacy and intellectual property rights. Not only must the data chart such long-term trends as climate change, but they also must be quickly available in sudden episodes of natural or man-made disasters, when the public response must be both quick and correct. Clearly, the challenges go far beyond the technical issues of sensing, storing, and analyzing the necessarily huge datasets required, but also require advancement of global civilization to a much more mature state than it has currently achieved.

46.4 Convergence in Education

To a significant degree, science and technology convergence requires overcoming the fragmentation of formal education into separate subject areas, without of course *dumbing down* the form or content of teaching. However, much more than this curricular transformation is required, because the convergence of science and technology with society will create a situation in which much education does not happen in schools, and may not even require educational institutions. While we can be optimistic, there is some reason to be concerned that the highly technical nature of nanoscience, and the fact that it often concerns phenomena that cannot directly be observed by human beings without the use of costly and complex instrumentation, may render educational convergence more difficult for this field.

46.4.1 Innovations in Formal Education

Nanoscience and nanotechnology have become so significant in the modern world that education about them is needed at many levels and in many subject areas, even in the social sciences. One contribution to the 2003 Los Angeles report showed how college students, and even high-school students, might do their own educational exercise in a social science or general science class by using the keyword *nanotechnology* while looking up books in Amazon.com, which was still in its early stages of development then [46.31]. For example, the *nanotechnology* category contained 246 books in November 2002. A student could check each one, instantly learn something about the subject merely by reading the description, and then discover convergence by looking at which additional categories the given book was assigned to. For example, 5.5% of these books were also in the *general business and investing* category, 3.7% in *electrical and optical engineering*, and 3.2% in *industrial engineering*. As of July 2016, merely entering *nanotechnology* into the Amazon.com search field for books reveals 9,861 books, 3,240 of which were also in *physics*, 2,599 in *ma-*

Table 46.2 Use of nano-related words in National Science Foundation grants

2005 Grants	BIO 1364	CISE 1382	ENG 1700	GEO 1672	MPS 2256	SBE 1163
Nanotechnology	0.6%	1.0%	10.1%	0.1%	4.0%	0.8%
Nanometer	0.1%	1.0%	3.8%	0.6%	2.9%	0.0%
Nanosecond	0.1%	0.0%	0.3%	0.0%	0.1%	0.0%
Nanoscience	0.1%	0.4%	2.6%	0.2%	1.9%	0.4%
Molecule	9.0%	1.7%	5.9%	0.7%	13.5%	0.2%
2015 Grants	**1331**	**2145**	**2673**	**1451**	**2669**	**1069**
Nanotechnology	1.1%	0.5%	4.1%	0.1%	2.8%	0.5%
Nanometer	0.3%	0.5%	2.2%	0.3%	2.7%	0.0%
Nanosecond	0.0%	0.0%	0.1%	0.0%	0.3%	0.0%
Nanoscience	0.0%	0.2%	1.0%	0.0%	1.5%	0.4%
Molecule	10.2%	0.8%	6.1%	1.8%	14.3%	0.1%

terials and material science, and 1,404 in *medical books*.

A contribution to the 2005 Hawaii conference showed how the public database of grants made by the National Science Foundation could be used for similar student assignments, since the database's search engine is versatile but easy to use, and students could read the abstracts describing the research grants, gaining some sense of satisfaction because they had found the abstracts themselves rather than merely reading the paragraphs from a teacher-assigned book [46.32]. The example given then tabulated the use of a long list of convergence-related terms across the six directorates of NSF: BIO (biology), CISE (computer science), ENG (engineering), GEO (earth sciences), MPS (math and physical sciences), and SBE (social and behavioral sciences). Table 46.2 reports the published data from fiscal year 2005 for the most nano-related keywords, followed by data from a decade later easily collated for this chapter, just as a student could do. In 2005, a student would discover, fully 10.1% of the 1700 grants made by the Engineering Directorate of NSF had the word *nanotechnology* in their online abstract.

A classroom discussion of results like those in Table 46.2 might go in any of several directions. The teacher would probably want to point out that the NSF directorates with the largest fraction of grant abstracts containing *nanotechnology*, ENG and MPS, had seen big increases in the number of grants, and new emphases in emerging areas might be partly responsible for the decline in the nanotechnology percentages. Also, the use of the term *nanometer* had not declined so much, and the word *nanotechnology* may have been used especially often when the field was new, being replaced by a variety of more specific terms in recent years. In any case the table does show connections to many fields, thus indicating the potential for nano fields to participate in convergence. Depending upon the scope of the course, and the instructor's expertise, a subsequent lecture might focus on future opportunities, such as greater involvement of nanoscience and nanotechnology in the geosciences [46.33].

Once a college or other suitable educational institution had sufficient expert staff, and material related to nanotechnology had been integrated into several different classes, then those resources could become part of a general plan to institutionalize convergence as a new paradigm for higher education more generally [46.34].

A common metaphor for science and technology convergence education is that it should produce professionals with T-*shaped expertise*, that is, possessing deep knowledge of one field, as represented by the vertical line of the T, but decent knowledge of adjacent fields, as represented by the horizontal line of the T. A more ambitious view argues that convergent professionals would need special abilities to integrate information from new fields as needed through life, and the social skills required to help other people achieve their own valuable degrees of convergence [46.35]. This perspective was widely shared at the New York convergence conference.

46.4.2 Integration of Learning with Life

The notion that learning takes place primarily in school may be one of the illusions of the 20th century that needs to be unlearned. A possibly sad fact about the field of computer science is that many of the best people in that diffuse field had already taught themselves to program computers long before they ever took a formal programming course, so people who enter the field only in college may be at a severe disadvantage. While much can be learned in school science and technology classes, students can also learn much about nature, machinery, and social relations outside educational institutions, simply by exploring their immediate environment with inquiring minds assisted by self-selected readings. Among the possibilities for transformation of

education in the 21st century are massive open online courses (MOOCs) that some innovative universities and colleges now offer to anyone anywhere [46.36]. This might be especially useful for early-level education in nanoscience and other disciplines in which the subject matter is not directly visible to the human eye. Yet it also could serve life-long learning, notably in refresher courses for people working in adjacent fields, who thus need to be brought up-to-date periodically, if they are to participate actively in convergence throughout their lives [46.37].

Much evidence indicates that passive learning is not effective, and students must be actively engaged with the aspects of reality they are currently seeking to learn about. Information technology can often provide a simulated environment to achieve greater student immersion in the more abstract fields of knowledge, a methodology currently called *cyberlearning* [46.38, p. 1008]:

> *Learning that is mediated or guided by interactions with the computer and done in an environment where both the computer and human agents can provide the help, or scaffolding needed, to foster deep understanding and masterful capabilities.*

Nanotechnology is a maturing field, in which some organizations and standards of learning have been established, but there is still much room for development of the best methods for education [46.39].

46.5 Conclusions

The 15 years of conferences and intensive written communications led to a firm consensus that science and technology convergence would involve essentially all aspects of society, potentially with revolutionary effects, and perhaps serving to protect society from some of the growing chaos in many nations and sectors of society. While information and communications technologies seem to be the most significant factor, in the context of convergence they might be reconceptualized as *cognitive technologies* [46.40]. Indeed, several contributors to the conferences suggested that our world could become a *cognitive society*, in which ordinary people were enhanced through their creative use of ubiquitous information, both individually and through their social interactions [46.41].

When the National Science Foundation surveyed the major research frontiers in 2016, from shaping the human–technology relationship, to predicting phenotypes from genotypes, and to the next quantum revolution at the nanoscale, it proclaimed that all required more convergent research [46.42, 43]:

> *The grand challenges of today – protecting human health; understanding the food, energy, water nexus; exploring the universe at all scales – will not be solved by one discipline alone. They require convergence: the merging of ideas, approaches and technologies from widely diverse fields of knowledge to stimulate innovation and discovery … Convergence blends scientific disciplines in a coordinated, reciprocal way and fosters the robust collaborations needed for successful inquiry. Convergence builds and supports creative partnerships and the creative thinking needed to address complex problems.*

It is reasonable to expect great diversity in the orientations toward convergence of scientists and engineers, and impossible to predict what consensus will emerge in the coming years. Given the uncomfortable social chaos that exists outside our universities and industrial laboratories, we could be forgiven for specializing in the narrow area that interests us most personally. Yet even from that modest perspective, the grand ideas of the convergence movement offer innumerable useful insights about the dynamic system that is science, technology, and society.

Acknowledgments. The views expressed in this essay do not necessarily represent the views of the National Science Foundation or the United States. [Legally required disclaimer]

References

46.1 M.C. Roco, W.S. Bainbridge (Eds.): *Societal Implications of Nanoscience and Nanotechnology* (Kluwer, Dordrecht 2001) p. 2

46.2 M.C. Roco, W.S. Bainbridge (Eds.): *Converging Technologies for Improving Human Performance* (Kluwer, Dordrecht 2003), p. ix

46.3 M.C. Roco, W.S. Bainbridge (Eds.): *Nanotechnology: Societal Implications – Maximizing Benefit for Humanity* (Springer, Berlin 2006)

46.4 M.C. Roco, W.S. Bainbridge (Eds.): *Nanotechnology: Societal Implications – Individual Perspectives* (Springer, Berlin 2006)

46.5 A. Nordmann: *Converging Technologies: Shaping the Future of European Societies* (European Commission, Brussels 2004) p. 14

46.6 D. Andler, S. Barthelme, B. Beckert, C. Blümel, C. Coeneu, T. Fleischer, M. Friedewald, C. Quendt, M. Rader, E. Simakova, S. Woolgar: *Converging Technologies and Their Impact on the Social Sciences and Humanities (CONTECS)* (Fraunhofer Institute, Munich 2008), pp. II–IV

46.7 O.N. Larsen: *Milestones and Millstones: Social Science at the National Science Foundation, 1945–1991* (Transaction, New Brunswick 1992)

46.8 W.S. Bainbridge: *Nanoconvergence* (Prentice-Hall, Upper Saddle River 2007)

46.9 M.C. Roco, W.S. Bainbridge, B. Tonn, G. Whitesides (Eds.): *Convergence of Knowledge, Technology and Society* (Springer, Dordrecht 2013), p. xiii

46.10 D.E. Stokes: *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Institution, Washington DC 1997)

46.11 IRGC: *White Paper on Risk Governance Toward an Integrative Approach* (International Risk Governance Council, Geneva 2006)

46.12 M.C. Roco: NBIC. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 209–226

46.13 W.S. Bainbridge: Human-technology collaboration. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 363–376

46.14 B. Tonn, D. Stiefel: The earth-scale system. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 533–544

46.15 J. Horgan: *The End of Science* (Addison-Wesley, Reading 1996)

46.16 F.J. Turner: *The Frontier in American History* (Holt, New York 1920)

46.17 V. Bush: *Science, The Endless Frontier* (U.S. Government Printing Office, Washington D.C. 1945)

46.18 W.S. Bainbridge: Twenty-first century society. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 713–727

46.19 H.-J. Choi, C. Montemagno: Convergence of nanotechnology and biotechnology. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 253–278

46.20 J.-W. Lee, M.M. Kim: Nanotechnology-neuroscience convergence. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 293–306

46.21 C.W. Stephan, W.G. Stephan: *Two Social Psychologies: An Integrative Approach* (Dorsey, Homewood 1985)

46.22 M.C. Roco: Convergence–divergence process. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 79–94

46.23 M.C. Roco: Principles and methods that facilitate convergence. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 17–41

46.24 H.C. White: *Chains of Opportunity: System Models of Mobility in Organizations* (Harvard Univ. Press, Cambridge 1970)

46.25 W.S. Bainbridge: Citizen science. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 377–390

46.26 J.C. Glenn: Collective intelligence systems. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 53–64

46.27 M. Lesk: The convergence of curation. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 95–112

46.28 T.J. Oden: Modeling and simulation. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 139–148

46.29 G.M. Olson: Collaboratories. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 391–400

46.30 B. Tonn, E. Rose, B. Hawkins: Whole-earth monitoring. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 691–710

46.31 W.S. Bainbridge: The evolution of semantic systems. In: *The Coevolution of Human Potential and Converging Technologies*, ed. by M.C. Roco, C.D. Montemagno (New York Academy of Sciences, New York 2004) pp. 150–177

46.32 W.S. Bainbridge: Transformative concepts in scientific convergence. In: *Progress in Convergence: Technologies for Human Wellbeing*, ed. by W.S. Bainbridge, M.C. Roco (New York Academy of Sciences, New York 2006) pp. 24–45

46.33 M.F. Hochella: The case for nanogeoscience. In: *Progress in Convergence: Technologies for Human Wellbeing*, ed. by W.S. Bainbridge, M.C. Roco (New York Academy of Sciences, New York 2006) pp. 108–122

46.34 R.T. Balmer: Converging technologies in higher education: Paradigm for the *new* liberal arts? In: *Progress in Convergence: Technologies for Human Wellbeing*, ed. by W.S. Bainbridge, M.C. Roco (New York Academy of Sciences, New York 2006) pp. 74–83

46.35 M.E. Gorman, J. Groves: Collaboration on converging technologies: Education and practice. In: *Managing Nano-Bio-Info-Cogno Innovations: Converging Technologies in Society*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2006)

46.36 D. Fisher: Online courses. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 1105–1118
46.37 A.C. McLaughlin, J. Feng, R. Pak: Life-long learning. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 1073–1087
46.38 J. Kolodner: Cyberlearning. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 1007–1022
46.39 J. Murday: Norms and standards of learning. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 1089–1104
46.40 J. Olds: Cognitive technology. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 227–238
46.41 A. Oliva, S. Teng: Cognitive society. In: *Handbook of Science and Technology Convergence*, ed. by W.S. Bainbridge, M.C. Roco (Springer, Berlin 2016) pp. 743–754
46.42 F.A. Cordova: *10 Big Ideas for Future NSF Investments* (National Science Foundation, Arlington 2016), http://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf
46.43 J. Mervis: NSF director unveils big ideas, Science **352**(6287), 755–756 (2016)

47. Global Perspectives of Nanotechnology Education

Kurt Winkelmann, Bharat Bhushan

The continued advancement of research and applications described in this book depends on the quality of the next generation of scientists and engineers who will lead the nanoetchnology revolution. This chapter first reviews the growth of nanotechnology education then describes the successful efforts of educators on six continents in developing the nanotechnology talents of their students. Examples of educational programs at the primary, secondary, undergraduate, and graduate levels, for teacher training, vocational education training, and for informal education of the general public are each presented.

47.1 **Growth and Trends of Nanotechnology Education** 1605
47.1.1 History of Nanotechnology Education .. 1605
47.1.2 Essential Nanotechnology Topics 1607
47.2 **Primary and Secondary Education** 1608
47.2.1 Nanotechnology Field Trip for Primary School Children in Taiwan . 1608
47.2.2 Secondary School Nanotechnology Education in Italy 1608
47.2.3 Nanotechnology School Programs in Germany 1609
47.2.4 Evaluation of Secondary School Nanotechnology Education in Europe .. 1609
47.3 **Vocational Education Training** 1610
47.3.1 Potential Areas of Growth for Vocational Education Training 1610
47.3.2 Vocational Education Training in North America 1611
47.4 **Undergraduate Education** 1611
47.4.1 Different Models of Undergraduate Education in Australia 1611
47.4.2 Undergraduate Nanoscience Education in Switzerland 1612
47.4.3 Research Experience for US Undergraduate Students 1613
47.4.4 Nanotechnology Education in South Africa 1613
47.5 **Graduate Education** 1614
47.5.1 Nanoscience Graduate Program in Switzerland 1614
47.5.2 Nanoscale Science and Technology Master's Program in Great Britain 1615
47.5.3 Laboratory Instrumentation Course in Taiwan 1615
47.5.4 Collaborative, Online Graduate Program in Israel and Europe 1616
47.6 **Teacher Professional Development** 1616
47.6.1 Teacher Training in Israel 1617
47.6.2 Teacher Training in South Korea 1617
47.6.3 Online Education for Teachers 1617
47.7 **Informal Education** 1618
47.7.1 Science Festival Activity in Israel 1618
47.7.2 NanoAventura in Brazil 1619
47.8 **Summary and Outlook** 1620
References 1622

Education and training are essential to produce a new generation of scientists, engineers, and skilled workers necessary for rapid progress in nanotechnology around the world. (For the sake of brevity in this chapter, the term *nanotechnology* encompasses all aspects of nanoscale science, engineering, and technology. Other terms are used only in cases when more specificity is necessary.) Nanotechnology education should be available to all people, not just future experts in this field. Politicians, policy makers, and voters need foundational knowledge in science, technology, engineering, and math (STEM) subjects because they make decisions that affect the growth of science and engineering, even though they do not study these technical fields. A trained nanotechnology workforce and a nanoliterate citizenry can be developed through both formal and informal education efforts, starting with the foundations in children and increasing in technical and specialized education through the university levels. Achieving these results requires teachers and education researchers to design curricula and programs that demonstrate the interdisciplinary nature of nanotechnology, including disciplines outside of the usual STEM fields. To achieve the best possible outcomes, activities should appeal to

© Springer-Verlag Berlin Heidelberg 2017
B. Bhushan (Ed.), *Springer Handbook of Nanotechnology*, DOI 10.1007/978-3-319-49347-3_47

and promote students' excitement and creativity at all educational levels: primary through graduate school, and continuing education programs, educator training, and informal science education.

With the decreasing number of people in western countries choosing to study science and engineering and with the rapid progress being made in nanotechnology, the problem of a trained workforce is expected to be acute. The question is being asked – is the traditional separation of academic disciplines into physics, chemistry, biology, and various engineering disciplines meaningful at the nanoscale? Arguably, nanotechnology education should begin by covering the broad science topics in order to facilitate new approaches to problems and new solutions based on a broader understanding of foundational science. However, business and entrepreneurial skills also are needed to transfer scientific knowledge into products, which increases the economic impact, importance, and continued support for scientific education. Therefore, it is essential that scientists and engineers no longer confine themselves strictly to the laboratory; broad and fundamental understanding of business, society, politics, and health and safety are required. To that end, it is also imperative that scientists and engineers prepare to work in a global environment, solve global problems, and address the implications of their work on a global scale. They must therefore collaborate with relevant experts to address the societal, ethical, political, and health and safety implications of their work for global humanity. Nanotechnology education should have as one focus a broad, interdisciplinary approach to science and engineering at the nanoscale. More on this subject can be found in the recently published book *Global Perspectives of Nanoscale Science and Engineering Education* [47.1].

To increase the pool of students interested in advanced education in science and technology, educators should convey STEM subjects as an exciting and creative endeavor from their first introduction. Innovative foundational education in schools and informal instruction opportunities can help to accomplish this goal. Since the mid-1990s, there has been a global push for better STEM education, particularly in the K-12 arena. The goal is to increase the number of people interested in pursuing advanced education in STEM fields, and to ensure sufficiently educated workers for STEM jobs. This has resulted in a generation of workers with content knowledge and enthusiasm, but many with limited ability to apply that knowledge to real-world scenarios and problems, and limited experience with thinking critically and creatively.

To address this challenge, there has been a drive to make technical fields more attractive to students by making STEM topics more relevant to students' interests. One approach, begun around 2010, includes the arts in STEM education, an initiative called STEAM, in order to encourage not only science content, but creativity and critical thinking skills derived from arts-based education as well. This approach is especially popular in the Republic of Korea (South Korea) [47.2, 3]. These arts-based skills are essential for translating discrete content knowledge into broad applications. If this STEAM education approach can be applied to nanotechnology, as well as more macroscale science and engineering fields, the beginnings of an interdisciplinary, broad education can be found [47.4]. However, interdisciplinary curricula relevant for nanotechnology need to be developed. This requires revamping the STEAM education model for nanotechnology, the development of new courses, and course material including textbooks and instruction manuals, and training of new instructors.

One might ask why a global perspective is needed for nanotechnology education. After all, scientific results should be reproducible around the world. Nanoscale phenomena in Brazil are the same as in Israel or Taiwan. A nanomaterial synthesis experiment conducted by scientists of either European or African heritage will have the same result. This is obvious to scientists; on the other hand, educators recognize that learning is a human endeavor so perspectives of students can differ due to their culture and environment. Students learn more when they care about the subject. Educators can motivate students to learn by designing activities that interest students [47.5]. For example, a student will be more interested in nanotechnology if the student identifies with aspects of people (gender, ethnicity, country of origin, etc.) associated with it. The many uses of nanotechnology that predate the scientific investigation of nanoscale features in the 20th century provide educators with culturally relevant examples that can increase student interest. Connections exist between a nanotechnology topic and many cultures since different cultures interacted and influenced each other throughout history. In addition, adding elements of history has the benefit of making future scientists more well-rounded in their education. History provides a cultural framework, making science more regionally relevant. By linking a cultural past with a cultural present, history makes science more enticing and relatable, perhaps sparking an interest in science among those students who are initially wary of studying it. A brief history of Damascus steel illustrates this point.

Damascus steel is commonly cited as an early use of nanomaterials [47.6]. Swords were made of this steel in Damascus, the current capital of Syria, until the 18th century [47.7]. The quality of the swords was unmatched during that time, and the particular iron ingots

used to forge the steel, known as wootz, were imported from India. In the 19th century, Michael Faraday studied Damascus steel. As the son of a blacksmith, the scientific study of what made Damascus steel superior in quality was relevant to Faraday's cultural framework. Later in 1857, as part of an unrelated study, he reported that metal colloids could have different colors depending on their particle size [47.8]. This is among the earliest observations that invoked nanostructures as an explanation. This foundational study in nanotechnology sparked further study, and we now know that carbon nanotubes and other nanostructures enhance the strength and sharpness of Damascus steel.

Development of nanotechnology curricula for primary, secondary, two-year vocational, traditional four-year university, and graduate education programs, continuing education for practicing teachers, and informal education, must continue to keep pace with research. It is hoped that this chapter provides a useful introduction to nanotechnology education for researchers with an interest in this subject and a survey of nanotechnology education as practiced in countries around the world. Educators and researchers can learn from the efforts of their peers and colleagues around the world who are pursuing the same goal of improving nanotechnology education.

Following Sect. 47.1, the remaining sections of this chapter provide summaries of nanotechnology education activities from around the world, with an emphasis on those occurring outside the United States. This approach was chosen because US-based programs are the most widely known, so there is less of a need to present them here. Chapter space limitations allow for only a somewhat random sampling, rather than a comprehensive review, of the many outstanding nanotechnology education activities. Long-standing programs are given preference for inclusion over short-term research studies of nanotechnology education. Information about the programs and curriculum materials is drawn primarily from articles published in the peer-reviewed academic literature and all referenced articles are written in English. Each chapter section addresses nanotechnology activities appropriate for a certain educational environment.

Not all of the programs described herein are still in existence. Nanotechnology courses may be absorbed into existing, traditional STEM courses. National priorities for STEM education can change, drawing funding from nanotechnology education to other subjects. Individual instructors may not sustain an activity that they created. This does not diminish the value of the successful programs, courses, and activities described here.

47.1 Growth and Trends of Nanotechnology Education

At the beginning of the 21st century, educators and policy makers assessed the need for including nanotechnology among the topics taught in schools [47.9, 10] as well as the funding mechanism to address that need [47.11, 12]. Due to the investment in the National Nanotechnology Initiative by President William J. Clinton in 2001, the United States led many of the early efforts to establish centers for nanotechnology research and education. Other countries soon established their own nanotechnology programs. The appearance of nanotechnology in school curricula is increasing around the world, as the examples cited in this chapter will show. In addition, readers may consult several reviews [47.13–15], books about this subject [47.1, 16–18], and ongoing published articles in the Journal of Nano Education.

47.1.1 History of Nanotechnology Education

Development of classroom nanotechnology activities preceded the establishment of these national programs. For example, Fig. 47.1 shows the number of publications related to nanotechnology appearing in the Journal of Chemical Education between the years of 1991 and 2015. Educators were already disseminating nanotechnology experiments such as the synthesis of ferrofluids [47.19] and demonstrating the use of scanning probe microscopy in the classroom [47.20]. Data was gathered by searching the website of the Journal of Chemical Education for the term *nano* in the titles and abstracts of articles. The resulting list of articles was reviewed to distinguish articles related to nanotechnology (shown in Fig. 47.1) from those that included other uses of *nano*, such as nanosecond spectroscopy or $NaNO_3$ (sodium nitrate).

A search for *nanotechnology education* within the SciFinder database provided by the Chemical Abstracts Service finds similar trends in other publications (excluding patents and including conference proceedings). SciFinder indexes not just chemistry journals but also those in materials science, physics, biology, and medicine. Figure 47.2 shows a similar trend for publication of nanotechnology education articles during the past twenty years. In fact, the data shown in

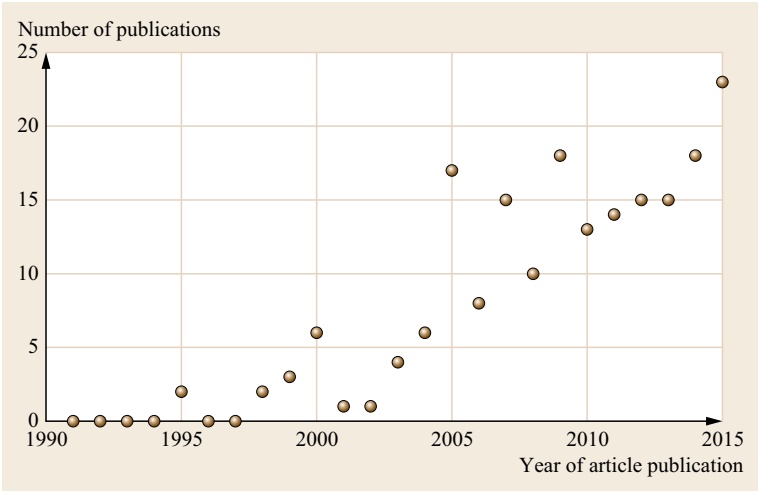


Fig. 47.1 Yearly number of nanotechnology-related publications in the Journal of Chemical Education

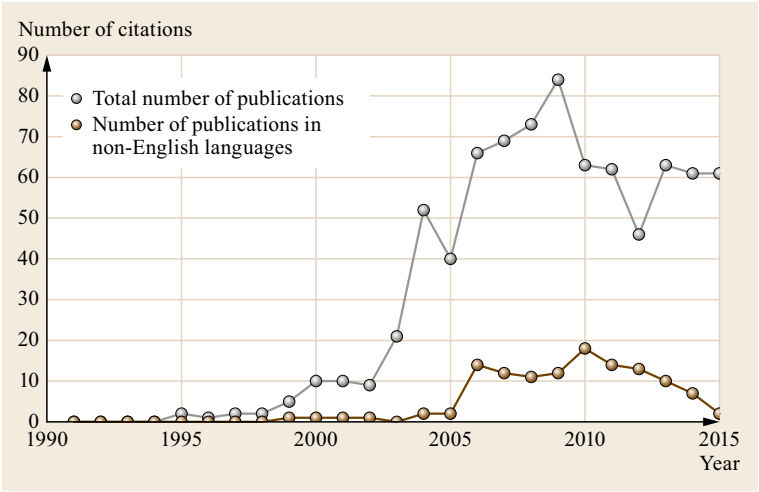


Fig. 47.2 Yearly number of nanotechnology education publications as indexed by the Chemical Abstract Service

Fig. 47.2 is certainly an undercounting of the number of publications because SciFinder does not list consistently relevant articles from smaller journals, such as the Journal of Nano Education. This may account for the apparent plateau in the number of published articles within the past six years. SciFinder does not index thoroughly journals of non-STEM-related subjects either, so articles about teaching nanotechnology history or ethics may not be included. Fig. 47.2 also shows that authors from around the world contribute to the academic literature of nanotechnology education research and practice. While most journals require authors to publish in English regardless of their native language, many nanotechnology education articles are published in other languages. The most common non-English languages are German (54 articles total) and Chinese (40 articles); articles written in a total of eight other languages constitute just 33 articles since 1995. Since 1999, 14% of the nanotechnology education articles accessible through SciFinder were written in a non-English language. It is clear that nanotechnology education publications are becoming more common and include contributions from many countries.

Implementation of a new topic requires a modification of the existing school curriculum, which raises questions. What is the goal of introducing nanotechnology to students? Which nanotechnology topics should be taught? Reviews of nanotechnology education activities published in the literature find two reasons that justify the inclusion of nanotechnology [47.14, 21]. First, from the US perspective, there is a looming shortage of nanotechnology workers so educators should prepare more students to fill these jobs of the future [47.9]. Fewer students are pursuing STEM degrees. Nanotechnology has applications in diverse fields, from electronics to pharmaceuticals to consumer goods, so learning

about nanotechnology could be a new way to get students interested in STEM subjects. To achieve the goal of a more nano-educated workforce, schools can add nanotechnology topics to existing STEM courses, such as a nanoparticle synthesis experiment to a chemistry lab course. Scientific literacy and more specifically nanoliteracy, is the second major reason cited for teaching nanotechnology in secondary schools. Due to its increasing number of applications, the general public should learn about nanotechnology so that they can have informed opinions about policies that involve the use of nanomaterials and related technologies. In order to introduce nanotechnology to as broad a student audience as possible, educators could add nanotechnology topics to both STEM and non-STEM classes. For instance, a discussion about the ethics of using nanotechnology could occur in a high school civics class.

47.1.2 Essential Nanotechnology Topics

Several research teams have addressed the question of which nanotechnology topics should be taught [47.22]. These are listed in Table 47.1. *The Big Ideas of Nanoscale Science and Engineering* by *Stevens* et al. [47.23] describes the following as the nine most important nanotechnology topics to be taught in US grades 7–12 (middle and high school):

- Size and scale
- Structure of matter
- Forces and interactions
- Quantum effects
- Size-dependent properties
- Self-assembly
- Tools and instrumentation
- Models and simulations
- Science, technology, and society.

A separate list for university-level students was also prepared. The authors gathered suggestions for list items from educators in US middle and high schools and higher education. Both groups gave similar feedback. The Big Ideas was the first and most well-known guide for educators in their development of new activities. A second survey of experts was conducted by *Huang* et al. [47.24] to determine the nanotechnology topics appropriate for Taiwanese primary school children. That list includes:

- Nanotechnology definitions
- Nanoscale features
- Nanophenomena in the natural world
- Nanomaterials
- The development of nanotechnology.

Most recently, *Sakhnini* and *Blonder* [47.25] reported a study of essential nanotechnology concepts to be taught to high school students, as chosen by Israeli nanotechnology researchers and high school science teachers. Topics are:

- Size-dependent properties
- Innovations and applications of nanotechnology
- Size and scale
- Characterization methods
- Functionality
- Classification of nanomaterials

Table 47.1 Important ideas about nanotechnology from four surveys of different groups of subject matter experts, organized by common features on each list

Primary School (Taiwan, 2011)	Middle and High School (US, 2009)	High School (Israel, 2015)	Undergraduate (US, 2009)
Nanotechnology definitions	Size and scale	Size and scale	Size and scale
Nanoscale features	Structure of matter Size-dependent properties Quantum effects	Size-dependent properties	Size-dependent properties Surface to volume ratio Quantum mechanics
	Forces and interactions	Functionality	Surface-dominated behavior
	Self-assembly		Self-assembly
Nanophenomena in the natural world		Innovations and applications of nanotechnology	
The development of nanotechnology	Tools and instrumentation	Characterization methods Fabrication approaches of nanomaterials	Tools and instrumentation/ characterization
	Science, technology, and society	The making of nanotechnology	Societal impact/ public education
Nanomaterials		Classification of nanomaterials	
	Models and simulations		Models and simulations

- Fabrication approaches of nanomaterials
- The making of nanotechnology.

Differences among these lists are likely due to the various ages of the target audiences (primary, middle, secondary, and undergraduate students). Cultural differences play a role and, given the rapid development of nanotechnology, the fact that the surveys were conducted in different years (2009, 2011, and 2015) may also contribute.

Still, similarities are obvious among the three studies. Entries in Table 47.1 are presented so that each row contains similar or related topics. Definitions of size and scale appear in all four lists. Different studies grouped nanoscale phenomena in different ways, either under a single category (e.g., nanoscale features or size-dependent properties) or separated into more specific items. It was assumed that undergraduates already understood the structure of matter so it was not included in their list. Ideas such as the impact of nanotechnology on society, the tools used to observe and measure nanoscale features, and applications of nanomaterials all appear in each list but these ideas overlap in a nonsimple way.

Of course, knowing the goal of introducing nanotechnology into a school curriculum and the general topics to teach are only the beginning of the process. Additional questions naturally follow, such as identifying the learning objectives, prerequisite knowledge required, and the appropriate pedagogy to use [47.14]. While critically important, these questions cannot be easily addressed in the present book chapter. Readers are urged to consult the STEM education research literature for a thorough presentation.

47.2 Primary and Secondary Education

Policies for funding primary and secondary education vary greatly around the world. Local, regional, and national governments can fund programs and individual educators can develop activities for their own students. Since the education of children is a high priority for every country, there is much interest in evaluating the effectiveness of nanotechnology education programs. The following examples of programs taking place around the world describe both top-down and bottom-up organizing approaches to secondary school education in nanotechnology and its overall evaluation. In preparing this chapter, little information was found regarding nanotechnology education in primary schools. Although many educators stated the need for this, there are few published accounts of nanotechnology classroom activities for young children. However, many educators describe outreach and informal learning programs for this audience. Readers will find informal learning discussed in Sect. 47.7 of this chapter.

47.2.1 Nanotechnology Field Trip for Primary School Children in Taiwan

Despite teaching nanotechnology in its primary schools, students in Taiwan lack an awareness of the subject [47.24]. To address this, educators at the Center for Nanotechnology of Ilan University collaborated with a local cable TV station and the county government to create a nanotechnology field trip to the university for primary school students [47.26]. Lesson plans allowed undergraduate students volunteering for a service learning program at Ilan University to help promote nanotechnology to sixth grade students attending four local schools. The half-day event focused on nanotechnology definitions, size-dependent effects, occurrences in nature, and applications.

Activities were designed to be interactive and entertaining, similar to activities used in informal education. This is especially appropriate since the children were attending the event away from their normal classroom. Although it was part of their day of formal schooling, it was also supposed to be a fun field trip. Students listened to a multimedia lecture that gave an overview of nanotechnology. They performed an experiment that allowed them to observe the effect of a hydrophobic surface on a water droplet. This was presented in the context of the nanoscale features that give a lotus leaf similar properties. Students played nanotechnology-themed games, visited a nanotechnology exhibition, and participated in other group activities.

Students completed a pre- and post-event quiz to measure changes in content knowledge. A comparison of these quiz scores showed that students gained approximately five points out of 15 questions (one point per question) on the quiz after participating in the activity. Among the four schools, average improvement ranged from gains of 3.4 points to 5.4 points.

47.2.2 Secondary School Nanotechnology Education in Italy

A project conducted at the Scientific-Technological Lyceum in northern Italy involved both secondary and primary school children [47.27]. Goals of the project

were to improve students' content knowledge about chemistry and nanotechnology, and enhance their communication and teamwork skills. After attending a seminar to learn about nanotechnology, a class of secondary school students developed a slide show presentation to show primary school students an application of chemical nanotechnology. Students chose to investigate a chemiluminescent sensor for detecting calcium ions in drinking water. Due to the cutting-edge nature of nanotechnology, no description of the subject appeared in the textbooks of either group of students. Therefore, the secondary school students had to research the topic using other sources. Students also performed laboratory exercises in which they prepared standards of calcium ion solutions and observed the chemiluminescence. Each student assumed a specific role within the team as they performed their research, created their own slide format, prepared the slides, and presented the research to the primary school children.

Students first chose a topic (chemical sensing of Ca^{2+} in drinking water), developed questions about nanotechnology to investigate, then answered the questions. In the second stage, students created a plan for teams to perform their research. Using online resources, students worked together to conduct their research, including the selection of images and information to include. Following this, they prepared their presentations and developed their own rubric to be used for assessing their work. Groups then gave their presentations to the primary school audience. The last stage was evaluating the groups' presentations by their peers. Students also reflected upon their own contributions to the project by answering a series of open-ended questions. They viewed the experience of researching, organizing, and presenting the information as worthwhile. Students answered questions about nanotechnology before the initial seminars, immediately after the project, and several months later. Students showed significant gains in content knowledge after completing the activity and they retained most of their knowledge over time.

47.2.3 Nanotechnology School Programs in Germany

Germany provides examples of support for nanotechnology education through various means. One common strategy for introducing students to nanotechnology is to bring the laboratory to them. This is the idea behind nanoTruck and Nanoshuttle [47.28]. For ten years, the nanoTruck program provided a mobile exhibit and teacher support within a two-story tractor trailer. Exhibits within the nanoTruck included examples of nanomaterials and their applications in consumer goods, transportation, electronics, building materials, and medicine. Visitors experienced multimedia presentations, hands-on experiments, and dozens of exhibits that helped them understand the basic science and how future careers will involve nanotechnology. Teachers could attend workshops where they learn about nanotechnology prior to a scheduled nanoTruck visit to their school. One million visitors participated in this program. Nanoshuttle is a current mobile exhibit touring the state of Bavaria. Like nanoTruck, students can perform experiments, learn about nanotechnology applications, and visualize nanostructures using advanced microscopes. The German Chemical Industry Association funds the dissemination of NanoSchoolBox, a kit containing materials for over a dozen experiments designed for high school chemistry and physics students. Online videos provide instructions and background information to students as they perform experiments individually or in groups. Other organizations, such as the Association of German Engineers and the Fund for the Chemical Industry, provide other electronic resources for students.

47.2.4 Evaluation of Secondary School Nanotechnology Education in Europe

The European Union's Nanotechnology Education for Industry and Society (NanoEIS) program recently published a report of the twelve most widely implemented and successful secondary school nanotechnology programs in EU countries as part of its 7th Framework Programme [47.29]. It indicates that compulsory activities were more common among successful programs compared to voluntary activities. Many programs lack hands-on activities for students and there is low community involvement outside of the school. There also existed a great variability in the involvement of industry and higher education as well as the degree to which the nanotechnology subject was integrated within existing science courses (or taught separately). The lack of hands-on experimentation, virtual environments, and involvement of other stakeholders were cited as deficiencies to be addressed.

Another NanoEIS report examined the extent that local secondary school nanotechnology programs and other tools encourage students to study science and technology at the university level [47.30]. As with the 2014 report [47.29], this report noted the general absence of university involvement in the secondary school nanotechnology programs. This leads to teachers not knowing what students should learn in high school if they want to study nanotechnology as undergraduates. In fact, the study found that in general, schools did not provide enough assistance helping students transition from secondary school to university. It found that the

exposure to nanotechnology among secondary school students was low, with a maximum of 20% of students in Ireland. Nanotechnology programs were rare and short-lived in all the EU countries studied. Once the program was initiated, it was expected that teachers should continue it on their own. This is likely a common problem in many regions of the world, as teachers are saddled with conflicting responsibilities to teach more information in different ways to more students. Despite these shortcomings, the report notes that most local nanotechnology programs were less than five years old. Some growing pains are common for young initiatives. It is hoped, even expected, that the dissemination of best practices of nanotechnology (and, in general, STEM) education along with the increasing need for nanotechnology workers will promote higher quality programs for secondary school students.

47.3 Vocational Education Training

Developing a skilled nanotechnology workforce requires more than talented K-12 teachers, new graduate and undergraduate university programs, and a knowledgeable public. Vocational education is needed to train technicians who need specific skills for industry. The audience for vocational education includes young people who do not attend a traditional four-year college or older workers who desire retraining so that they can advance in their current job or find a new job in the nanotechnology field. Specific skill needs of industry can change quickly as new products are developed or become obsolete and as new government regulations affect manufacturing processes. Universities are usually not nimble enough to keep up with these changes. Vocational education training (VET) can fill this gap. It should also be noted that industry and government seek nontechnical workers with an understanding of nanotechnology in order to devise policies governing waste, toxicology, risk assessment, and legal aspects of nanomaterials [47.31].

47.3.1 Potential Areas of Growth for Vocational Education Training

The approach to overseeing a nation's education program varies by country. In the US, the state and local governments play a larger role over the education priorities than the federal government, although higher education institutions often rely on indirect federal funding. In other countries, such as Germany, the national government takes a more active role. In order to develop an effective, single VET curriculum, the German Federal Ministry of Education and Research (BMBF) assessed the current and anticipated skills needed by nanotechnology workers. Following this, the BMBF began to modernize the existing curriculum for trades related to nanotechnology and set up new interdisciplinary degree programs and continuing education programs. They also work to promote a strong relationship between VET schools and the businesses that employ those graduates. Government plays a role in promoting relevant employment opportunities that VET can provide to secondary school students and current workers [47.32].

Other countries should follow this example. An assessment of relevant skills needed for industrial nanotechnology workers would provide guidance to schools seeking to create and maintain effective VET programs. Without ongoing assessment of these needs, educators will only react, often slowly, rather than be proactive in designing curricula to better suit their students.

Attendees at a 2014 National Science Foundation (NSF)-sponsored workshop, Nanoscale Science and Engineering Education: The Next Steps, identified several areas of improvement for vocational training [47.33]. Academic researchers and educators from many countries, policy makers, and representatives from other US funding organizations noted that a best practices manual that builds on existing VET programs could contribute to the effort to create nationwide programs that are sufficiently rigid to ensure a consistent quality of education and training in all areas, but sufficiently flexible to account for local and regional technical needs. General recommendations from the NSF workshop recognized the need to promote the societal impacts and interdisciplinary nature of nanotechnology education as a way to increase diversity in the student population.

A third area of growth is the dissemination of research studies and descriptions of successful VET programs. There are very few reports of nanotechnology programs for vocational training in the literature. This could be due to a lack of attention paid to vocational education by education researchers or the relatively low numbers of nanotechnology vocational programs compared to bachelor's and graduate degree programs. Whatever the cause, this should become less of a problem as the need for technicians of nanotechnology-related instrumentation and infrastructure increases.

47.3.2 Vocational Education Training in North America

There are two strong examples of vocational-level programs developed by postsecondary education institutions in the United States. The first example is the one-semester, 18 credit-hour capstone style program pioneered by Pennsylvania State University (Penn State). This program is intended to complete a student's vocational or bachelor's degree at a home institution by drawing traditional, foundational education disciplines together and integrating nanotechnology into them. In addition to this capstone program, Penn State developed the Nanotechnology Applications and Career Knowledge (NACK) network to develop a standardized curriculum to assist vocational institutions in those foundations, as well as to train practitioners.

The second example is the program developed by the Dakota County Technical College (DCTC) in Rosemount, Minnesota [47.34]. Inspired by information at a National Career Path Network conference, DCTC administrators began asking local industry (including such businesses as 3M, Honeywell, Synovis, Hysitron, General Mills, Seagate, Medtronic, among other, smaller companies) what they needed in vocational-level training for students. As a result of their inquiries, DCTC developed their Associate of Applied Science (AAS) degree Nanoscience Technician program in 2004, the first such two-year program in the US. This program takes a multidisciplinary approach to teaching not only tool usage and practical applications, but critical thinking capabilities as well. Developed from this program, and with funding from the NSF, DCTC created Nano-Link in 2008 to guide and mentor schools in the evaluation and validation of nanotechnology content.

The Northern Alberta Institute of Technology (NAIT) offers the first Nanotechnology Systems degree program in Canada [47.35]. Many of the 24 courses students complete are traditional subjects (e.g., chemistry, physics, biology, math, and materials science), some of which have an emphasis on nanotechnology. They also complete courses in manufacturing, project management, cleanroom practices, and a capstone project. Nanotechnology courses cover synthesis, characterization, and lithography. Students gain experience working in a cleanroom and in instrumentation such as a scanning electron microscope. They also visit industrial nanotechnology facilities in the area. The program advertises that students will have skills needed in the pharmaceuticals, electronics, manufacturing, and energy industries.

47.4 Undergraduate Education

Advancing nanotechnology skills and understanding requires comprehensive inclusion in four-year undergraduate programs. Existing nanotechnology courses and curricula can be incorporated into traditional science and engineering majors. Additionally, interdisciplinary nanotechnology majors may provide more focused education for interested students. Research opportunities may enhance the experiences of students both in terms of content and cultural awareness, and may provide impetus to continue to graduate school.

47.4.1 Different Models of Undergraduate Education in Australia

Two Australian schools are among the first in the world to offer Bachelor of Science degrees specializing in nanotechnology. Flinders University began offering its degree program in 2000 and the University of New South Wales started its degree program in 2002 [47.36]. Both degrees were developed by faculty in multiple disciplines to teach the relevant nanotechnology course material, important business skills, and soft skills such as communication and teamwork. Enrollment by Australian college students in science was not growing as fast as expected and nanotechnology was viewed as a potential way to attract more students to this cutting-edge field.

Although the goals and development of the two degrees have similarities, they are based on different models. Flinders University created a nanotechnology degree program housed entirely within a single school, the School of Chemistry, Physics, and Earth Science, with no others involved. This is the traditional and most common scenario for introducing new degrees. It allows more tight control over the curriculum by a single group of faculty and administrators. There are fewer stakeholders involved, which simplifies the decision making process. All students and faculty involved in the single-school program share a common set of scientific vocabulary and share a common academic culture. However, the topics covered within the degree program are limited by the expertise, space, and equipment of the faculty and infrastructure of that particular school. It also diminishes the scope of what is truly a widely interdisciplinary field.

The University of New South Wales (UNSW) offers its nanotechnology degree through four schools: Chemistry, Physics, Materials Science, and Biotech-

nology and Biomolecular Science. This gives students a broader perspective of what the subject nanotechnology includes, especially biology (Flinders University does require students to enroll in a biology course but biology faculty are not involved in administering the program). More resources are available and more input from different disciplines can lead to better decisions. Disadvantages of this approach are the difficulty of administering the program and keeping independent schools working well together. Academic disciplines are traditionally located in separate departments or schools so it can seem somewhat unnatural for them to cooperate closely. Students belong both to many schools and also to none, so they lack a common cohort of peers.

A third school, the University of Technology, Sydney, followed a different path by creating its nanotechnology degree program in a separate school. It is controlled by a single administrative unit and students can identify a single school to which they belong – two of the benefits of a single-school approach. Although the degree program is housed separately from other disciplines, faculty do not teach duplicates of existing courses, such as materials science, to the nanotechnology students. Rather, the curriculum calls upon other schools to provide the many core STEM classes that the nanotechnology students need to complete. This is clearly efficient but the program coordinators must ensure that faculty teaching those other subjects (who are affiliated with other schools) continue to teach topics important to the nanotechnology curriculum. This problem and challenge of creating a new, independent administrative entity is solved with financial support from the university. Funding, in turn, represents a strong commitment by the university to the long-term success of the program.

Development of any nanotechnology degree, including the three discussed here, requires answering the question of what topics to include. Nanotechnology is a broad field with each of its constituent fields of chemistry, physics, biology, and engineering having a subfield of nanochemistry, nanophysics, etc. Not all students who are interested in nanotechnology will necessarily want to study all of these subjects, even if time were available. Both Flinders and UNSW elected to create a common first-year curriculum for all nanotechnology students that emphasizes basic sciences and mathematics. Students can pursue more specialized tracks with more specialized classes in engineering, math, and science during subsequent years. At Flinders University, students choose between biodevices and nanostructures starting in their second year. Students at UNSW can select an emphasis in bionanotechnology, molecular electronics, or nanomaterials beginning in their third year. Many classes appear in the curricula of both tracks. The University of Technology, Sydney, follows a more general curriculum for its students with individual classes, such as nanoscale electronics and bionanotechnology, which Flinders and UNSW students can study in more depth.

A second challenge is how to offer all classes for a new degree program to students, given the practical and financial constraints that are encountered at every university. Students enroll in new classes created especially for the nanotechnology degree programs but these new courses account for only a small fraction (between ten and 17%) of the total number of credit hours that they must complete. For all three universities, the students begin their nanotechnology courses during the first year, some during their first semester in the program. This is important for several reasons. First, students may enter a program without knowing exactly what nanotechnology is, so it is important to correct any misconceptions as soon as possible. Students at Flinders and UNSW must choose a concentration for their degree during their second or third year respectively, so they need to know about the different aspects of nanotechnology as early as possible. Finally, faculty felt that nanotechnology could be a subject that increased student interest in science so they should not wait to introduce it.

Despite the very different designs of their nanotechnology programs, students at both Flinders and UNSW graduate and go on to successful careers in a variety of fields, including those outside of science. They also seek graduate degrees at a higher rate than students from other science programs.

47.4.2 Undergraduate Nanoscience Education in Switzerland

To address the need for a more interdisciplinary approach to nanotechnology education, the University of Basel Swiss Nanoscience Institute (SNI) offers a bachelor's degree in Nanoscience that prepares undergraduates for a graduate program in Nanoscience or in a traditional science field [47.37]. The flexible curriculum includes chemistry, physics, and biology courses. Additional classes teach students communication skills and discuss ethical dilemmas encountered in nanotechnology.

During the first semester of their Nanoscience program, undergraduate students enroll in Nano I, a seminar course that introduces them to nanotechnology research performed at the University of Basel. In the Nano II class taken in the second semester, students attend site visits to Swiss companies that are active in nanotechnology manufacturing and research so that they

can see the application of the course material learned in class. The remainder of their first-year courses are math, physics, chemistry, and biology. Second-year courses emphasize nanotechnology and physics. Students become more involved in a particular field of nanotechnology (e.g., nanophysics). These classes prepare students for research experiences gained in the third year. At that point, students enroll in a series of short (one to three week) courses taught at the University of Basel or partnering institutions. Students participate in ongoing research with faculty and graduate students. These research courses prepare students for graduate-level research, which they can pursue at the SNI as well (graduate nanoscience programs at the University of Basel are discussed in the next section). As with students in many STEM programs, they often cite the research experience as a highlight of their academic career. Students form a strong, supportive network among themselves and classes are often small. The three-year curriculum is rigorous so roughly half of the students switch to other science programs before graduating with a bachelor's degree. Those who go on to study for a master's degree are well prepared. Over the past five years, 76–100% of these students entering the Nanoscience master's degree have successfully completed the program.

47.4.3 Research Experience for US Undergraduate Students

Besides individual nanotechnology topics incorporated into traditional courses, separate nanotechnology courses, and academic major and minor programs, undergraduates should have the opportunity to perform undergraduate research. This is known to be an effective means for students to learn about STEM careers and remain interested in STEM subjects [47.38–41]. The US NSF funds a Research Experience for Undergraduates (REU), bringing 80–90 students from across US universities together each year at 14 National Nanotechnology Infrastructure Network (NNIN) sites for a summer research experience [47.42]. This provides students with nanotechnology-specific training and offers them an opportunity to present their work and develop professional connections for advanced research opportunities.

The NNIN REU program also has an international component [47.42]. The formation for such a program was in response to the recognition that competence in STEM fields is necessary but not sufficient. Students need to develop the ability to work with colleagues from other countries and cultures and address global STEM problems. Beginning in 2008, the most highly qualified REU participants were offered the opportunity to apply for an international research (i-REU) experience during the subsequent summer. These students perform nanotechnology research in the laboratories of NNIN partners located in Japan, France, Germany, Belgium, and The Netherlands. Students from previous years and NNIN staff members create online resources to help future cohorts. Students travel in groups to the partner institutions and spend eleven weeks performing research. The NNIN program funds their travel, stipends, and accommodations while the host university covers all research costs. Of the 40 to 45 applications, 12 to 15 students are selected each year for a total of 109 completed i-REU appointments. Most of those who have graduated went on to graduate school and a significant portion of those graduate students have received additional research awards. Approximately half of the i-REU participants are women and about a third attended smaller, nonresearch colleges or universities.

Given that the goal of the i-REU program is to develop culturally aware STEM students, the program was evaluated using an assessment tool from the Cultural Intelligence Center that measures cultural intelligence. This self-assessment survey evaluates a person's interest and enthusiasm for multicultural experiences; knowledge of business practices, values, and behaviors observed in different cultures; the ability to understand the person's own cultural knowledge and plan for multicultural encounters accordingly; and the ability to modify verbal and nonverbal communications in different cultural settings. Participants completed the survey prior to and after completing the i-REU program. They received feedback about their scores on the preprogram survey and these results can be used by the program leaders to help better prepare participants for their international research experience. Comparing pre- and post-summer tests, results improved from moderate to high levels in three of the four areas evaluated (knowledge, strategy, and action). Participants' scores for enthusiasm/interest initially were very high and increased slightly after they completed their i-REU appointments. It is clear that this experience greatly benefited the students by raising their cultural awareness.

47.4.4 Nanotechnology Education in South Africa

Nanotechnology research leads nanotechnology education in all countries but college-level curriculum materials, courses, and programs for nanotechnology are now well known in developed countries. This is not the case in the developing world. The need to train nanotechnology workers is still acute, especially given that nanotechnology can have a relatively greater

impact on the developing world by providing clean water, better medicines, and sustainable energy. In 2009, a senior-level, introductory nanotechnology course was taught in the School of Chemistry at the University of KwaZulu-Natal (UKZN) Pietermaritzburg, a leading South African university for chemistry students [47.43]. The course content was based on a course offered in the United States and it was taught by Dr. Brian Augustine, a professor at James Madison University. This arrangement was made possible through the Fulbright Program.

Students at UKZN can major in chemistry and chemical technology and enroll in an honors program. The curriculum is typical of any college-level chemistry degree. A new nanotechnology curriculum was introduced into a materials course taught in the chemical technology program. Students learned about quantum mechanics, material properties, characterization methods, lithography, self-assembly, and other topics. Case studies were used to teach DNA self-assembly, carbon nanomaterials, and nano-imprint lithography. For the laboratory portion of the course, students synthesized gold nanoparticles, imaged surfaces with an atomic force microscope (AFM), viewed AFM tips with an electron microscope, synthesized and determined the critical micelle concentration of amphiphilic molecules, and fabricated PDMS microcontact printing masters, which they then used to create microstructures on gold surfaces.

UKZN possesses chemical research instrumentation appropriate for a Ph.D.-granting university. Interestingly, a lack of basic, relatively inexpensive laboratory supplies was a challenge. Chemicals were also more expensive than in the United States, in part due to shipping fees. The class was taught in English and students had to write laboratory reports in English, but this did not have a significant effect on grades for students who learned English as a second language. Student performance in the class (as in other UKZN classes) correlated strongly with race. Low income and poor living conditions continue for black South Africans even after the official end of apartheid. This has a negative impact on their educational preparation for college. Since South Africa is more prosperous, these problems could be even more significant in other developing countries.

Three assessment instruments were used to evaluate students' learning and attitudes. The Nanotechnology Learning Outcomes Survey developed at Northwestern University was given at the beginning and end of the semester in order to measure students' attitudes about nanotechnology. Students completed a longer and more challenging multiple choice exam to determine their knowledge of nanotechnology at the beginning of the semester and as part of the third course exam. A size and scale assessment was included in the third exam also. Students in the new nanotechnology course at UKZN responded to the size and scale activity and the Nanotechnology Learning Outcomes Survey similarly to nonscience majors at Northwestern University in the US. The UKZN students made significant pre- to postsemester gains on the nanotechnology knowledge assessment. Students improved their knowledge related to general science questions, not just those addressing nanotechnology, suggesting that a nanotechnology course can help students learn more fundamental science information as well.

47.5 Graduate Education

The previously mentioned 2014 NSF workshop identified deficiencies in graduate training for nanotechnology including interdisciplinary education and diversity [47.33]. As well, the workshop noted that graduate programs will improve as the pipeline of students funneling into the programs are better equipped with foundational nanotechnology knowledge due to strengthening of nanotechnology education in the secondary and undergraduate levels. The University of California, Santa Barbara (UCSB), is addressing the deficiencies in interdisciplinary focus with a unique research opportunity. Nanotechnology graduate students are awarded fellowships and are embedded within social science and humanities social science projects. This arrangement reveals to students the connections between the laboratory and broader societal forces. Schools around the world take different approaches to the format, structure, and curriculum of graduate nanotechnology degrees. All seek to provide a program that addresses the interdisciplinarity of nanotechnology.

47.5.1 Nanoscience Graduate Program in Switzerland

The University of Basel's SNI, mentioned earlier, offers a Nanoscience bachelor's degree as well as both a master's and Ph.D. in Nanoscience [47.37]. The University of Basel was the first European university to offer both levels of nanoscience degrees. A master's program is required for students who wish to eventually earn a Ph.D. Students with an undergraduate degree in a traditional field of science pursuing the master's degree in Nanoscience may need to complete some undergraduate courses in order to qualify for entrance in

the program. Like the SNI undergraduate degree, the master's degree is designed to provide an interdisciplinary education. Students devote more than half of their efforts to studying nanoscience within the field of either molecular biology, physics, or chemistry, and must complete classes in fields outside of their chosen discipline. Most of their research is devoted to the completion of a thesis project but they also complete two smaller projects in different disciplines. The program lasts for a year and a half.

The Ph.D. program is unusual in its design. Science faculty at the University of Basel or faculty at other institutions (many outside of Switzerland) who collaborate with them can submit a research project idea to the SNI. A committee selects the projects to be supported. Each project requires both a faculty principle investigator (PI) and coPI. These faculty are frequently from different disciplines so most projects are inherently interdisciplinary. Based on the committee's recommendations, the research opportunities are advertised around the world to potential Ph.D. students. The ratio of applicants to positions is about 50 : 1.

47.5.2 Nanoscale Science and Technology Master's Program in Great Britain

Another example of a nanotechnology postgraduate degree is the Nanoscale Science and Technology master's program, taught for fifteen years at the University of Leeds in partnership with the University of Sheffield in the United Kingdom [47.44]. The joint program recently stopped admitting students due to a change in the university's focus. (The University of Leeds now offers a Nanomaterials and Materials Engineering graduate program.) This one-year Nanoscale Science and Technology master's program followed the typical three-year undergraduate degree. While most master's degrees narrow the student's academic focus on a particular STEM topic, this program widened the scope of study so that a student progressed from a bachelor's degree in any single science or engineering discipline to an interdisciplinary nanotechnology master's degree. Students learned about electronics, materials, biology, chemistry, and physics with the emphasis on the latter two fields. Individual courses were designed to be interdisciplinary, present advanced topics, and simultaneously provide the fundamental knowledge that some students might not possess due to their varied academic backgrounds. The faculty examined the bachelor's-level courses of applicants to the program in order to determine if they have the proper preparation. The curriculum was organized so that a single course covered a broad topic, such as characterization techniques, with both theoretical and experimental information. Top-down and bottom-up approaches to nanotechnology were covered in separate semesters.

Students' coursework was evaluated using a variety of methods, including presentations, exams, essays, and laboratory experiments. Experiments were designed to be relatively brief and not require advanced instrumentation or an inordinate amount of training before use. In addition to classes, students completed a research project. With multiple STEM departments from the two universities contributing, students chose one of many projects. Preparation for this began early in the program with the student selecting a topic (and several alternate choices) and planning the research in the first semester, then collecting and reporting preliminary results in the spring. The student completed the research during the summer and gave a presentation of the work. Prior to the fall semester, the student wrote a thesis. This curriculum could be modified to accommodate students wishing to pursue the master's degree on a part-time basis or a certificate was available if students did not wish to complete the entire program. The curriculum topics were also adaptable for short courses, which better suited the needs of industrial nanotechnology workers.

47.5.3 Laboratory Instrumentation Course in Taiwan

The interdisciplinarity of nanotechnology creates a challenge for faculty and staff who train students to operate research instruments [47.45]. Potential users start with a wide range of background knowledge and they have different needs for the instruments. The National Taiwan University addressed this problem by creating a graduate-level Applied Mechanics Laboratory course for students studying bionanotechnology. Course topics covered fluid mechanics, dynamics, mechanics of materials, and nanobiomechanics. There is some flexibility in the course topics so a student can customize the curriculum to match his or her interests. Course content includes theoretical background, operation of the instrument, and the interpretation of results for device fabrication. The nanoelectromechanical systems (NEMS), nanotechnology, and nanobiology teaching laboratories are all involved in the course. Faculty, research staff, technicians, and graduate students from each lab collectively play the role of facilitators for other graduate students enrolled in the class. In total, 50 faculty members from many departments and colleges were involved in developing the learning materials. A virtual portal for students to learn about the instruments is a major component. Online content includes recorded lectures, multimedia, digitized copies of printed manuals, demonstrations, and remote access to instruments. Since each student can choose different

course material to study, facilitators do not give organized lectures. Instead, they provide assistance to the students in the lab and expect the students to seek help when they need it. This self-paced learning environment is especially appropriate for graduate students who need to develop the intrinsic motivation to learn in order to succeed as independent researchers. Students are evaluated based on the quality of their lab work, written lab reports, and an oral exam.

Students responded favorably in a survey of their perceptions of the Applied Mechanics Laboratory course, including the course materials, virtual and face-to-face learning environments, and the quality of instruction. Students were less pleased with the way the instructors taught the course. This may be due to their expectation of a traditional lecture format, which was not used. Students took advantage of the online course content, especially for the experiments that they were most interested in learning. Student felt that the virtual portal improved their learning of both the principles of the instruments as well as the operations. They felt that the online information reduced the number of potential mistakes they might make. This is an important consideration when designing a hands-on laboratory instrumentation course.

47.5.4 Collaborative, Online Graduate Program in Israel and Europe

EduNano is a comprehensive, online academic program developed and supported by a consortium of universities in Europe and Israel [47.46]. Course material is available for professional development of K-12 teachers, undergraduate and graduate degree programs, and continuing education. The online structure of the program allows multiple collaborating institutions to teach geographically dispersed students. Six universities in Israel participate along with universities in Bulgaria, France, and Italy. An electronics company, Elbit Systems in Israel, is an active partner and provides an industry perspective. Faculty at each university prepare and contribute online courses covering all fields of nanotechnology. Students completing these courses at any of the universities receive equivalent credit towards their degree.

The curriculum was developed based on the availability of existing courses taught at the member universities, the perceived need for the courses by academic researchers and industry, and a list of agreed-upon learning outcomes. Learning outcomes state the expected tasks that students should accomplish upon completing the course and are an essential guide to developing curricula and assessment. Online class sessions consisted of video lectures, a combination of video lecture and slide show, and text-based information on a web page. Video lectures were created by recording regular, face-to-face class sessions, recording a lecture in a studio, or recording the lecturer sitting at a computer with the lecture video combined with animations, a slide show, or other imagery. Each of these approaches has advantages and drawbacks, depending on the subject matter being taught and the talents of the lecturer. Assessment of student performance may take place through traditional examinations if the students are all located near the school teaching the online course. If students are far from the school, they may complete another type of assignment, such as online quizzes and presentations given to the class using video conferencing software. The EduNano program recently completed its pilot test and began offering classes in 2016.

47.6 Teacher Professional Development

In order to further nanotechnology education, educator training is essential. Educators provide the technical content and encourage excitement and creativity. It is critical to attract and educate students in the growing field of nanotechnology in order to develop a trained nanotechnology workforce. Therefore, it is imperative to ensure appropriate and sufficient educator training.

Presently, education degree programs focus on traditional STEM fields without specific training in nanotechnology. Teachers typically learn one STEM subject in college and may teach only that subject. If they are not knowledgeable in another subject, they will hesitate to teach it. Therefore, postgraduate courses, summer classes, and professional development activities are necessary to help teachers learn about nanotechnology. Reviews of education research related to nanotechnology professional development and preservice teacher training provide more background on this subject [47.15, 47]. While education students (i. e., future teachers) could enroll in nanotechnology courses in college, a course specifically designed to address their needs as STEM teachers (rather than researchers) would be more useful.

There is a wealth of information available from a variety of sources on nanotechnology [47.13]. In many cases, the deficiency is in the packaging. Teachers seeking to stay up-to-date with nanotechnology do not have the time to find, organize, and learn new information on their own. Assembling the information into courses, short courses, or learning modules is time consuming,

but necessary in order to make the information accessible and to disseminate the information to those who need it quickly. Further, assembling the information into discrete packages allows for flexibility and targeted training.

47.6.1 Teacher Training in Israel

The Chemistry Department of the Weizmann Institute of Technology in Israel provides an example of such packaging. High school teachers seeking an advanced education degree can enroll in Introduction to Materials and Nanotechnology [47.48]. Lectures focus on the essential knowledge that students need in order to understand nanotechnology. These include a brief history of materials science, a qualitative description of quantum mechanics, and material characterization methods such as probe microscopy and x-ray measurements (instrumentation was selected based on its importance in nanotechnology and its availability at the university). These lectures are given by research faculty at the university and include visits to the research laboratory where the instruments are housed. Each student then selects a topic to study in greater depth. They share what they learned with the class in a lecture-style presentation and a report, including the history of the topic, its scientific description, technological applications, and contributions made by Israeli scientists. Nanotechnology research faculty provide opportunities for the students to work in their laboratories so that students receive an authentic research experience. While completing these assignments, students meet for weekly recitations where they can ask questions about the scientific information presented in class, their individual topic and presentation, and how to apply their new content knowledge to the classroom. Students complete pre- and postsemester quizzes that cover content knowledge taught in the course. In a study of this program's effectiveness, all participating students were current high school teachers seeking an M.Sc. degree for educators. Students initially scored low, but showed significant improvement on the postsemester quiz. In interviews, students related that they felt that they gained a lot from the course and they expressed greater confidence in their ability to teach nanotechnology to their students.

47.6.2 Teacher Training in South Korea

A short course offered in the Republic of Korea (South Korea) provides training for existing high school teachers to become more fluent in new fields of technology that appear in new editions of their students' textbooks [47.49]. Begun in 2006 at Chonbuk National University with 20 students enrolled, the program quickly grew to involve other university partners and now educates more than 500 students each year. The ten-day course devotes 40% of its time to nanotechnology, with biotechnology, environmental science and technology, and other activities each comprising 20% of the course. Teachers receive a lecture about a topic then perform a related laboratory experiment. Laboratory activities are especially important since college education in South Korea focuses on memorization of knowledge rather than gaining hands-on experience. Topics covered include structural properties of nanomaterials, lithography, imaging, optical properties of light-emitting diodes (LEDs), synthesis of a variety of nanostructures, and applications of these materials. The program has been well received by its students. A drawback to these courses is that students must be geographically close to the campus.

47.6.3 Online Education for Teachers

An alternative to a traditional face-to-face class is an online course, such as Nanoscience for Teachers conducted at the University of Wisconsin–Madison [47.50]. This course is exemplary for its well-designed structure, content, and emphasis on using education research findings to improve the learning of students (in this case, teachers). Because it is an online course, students from around the world could enroll in the class.

This graduate course helps middle and high school teachers learn the fundamental knowledge required for understanding nanotechnology and helps them to stay up to date on the current state of the discipline. The online course is taught asynchronously, so the students have greater flexibility in their schedules. Based on previous research of student learning in online environments [47.51], course designers created a structured, easy-to-follow format for the course in order to maximize student learning. Instructors provide many opportunities and means for communication among students and between the instructor and students, which leads to greater student satisfaction and is consistent with social constructivism pedagogy.

During the eight-week course, teachers learn about nanotechnology through readings and online discussions activities. They complete weekly quizzes to assess their learning. Each week, a guest lecturer provides more in-depth information about a nanotechnology topic via an online chat. Weekly topics include all of the Big Ideas discussed in Sect. 47.1.2, with a significant amount of time devoted to the occurrence of nanostructures in nature, societal impacts, and applications. Each student in the course develops a module that they could present to their own students. Some teachers create a se-

ries of short lessons that would be taught throughout the year, while others design a single, in-depth lesson about nanotechnology. The goal of this final project is similar to that of the assignment given to students in the face-to-face graduate course at the Weizmann Institute of Technology discussed previously but the format differs due to the online environment. Each student creates a complete lesson plan (documents or presentation slides) and uploads them to the course website. Students engage in anonymous peer review of their classmates' work.

Since most of the students in the class lived relatively close to the university, the instructors organized a site visit to the University of Wisconsin–Madison campus during the middle of the course. The visit consisted of a lab facility tour, instrument demonstrations, and laboratory activities, which teachers could perform with their students. Materials were posted online after the tour for the benefit of those who could not attend the site visit.

Student satisfaction was assessed with weekly surveys that allowed the instructors to make small modifications as the course progressed. The Constructivist On-Line Learning Environment Survey tool (COLLES) evaluates six aspects of a learning community: professional relevance, interactivity, peer support, instructor support, reflective thinking, and understanding between the instructor and students [47.52]. Results of COLLES collected prior to beginning the course (measuring students' expectations) and after completing it (measuring their view of the course experience) showed that the course met teachers' expectations held prior to beginning it with the exception of the degree of interactivity. Teachers showed statistically significant gains in content knowledge, based on pre- and postcourse quizzes. A follow-up survey conducted a year later revealed that most of the teachers actually did introduce their nanotechnology modules created in the online class to their students.

47.7 Informal Education

Education for the general public on the benefits of nanotechnology is an important part of improving nanotechnology education. This requires supporting and engaging in community-based opportunities for science and education and making them approachable, affordable, inclusive, and interesting. Currently, this is being done mostly by a mix of science museums and university groups, separately or in collaboration [47.53, 54]. In the US, these efforts are largely funded by the NSF.

A major effort has been carried out by the Nanoscale Informal Science Education Network (NISE Net) to organize and implement broadly informal nanotechnology education to the public at large [47.55]. Public understanding and attitudes affect the environment for research and development (R&D), reputation, and market acceptance of nanotechnology products [47.56], so informal education is an important area and requires funding. There is a need to expand informal science education to ensure that all corners of society are being engaged in these topics, especially from a young, formative age. Below are two examples of informal education programs that illustrate the range of ways to inform the public about nanotechnology – one is a single activity developed for a science festival and the other is a major museum exhibition.

47.7.1 Science Festival Activity in Israel

A one-hour activity, *It's a Small World After All*, was designed for an Israeli science festival [47.57]. Four hundred children participated, and assessment data was collected for 70 children, aged 9, 10, or 11. To evaluate any changes in knowledge and opinions of nanotechnology, each child engaged in personal meaning mapping (PMM), a process in which a person writes down (in text or pictures) all thoughts or known facts about a particular word, in this case *nanotechnology*, before the activity, and then again afterwards [47.58]. This is a quick process that shows how the learner's views and knowledge of the subject changes and the extent of the change. The activity introduced children to two topics: size and scale, and applications of nanotechnology, appropriate for children. Using pre- and postactivity PMM data, instructors measured changes in children's understanding of size and scale, applications, and the definition of nanotechnology as well as changes in their attitudes about nanotechnology. Children were not given any specific instructions about what to write on their maps.

Following best practices developed through education research, the designers used many teaching methods, such as experiments, videos, hands-on activities, demonstrations, and constructing models. Instructor-led presentations focused on analogies to help the young audience understand the relative sizes of nanoscale and everyday objects. To illustrate the vast difference between a value of 1 and 1×10^{-9}, children performed a ten-fold dilution of food coloring nine times. They wetted a cloth that contained nanoparticles that imparted a superhydrophobic property to its surface and

observed that the fabric remains dry. Children held and smelled used socks that remained odor-free due to copper nanoparticles embedded in the threads. They also received materials for building a cardboard fullerene structure.

Children and their accompanying families responded positively to the activity, based on their questions asked after the event was over. Children showed improvement in all four dimensions of the PMM assessment (definition, size and scale, applications, and attitudes). For all form categories, more boys showed an initial understanding or positive attitudes about nanotechnology than girls. After the activity, the difference between the number of boys and girls responding correctly or with positive attitudes had disappeared and in three of the four dimensions, more girls provided accurate responses or showed positive perceptions of nanotechnology. Based on a percentage of possible gain, children showed a 16% increase in providing a correct definition of nanotechnology (with all of the gain due to girls' responses), a 77% increase for size and scale understanding, a 59% increase in correct information about applications, and an increase of 40% in expressing positive attitudes. Since size and scale is a big idea of nanotechnology that all other big ideas build upon, it is noteworthy that the activity led to a majority of children understanding this concept [47.23]. No children expressed negative attitudes about nanotechnology in the postactivity PMM. It is clear that the use of multiple teaching methods as well as interesting and relevant examples and applications of nanotechnology in this one-hour event led to a significant, positive effect on the children's knowledge and attitudes about nanotechnology.

47.7.2 NanoAventura in Brazil

Museums around the world have developed various exhibits portraying nanotechnology phenomena that can be experienced by visitors of all ages. Exhibits can be located in a single museum or they can tour many locations. An example of the latter type is NanoAventura (NanoAdventure), a traveling exhibit developed by the Museu Exploratório de Ciências in Brazil in 2005. The hour-long exhibit was open to the general public and designed for a target audience of children aged 9–14. Guides led the audience as they viewed videos and played nanotechnology-themed games and simulations together and in smaller groups [47.59–61].

Size and scale, the composition of matter, and applications were viewed as important topics for introducing nanotechnology to the audience. Nanotechnology presents additional challenges to designing a museum exhibit due to the new terminology (e.g., *nano*) and lack of familiarity with nanoscale objects. To address this, an introductory video compared sizes of different objects with an emphasis on biological materials. The audience, originally consisting of up to 48 visitors, was divided into 12 member groups to play four video games.

Video games were chosen as a medium for learning because children and teenagers are often regular video game players. Games required all team members to work together but the game recorded both individual and group scores, providing cooperative and competitive aspects [47.59]. Topics for the four games are nanomedicine, removing impurities from a surface using AFM, a nanomanufacturing assembly line, and a virtual tour of UNICAMP research facilities.

Previous studies of the public's attitudes towards nanotechnology indicate that a positive view is associated with the expectation that nanotechnology will improve their quality of life [47.62]. Medicine is a field that is associated with quality of life and one that nanotechnology can improve. Therefore, the NanoAventura designers felt that medicine could be a concept that would keep the audience's attention. Also, biological cells were often cited by participants as the smallest object they know about so this was viewed as a useful reference point for presenting nanoscale objects.

Not surprisingly, the general public was found to lack a basic understanding of these topics. An evaluation of NanoAventura showed that visitors more frequently provided correct answers to questions about the particulate nature of matter and applications of nanomaterials after completing the exhibit [47.61, 63]. They also changed their description of a material's composition from stating its origin (e.g., sugar is made of sugar cane) to an answer that included atoms and molecules. Visitors more often favorably associated nanotechnology with scientific advances, progress, and the future. Overall, they found the exhibit interesting and educational.

47.8 Summary and Outlook

Nanotechnology is an area of rapid growth in research, development, and production. Nanotechnology R&D and applications affect industry, academia, and consumers as well. Unfortunately, the demand for scientists, engineers, and skilled workers worldwide in nanotechnology has outpaced the supply. Additionally, research has shown a fundamental lack of nanoliteracy in the general public, further reducing the likelihood of students choosing to study nanotechnology. In order to meet the demands of the burgeoning fields, an increase in the numbers of researchers and technicians available is required, and to that end, formal and informal education efforts have been implemented in various locations around the world. These efforts have targeted all levels of education, from primary and secondary through technical and university-level education, and including continuing education for educators and informal education for the general population. The goals of introducing students to nanotechnology are to fill the jobs that require an educated and skilled workforce and to promote science literacy in the general population who participate in policy decisions and engage with nanotechnology applications as consumers. This chapter has identified creative ways of increasing excitement, creativity, content retention, and improving attitudes towards nanotechnology in potential students at all levels. Various strategies have been used to make education inclusive, accessible, and culturally relevant to students of all ages.

In the primary and secondary school arena, Taiwan, Italy, and Germany have implemented programs to increase nanotechnology understanding. In Taiwan, a field trip to Ilan University was created to engage students in hands-on lab activities promoting better awareness and understanding of nanotechnology. In Italy, educators aimed to combine content learning and soft skills to provide students a broad nano-experience. Additionally, secondary students learned by teaching, in that they researched and prepared lessons to teach to primary students. This promoted the content understanding, as well as communication, teamwork, and problem-solving skills. In Germany, the government implemented a traveling exhibit bringing hands-on nano-experience of materials and applications to students. The exhibit encouraged basic science and nanotechnology mastery with educator support, as well as industry-supported curricula with online support for additional learning opportunities. The EU examined such programs resulting in NanoEIS reports on the strengths and opportunities of them. The reports identified that compulsory activities with a lack of hands-on interaction were common, and that there was low community support and variable industry support. Opportunities exist for university and industry involvement to promote transition to university study, an objective sorely needed as current programs show this transition is difficult in the nanotechnology arena. Further, exposure to nanotechnologies and support for educators could increase successes of programs and further funnel students to technical and university STEM programs.

VET, Penn State, DCTC in Minnesota, and NAIT in Canada have created unique programs and opportunities. Penn State has created a capstone program for vocational and bachelor's degree students that inserts nanotechnology into foundational disciplines. Further, Penn State developed the NACK network to advance vocational education in nanotechnology. The DCTC relied heavily on industry needs to create a technician training program that combines hard content skills with soft skills of problem solving and critical thinking. The program serves as a model to others, encouraging local industry inquiry to tailor the program to local needs. NAIT's Nanotechnology Systems degree focuses on not just traditional content subjects, but also applied knowledge courses such as project management and manufacturing. These examples provide a foundation, but improvement in VET is needed in the development of best practices manuals and development of national curricula with sufficient rigidity to promote learning and flexibility to meet local needs. As well, diversity of the student population should be encouraged.

Undergraduate education initiatives have taken a variety of forms around the world. Australia, Switzerland, the US, and South Africa have all tackled the problem of nanotechnology education at this level. In Australia, Bachelor of Science degrees in nanotechnology were created at Flinders University and the UNSW, among the first of their kind in the world, by drawing together faculty and curricula in several existing departments. The University of Technology, Sydney also created a degree program, but did so by creating an entirely new department. These degree programs create specialized nanotechnology education, deepening students' understanding. The University of Basel's SNI improved the interdisciplinary approach with a B.S. in Nanoscience. Traditional content is combined with soft skills. Short courses emphasize specialized and cross-disciplinary content, with a research emphasis. Strong support for transition to graduate education was indicated. The US NSF has created a national summer REU providing training, research education, presentation skills, and professional development connections in nanotechnology for undergraduates. If students excel, they may pursue an international version of the program

in the subsequent summer. The international program is aimed to position US students to solve global problems on a global stage by developing culturally aware STEM students. Assessment showed significant improvements in knowledge, strategy, and action in participants. In South Africa, rather than a separate degree, the UKZN elected to develop a course in nanotechnology for chemistry students as a way to introduce the topic in the developing world. Not surprisingly, socioeconomic disparity and cultural disparities played a role in the outcomes of the program. Despite these successful programs, opportunities at the undergraduate level exist in the need to promote interdisciplinary cooperation, societal impacts, and diversity in the student population.

At the graduate level, UCSB, the University of Basel, universities in the United Kingdom, the National Taiwan University, and a consortium of European and Israeli universities have all developed interesting approaches. UCSB has embedded graduate students in nanotechnology Ph.D. programs into social science and humanities projects to reveal connections between the laboratory and broader societal forces. Further, the program promotes strong cross-disciplinary cooperation. The University of Basel's SNI graduate program covers both master's and Ph.D. students. At the master's level, students must complete a thesis in their chosen discipline, as well as complete two projects that are strongly cross-disciplinary in nature. The Ph.D. program is project driven, in that faculty PIs and coPIs propose projects, gain approval from the SNI, and then advertise worldwide for project positions (Ph.D. student openings). In this way, the projects drive the student selection. At the University of Sheffield and University of Leeds, a master's in Nanoscale Science and Technology broadens, rather than concentrates, an undergraduate degree in a single science to an interdisciplinary master's. Courses focus on cross-disciplinary topics, rather than scientific categories, e.g., characterization techniques, rather than chemistry. At the National Taiwan University, the need for students to learn instrumentation techniques coming from different science backgrounds is being addressed. The course created for biotechnology students flips the classroom: reading and study materials are examined outside the classroom, and classroom time is focused on implementation of learning, with instructor support. The European-Israeli consortium is developing an online program for continuing education incorporating university educators and researchers with industry-partnering support. Online classes include video lectures and/or text-based information (slides or web content), with first classes beginning in 2016. Missing from all these programs is an opportunity to address diversity and inclusion. While some programs target worldwide audiences, little effort can be seen in the programs identified to target all corners of society and all cultural and socioeconomic backgrounds.

In the area of professional development, the Weizmann Institute of Technology in Israel, Chonbuk National University in South Korea, and The University of Wisconsin–Madison in the US have all identified strategies for teacher professional development. At the Weizmann Institute, teacher training in their university education includes a focus on learning by teaching with essentials in nanotechnology knowledge, and then projects to increase broad understanding. The program features a laboratory research experience to provide hands-on education. At Chonbuk National University, a short course for existing high school teachers with lectures on a topic and then a laboratory experience implementing the topics has been developed. The hands-on nature of the laboratory experience is essential in South Korea where an overall lack of hands-on focus exists in education. The University of Wisconsin–Madison developed an online course *Nanoscience for Teachers* with readings, online discussions, weekly quizzes, and guest lecturers for each topic. Teachers developed a module for their own classrooms, tailored to their teaching style and the needs of their students, then anonymously peer reviewed their classmates' modules for improvement opportunities. Strong assessment showed significant gains in content knowledge. While there is a wealth of information and content for professional development currently existing, the problem is in the packaging. Materials need to be packaged for flexibility and targeted training. Further, the university education for teachers needs to be designed to meet their needs as STEM educators, rather than researchers.

In the area of informal education, Israel and Brazil provide examples of bringing nanotechnology to the general public. In Israel, a one-hour activity was developed for a science festival for children aged 9–11, and included presentations and hands-on activities. The activity focused on concepts of size, scale, and applications. Age-appropriate assessment showed significant gains in understanding of size and scale, applications, and attitudes towards nanotechnology. In Brazil, a traveling museum exhibit featuring a one-hour interactive activity was aimed at children ages 9–14, and focused on using existing public competencies in size and application to leverage new understandings and knowledge. Participants watched videos and then played nanotechnology-themed video games. Cooperative and competitive learning was highlighted, and assessment showed gains in understanding of size and applications. In general, it seems that most informal education is aimed at younger audiences, who may already be exposed to these subjects and ideas in other arenas, such

as formal schooling. A significant opportunity exists to educate an older population that may not have formal education opportunities or interest in nanotechnology. To promote the goal of nanotechnology literacy, development of programs aimed at adult populations should be encouraged.

References

47.1 K. Winkelmann, B. Bhushan (Eds.): *Global Perspectives of Nanoscience and Engineering Education* (Springer, Cham 2016)

47.2 G. Yakman, H. Lee: Exploring the exemplary STEAM education in the U.S. as a practical educational framework for Korea, J. Korea Assoc. Sci. Educ. **32**, 1072–1086 (2012)

47.3 J.-E. Jon, H.-I. Chung: *Consultant Report Securing Australia's Future Stem: Country Comparisons, STEM Report – Republic of Korea* (Australian Council of Learned Academies, Melbourne 2013), available online

47.4 D.J.C. Herr: The need for convergence and emergence in twenty-first century nano-STEAM+ educational ecosystems. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 83–115

47.5 G. Ladson-Billings: But that's just good teaching! The case for culturally relevant pedagogy, Theory Pract **34**, 159–165 (1995)

47.6 R. Eglash: Nanotechnology and traditional knowledge systems. In: *Nanotechnology and Global Sustainability*, ed. by D. Maclurcan, N. Radywyl (CRC Press, Boca Raton 2012) pp. 45–66

47.7 M. Reibold, P. Paufler, A.A. Levin, W. Kochmann, N. Pätzke, D.C. Meyer: Discovery of nanotubes in ancient Damascus steel. In: *Physics and Engineering of New Materials*, ed. by D.T. Cat, A. Pucci, K.R. Wandelt (Springer, Berlin 2009) pp. 305–310

47.8 P.P. Edwards, J.M. Thomas: Gold in a metallic divided state – from Faraday to present-day nanoscience, Angew. Chem. Int. Ed. **46**, 5480–5483 (2007)

47.9 M.C. Roco: Broader societal issues of nanotechnology, J. Nanopart. Res. **5**, 181–189 (2003)

47.10 European Commission: *Towards a European Strategy for Nanotechnology* (Office for Official Publications of the European Communities, Luxemburg 2004), available online

47.11 K. Winkelmann: A ten year review of the NSF Nanotechnology in Undergraduate Education (NUE) program, J. Nano Educ. **6**, 109–116 (2014)

47.12 T. Amjad, J. Nielsen, L. Osborne, S. Eng, T. Jarosz: *Analysis of Reports of the Nanotechnology Undergraduate Education in Engineering Program* (Manhattan Strategy Group, Bethesda 2012), available online

47.13 K. Winkelmann, L. Bernas, M. Saleh: A review of nanotechnology learning resources for K-12, college and informal educators, J. Nano Educ. **6**, 1–11 (2014)

47.14 M. Orgill, S.A. Wood: Chemistry Contributions to nanoscience and nanotechnology education: A review of the literature, J. Nano Educ. **6**, 83–108 (2014)

47.15 L.A. Bryan, A.J. Magana, D. Sederberg: Published research on pre-college students' and teachers' nanoscale science, engineering, and technology learning, Nanotechnol. Rev. **4**, 7–32 (2015)

47.16 J. Light Feather, M.F. Aznar: *Nanoscience Education, Workforce Training, and K-12 Resources* (CRC Press, Boca Raton 2011)

47.17 A.E. Sweeney, S. Seal (Eds.): *Nanoscale Science and Engineering Education* (American Scientific Publishers, Stevenson Ranch 2008)

47.18 K.A.O. Pacheco, R.W. Schwenz, W. Jones Jr. (Eds.): *Nanotechnology in Undergraduate Education* (American Chemical Society, Washington 2009)

47.19 P. Berger, N.B. Adelman, K.J. Beckman, D.J. Campbell, A.B. Ellis, G.C. Lisensky: Preparation and properties of an aqueous ferrofluid, J. Chem. Educ. **76**, 943–948 (1999)

47.20 C.S. Rapp: Getting close with the instructional scanning tunneling microscope, J. Chem. Educ. **74**, 1087–1089 (1997)

47.21 B. Hingant, V. Albe: Nanosciences and nanotechnologies learning and teaching in secondary education: A review of literature, Stud. Sci. Educ. **46**, 121–152 (2010)

47.22 R. Blonder, S. Sakhnini: What are the basic concepts of nanoscale science and technology (NST) that should be included in NST educational programs? In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 117–127

47.23 S. Stevens, L.M. Sutherland, J.S. Krajcik: *The Big Ideas of Nanoscale Science and Engineering: A Guidebook for Secondary Teachers* (NSTA, Arlington 2009)

47.24 C.Y. Huang, L.R. Hsu, H.C. Chen: A study on the core concepts of nanotechnology for the elementary school, J. Natl. Taichung Univ. Math. Sci. Technol. **25**, 1–22 (2011)

47.25 S. Sakhnini, R. Blonder: Essential concepts of nanoscale science and technology for high school students based on a Delphi study by the expert community, Int. J. Sci. Educ. **37**, 1699–1738 (2015)

47.26 M.-T. Wu, S.-Y. Lin, Y.-I. Cho, H.-H. Chen: The service-learning courses in Ilan University to promote nanotech's popular science education, Int. J. Info. Educ. Technol. **5**, 814–817 (2015)

47.27 P. Ambrogi, M. Caselli, M. Montalti, M. Venturi: Make sense of nanochemistry and nanotechnology, Chem. Educ. Res. Pract. **9**, 5–10 (2008)

47.28 R. Nonninger, J. Dege, T. Wilke, T. Waitz: Nanoscience education in school chemistry: Perspectives for curricular innovations in context of an education for a sustainable development. In: *Global Per-*

spectives of Nanoscience and Engineering Education, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 237–274
47.29 M. Talesnik, D. Rosenberg, K. Griffin, B. Szafran, A. Duschl: Report on best practices in nanotechnology education at the secondary school level (NanoEIS 2014)
47.30 M. Talesnik, D. Rosenberg, C. Aberg, I. Lynch: Secondary school education and its contribution to facilitating transition into university (NanoEIS 2013)
47.31 I. Malsch: Nano-education from a European perspective: Nano-training for non-R&D jobs, Nanotechnol. Rev. **3**, 211–221 (2014)
47.32 R. Yawson: Skill needs and human resource development in the emerging field of nanotechnology, J. Vocat. Educ. Train. **62**, 285–296 (2010)
47.33 G. Cibuzar: A university-technical college nanoscience training program. In: *16th Bienn. Univ./Gov./Ind. Microelectron. Symp., San Jose* (2006) pp. 101–104
47.34 Northern Alberta Institute of Technology: *Nanotechnology Systems*, available online at http://www.nait.ca/program_home_77412.htm
47.35 J. Murday: Nanoscale Science and Engineering Education (NSEE) – The Next Steps, J. Nano Educ. **7**, 85–147 (2015)
47.36 J.G. Shapter, J.J. Gooding: Teaching undergraduates nanotechnology. In: *Nanoscale Science and Engineering Education*, ed. by A.E. Sweeney, S. Seal (American Scientific, Stevenson Ranch 2008) pp. 421–457
47.37 C. Moller, K. Spieler, M. Calame, E. Meyer: From bachelor to PhD: The Swiss nanoscience institute at the university of Basel offers excellent interdisciplinary education at all levels. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 351–373
47.38 National Science Foundation: *NSF's Research Experiences for Undergraduates (REU) program: An Assessment of the First Three Years (Report 90–58)* (National Science Foundation, Washington DC 1990) pp. 1–29
47.39 D.A. Willis, P.S. Krueger, A. Kendrick: The influence of a research experiences for undergraduates program on student perceptions and desire to attend graduate school, J. STEM Educ. **14**, 21–28 (2013)
47.40 E. Horsch, M.S. John, R.L. Christensen: A case of reform: The undergraduate research collaboratives, J. College Sci. Teach. **41**, 38–43 (2012)
47.41 T.D. Sadler, L. McKinney: Scientific research for undergraduate students: A review of the literature, J. College Sci. Teach. **9**, 43–49 (2010)
47.42 N. Healy, L. Ratbun: Education and outreach of the national nanotechnology infrastructure network (NNIN) 2004–2015: History and accomplishments of undergraduate programs. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 323–349
47.43 B.H. Augustine, O.Q. Munro: In to Africa: Teaching nanoscience to undergraduates in KwaZulu-Natal, South Africa, Mater. Res. Soc. Symp. Proc. **1320**, 48–58 (2011)
47.44 R.W. Kelsall: Master's level nanotechnology education. In: *Nanoscale Science and Engineering Education*, ed. by A.E. Sweeney, S. Seal (American Scientific, Stevenson Ranch 2008) pp. 649–661
47.45 H.-P. Yueh, H.-J. Sheen: Developing experiential learning with a cohort-blended laboratory training in nano-bio engineering education, Int. J. Engng. Ed. **25**, 712–722 (2009)
47.46 S.R. Cohen, R. Blonder, S. Rap, J. Barokas: Online nanoeducation resources. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 171–194
47.47 R. Blonder, I. Parchmann, S. Akaygun, V. Albe: Nanoeducation: Zooming into teacher professional development programmes in nanoscience and technology. In: *Topics and Trends in Current Science Education*, ed. by C. Bruguière, A. Tiberghien, P. Clément (Springer, Dordrecht 2014) pp. 159–174
47.48 R. Blonder: The Story of nanomaterials in modern technology: An advanced course for chemistry teachers, J. Chem. Educ. **88**, 49–52 (2011)
47.49 S.-W. Han, J.-S. Jeon, B.J. Ahn: Training programs of advanced sciences for high school physics teachers, J. Mater. Educ. **31**, 45–50 (2009)
47.50 J.H. Tomasik, S. Jin, R.J. Hamers, J.W. Moore: Design and initial evaluation of an online nanoscience course for teachers, J. Nano Educ. **1**, 48–67 (2009)
47.51 K. Swan: Building learning communities in online courses: The importance of interaction, Educ. Commun. Inf. **2**, 23–49 (2002)
47.52 P.C. Taylor, D. Maor: Assessing the efficacy of online teaching with the constructivist on-line learning environment survey. In: *Flexible Futures in Tertiary Teaching/Proceedings of the 9th Annual Teaching Learning Forum*, ed. by A. Herman, M. Kulski (Curtin University of Technology, Perth 2000)
47.53 A.C. Payne, W.A. DeProphetis, A.B. Ellis, T.G. Derenne, G.M. Zenner, W.C. Crone: Communicating science to the public through a university-museum partnership, J. Chem. Educ. **82**, 743–750 (2005)
47.54 W.C. Crone: Bringing Nano to the public: A collaboration opportunity for researchers and museums, J. Nano Educ. **2**, 102–116 (2010)
47.55 L. Bell: Nanoscale informal science education in the U.S. – NISE Net. In: *Global Perspectives of Nanoscience and Engineering Education*, ed. by K. Winkelmann, B. Bhushan (Springer, Cham 2016) pp. 277–311
47.56 J.F. Sargent Jr.: Nanotechnology: A policy primer, Tech. Rep. RL34511 (Congressional Research Service, Washington DC 2013)
47.57 R. Blonder, S. Rap: It's a small world after all: A nanotechnology activity in a science festival, J. Nano Educ. **4**, 47–56 (2013)
47.58 M. Adams, J.H. Falk, L.D. Dierking: Museums, learning, and research. In: *Researching Visual Arts Education in Museums and Galleries: An International Reader*, ed. by M. Xanthoudaki, L. Tickle, V. Sekules

(Springer, Dordrecht 2012) pp. 14–32
47.59 S. Murriello, D. Contier, M. Knobel: Challenges of an exhibit on nanoscience and nanotechnology, J. Sci. Commun. **5**, 1–11 (2006)
47.60 S.E. Murriello, M. Knobel: Encountering nanotechnology in an interactive exhibition, J. Museum Educ. **33**, 211–230 (2008)
47.61 S. Murriello, D. Contier, M. Knobel: NanoAventura: An Interactive exhibition on nanoscience and nanotechnology as an educational tool, J. Nano Educ. **1**, 1–10 (2008)
47.62 E. Einsiedel: In the public eye: The early landscape of nanotechnology among Canadian and U.S. publics, AZojono J. Nanotechnol. Online (2005) doi:10.2240/azojono0110
47.63 M. Knobel, S.E. Murriello, A. Bengtsson, A. Cascón, R. Zysler: The perception of nanoscience and nanotechnology by children and teenagers, J. Mater. Educ. **32**, 29–38 (2010)

About the Authors



Aditya Aryasomayajula

McMaster University
Dept. of Mechanical Engineering
Hamilton, Canada
adityaarya83@gmail.com,
aryasoma@mcmaster.ca

Chapter C.16

Dr Aditya Aryasomayajula received his PhD in Electrical Engineering from the Technical University Dresden. He worked as a Postdoctoral Fellow at the European Molecular Biology Laboratory on microfluidics and genomics, before joining McMaster University as a Postdoctoral Fellow researching microfluidics, electrochemical sensors, and bio-sensing technology.

W. Robert Ashurst

Auburn University
Dept. of Chemical Engineering
Auburn, USA
ashurwr@auburn.edu



Chapter I.39

Dr W. Robert Ashurst is currently the Uthlaut Family Associate Professor of Chemical Engineering at Auburn University. He received his PhD in Chemical Engineering from the University of California at Berkeley. His research focuses on design of molecular precursors for advanced monolayer films, tribology at the micro and nanoscale, and the influence of surface chemical treatments on micro and nanoscale devices.

Massood Z. Atashbar

Western Michigan University
Dept. of Electrical & Computer Engineering
Kalamazoo, USA
massood.atashbar@wmich.edu



Chapter A.3

Professor Massood Z. Atashbar received degrees in Electrical Engineering from Isfahan University of Technology and Sharif University of Technology and in 1998 a PhD from RMIT University, Melbourne. After a postdoctoral at Pennsylvania State University, University Park, he joined the Electrical and Computer Engineering Department at Western Michigan University, where he currently is a Professor. His research interests include physical and chemical microsensors development, advanced signal processing, and engineering education.



Wolfgang Bacsa

Centre National de la Recherche Scientifique, Université de Toulouse
Centre d'Elaboration de Matériaux et d'Etudes Structurales (CEMES), UPR #8011
Toulouse, France
wolfgang.bacsa@cemes.fr

Chapter B.8

Professor Wolfgang S. Bacsa received his PhD from the Swiss Federal Institute of Technology (Zurich, 1990). He worked at Pennsylvania State University and Swiss Federal Institute of Technology in Lausanne and was a Visiting Professor at Boston University and at UNIST Korea. He is now a Professor at CEMES CNRS and the University of Toulouse. His research is dedicated to optical spectroscopy on nanostructured materials and novel optical microscopy techniques.



William S. Bainbridge

National Science Foundation
Division of Information & Intelligent Systems
Arlington, USA
wbainbri@nsf.gov

Chapter J.46

William Sims Bainbridge is the Director of the Cyber-Human Systems program of the National Science Foundation. He earned his PhD in Sociology from Harvard (1976). He has participated centrally in the Nanotechnology and Convergence activities and has co-edited and edited several books.

Antoni Baldi

Barcelona Microelectronics Institute (IMB-CNM-CSIC)
Chemical Transducers Group
Bellaterra, Spain
antoni.baldi@imb-cnm.csic.es



Chapter A.3

Dr Antonio Baldi received a degree in Telecommunication Engineering from BarcelonaTech (UPC) and a PhD in Electronics Engineering from Universitat Autònoma de Barcelona (UAB), followed by a postdoctoral at the University of Minnesota, working in the field of bioMEMS. In 2003, he joined the Chemical Transducers Group at the Barcelona Microelectronics Institute (IMB-CNM, CSIC). His current research is focused on electrochemical transducers and lab-on-a-chip devices for chemical sensing and biodetection.

Wilhelm Barthlott

University of Bonn
Nees-Institute for Biodiversity of Plants
Bonn, Germany
barthlott@uni-bonn.de



Chapter H.36

Dr Wilhelm Barthlott is a Professor Emeritus and former Head of the Biodiversity Institute at the University of Bonn. He is interested in structure, function, and biomimetic application of biological surfaces and has published more than 400 papers. His discovery of the Lotus Effect in 1997 has led to innovation in surface technologies. His current interests concentrate on air retaining surfaces and the relation between biodiversity and bionics.



Pouriya Bayat

York University
Dept. of Mechanical Engineering
Toronto, Canada
pouriya@yorku.ca

Chapter C.16

Pouriya Bayat obtained his BSc in Mechanical Engineering from Isfahan University of Technology. During his studies, he focused on bio-fluid mechanics. He joined the ACUTE Lab at York University in 2015, where he works on solution exchange of microparticles.



Mehmet Z. Baykara

Bilkent University
Dept. of Mechanical Engineering & UNAM
Ankara, Turkey
mehmet.baykara@bilkent.edu.tr

Chapter E.24

Mehmet Z. Baykara received his degree in Mechanical Engineering from Boğazici University and his PhD in Mechanical Engineering and Materials Science from Yale University. He is currently an Assistant Professor at the Department of Mechanical Engineering and the National Nanotechnology Research Center at Bilkent University, where he leads the scanning probe microscopy research group.

Bharat Bhushan Chapters 1, D.22, E.23, F.27, F.32–G.35, H.36–I.38, J.47 For biographical profile, please see the section "About the Editor".

Marcie R. Black

Advanced Silicon Group
Lincoln, USA
mrkarty@yahoo.com



Chapter B.9

Dr Marcie Black is the CEO of Advanced Silicon Group (ASG). ASG commercializes silicon nanowire arrays for applications, including black silicon solar cells, biosensors for less invasive testing, and lithium ion batteries. She has had over 25 patents issued and has numerous patents pending on using nanotechnology. Marcie earned her BS, MEng, and PhD degrees from the Massachusetts Institute of Technology.

Maarten P. de Boer

Carnegie Mellon University
Dept. of Mechanical Engineering
Pittsburgh, USA
mpdebo@andrew.cmu.edu



Chapter I.39

Dr Maarten P. de Boer received his PhD in Chemical Engineering and Materials Science from the University of Minnesota in 1996. His work at Carnegie Mellon University, and prior to that at Sandia National Labs, addresses challenges in tribology and thin film mechanical properties. Current research topics include reliability of nanomechanical switches, irradiation and mechanical properties of nanocrystalline metals, strength of semi-crystalline polymers, and adjustable coefficient of thermal expansion metamaterials.



Donald W. Brenner

North Carolina State University
Dept. of Materials Science & Engineering
Raleigh, USA
brenner@ncsu.edu

Chapter F.30

Professor Donald Brenner is a Kobe Steel Distinguished Professor of Materials Science and Engineering at North Carolina State University. He received his degrees in Chemistry from the State University of New York and Penn State University. His research interests focus on using atomic and multiscale modelling to develop and understand technologically important processes and materials.



Jean-Marc Broto

Centre National de la Recherche Scientifique, Institut National des Sciences Appliquées de Toulouse, Université de Toulouse
Laboratoire National des Champs Magnétiques Intenses de Toulouse (LNCMI-T), UPR #3228
Toulouse, France
jean-marc.broto@univ-tlse3.fr

Chapter B.8

Jean-Marc Broto is a Professor at the University Paul Sabatier Toulouse III. He is a specialist in electronic transport and magnetization properties under high magnetic fields and contributed to the discovery of giant magnetoresistance in 1988.

Patrizio Candeloro

University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Catanzaro, Italy
patrizio.candeloro@unicz.it



Chapter D.21

Dr Patrizio Candeloro is an Assistant Professor in Physics at the University Magna Graecia of Catanzaro and a Senior Researcher at BioNEM laboratory. Throughout his scientific career, he worked in the field of micro and nanotechnologies for different purposes and applications. Currently, he develops plasmonic nanostructures for SERS spectroscopy and investigates the applications of Raman techniques to biomedical issues.

Brigitte Caussat

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire de Génie Chimique (LGC), UMR #5503
Toulouse, France
brigitte.caussat@ensiacet.fr



Chapter B.8

Brigitte Caussat is a Professor at INP Toulouse, ENSIACET. Since 1991 she has been working on thermal chemical vapor deposition (CVD) processes, either on planar substrates or on powders. Her recent research activities have dealt with CNT oxidation in dry mode using gaseous mixtures containing ozone and CNT decoration by silicon, iron, nickel, and copper using the fluidized bed CVD process.



Chunying Chen

National Center for Nanoscience & Technology of China
CAS Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety
Beijing, China
chenchy@nanoctr.cn

Chapter J.45

Dr Chunying Chen received her PhD from Huazhong University of Science and Technology. She has worked at the Institute of High Energy Physics of the Chinese Academy of Sciences, the Medical Nobel Institute for Biochemistry at Karolinska Institute of Sweden, and at the National Center for Nanoscience and Technology of China. She was one of the earliest researchers in the field of nanosafety in China.



Rui Chen

National Center for Nanoscience & Technology of China
CAS Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety
Beijing, China
chenr@nanoctr.cn

Chapter J.45

Dr Rui Chen received his PhD from the Chinese Academy of Science in 2005. He has worked at the School of Life Sciences of Tsinghua University and at the National Center for Nanoscience and Technology of China. His research on nanoscience and nanotoxicology focuses on the risk assessment of nanomaterials to the environment and human health.

Yu-Ting Cheng

National Chiao Tung University
Dept. of Electronics Engineering
Hsinchu, Taiwan
ytcheng@mail.nctu.edu.tw



Chapter I.43

Yu-Ting Cheng received degrees in Materials Science and Engineering from National Tsing Hua University and Carnegie Mellon University and a PhD in Electrical Engineering from the University of Michigan Ann Arbor. In 2002, he joined the Department of Electronics Engineering at National Chiao Tung University. His research includes the fundamental study of materials for microsystem integration and MEMS applications, SoP, and the design and fabrication of microsensors and microactuators.

Mu Chiao

University of British Columbia
Dept. of Mechanical Engineering
Vancouver, Canada
muchiao@mech.ubc.ca



Chapter I.43

Mu Chiao received his degrees in Mechanical Engineering from National Taiwan University and the University of California Berkeley. After a postdoctoral at the Berkeley Sensor and Actuator Center and a position as Senior MEMS Engineer at Intpax Inc., he joined the Department of Mechanical Engineering of the University of British Columbia as an Assistant Professor. His current research interests include design and fabrication of MEMS and nanodevices for biomedical applications.



Ciro Chiappini

King's College London
Craniofacial Development and Stem Cell Biology
London, UK
ciro.chiappini@kcl.ac.uk

Chapter D.21

Dr Ciro Chiappini is a Lecturer in Nanomaterials and Biointerfaces at King's College London. He received a PhD in Biomedical Engineering from the University of Texas at Austin in 2011. He worked as Newton Fellow and Marie Curie Fellow at Imperial College London until 2016. Dr Chiappini develops functional interfaces for tissue engineering and precision medicine.



Lilia A. Chtcheglova

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
lilia.chtcheglova@jku.at

Chapter E.25

Dr Lilia Chtcheglova received her PhD from the Swiss Federal Institute of Technology in Lausanne. Afterwards, she joined the group of Peter Hinterdorfer and continues to work at Johannes Kepler University and Center for Advanced Bioanalysis. Her research interest is focused on nanoscopic approaches in molecular recognition studies and their applications in molecular cell biology.

Alessandro Coclite

Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Genova, Italy
alessandro.coclite@iit.it



Chapter D.21

Dr Alessandro Coclite received his PhD from Politecnico di Bari in 2015. He is now a Postdoctoral Fellow in the Laboratory of Nanotechnology for Precision Medicine. His research deals with the vascular transport of nanoconstructs via lattice Boltzmann-immersed boundary models, analyzing their properties in terms of shape, adhesive coating, and stiffness.

Maria Laura Coluccio

University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Catanzaro, Italy
coluccio@unicz.it



Chapter D.21

Dr Maria Laura Coluccio received her PhD in Chemical and Materials Engineering from the University of Pisa. She worked at the Italian Institute of Technology, the University of Reggio Calabria, and the University of Pisa before joining the University Magna Graecia of Catanzaro. Her research focuses on nanotechnology applied to biomedicine, including membranes, plants for bioextractions, drug delivery, and tissue engineering.



Paolo Decuzzi

Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Genova, Italy
paolo.decuzzi@iit.it

Chapter D.21

Professor Paolo Decuzzi is a Senior Researcher at the Italian Institute of Technology in Genova, where he founded the Laboratory of Nanotechnology for Precision Medicine in July 2015 with the support of a European Research Council Consolidator Grant. He investigates how to rationally design polymeric nanoconstructs for multi-modal imaging and combination therapy in cancer, cardiovascular, and neurological diseases.



Frank W. DelRio

National Institute of Standards & Technology
Material Measurement Laboratory
Boulder, USA
frank.delrio@nist.gov

Chapter I.39

Dr Frank W. DelRio is a Research Scientist in the Material Measurement Laboratory at the National Institute of Standards and Technology. His research interests include measurements and standards related to the elastic, plastic, fracture, interfacial, and transport properties of thin films and small-scale structures, with an emphasis on advance dmaterials in nanoelectronics, biomedical and health, and energy applications.

Enzo Di Fabrizio

King Abdullah University of Science and Technology
Dept. of Physical Science and Engineering
Thuwal, Saudi Arabia
enzo.difabrizio@kaust.edu.sa



Chapter D.21

Professor Enzo Di Fabrizio is a Full Professor at King Abdullah University of Science and Technology and Director of the BioNEM Laboratory at the University of Catanzaro. He is coauthor of more than 400 papers and inventor of more than 13 international patents. His scientific activity is interdisciplinary research between physics and biology, which includes basic and applied studies in nanotechnology.

Daniele Di Mascolo

Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Genova, Italy
daniele.dimascolo@iit.it



Chapter D.21

Daniele Di Mascolo received his PhD in Biomedical and Computer Engineering from Magna Graecia University in 2014. He worked at the Houston Methodist Research Institute before joining the Italian Institute of Technology. His research focuses on nano and micro-drug delivery systems for the pharmacological treatment of different pathologies.



Lixin Dong

Michigan State University
Dept. of Electrical & Computer Engineering
East Lansing, USA
ldong@egr.msu.edu

Chapter C.18

Lixin Dong is an Associate Professor at Michigan State University. Prior to joining MSU in December 2008, he held a Senior Research Scientist position at ETH Zurich, where he led the NanoRobotics Group in the Institute of Robotics and Intelligent Systems. He is Associate Editor of IEEE Trans. on Nanotechnology and the IEEE Trans. on Automation Science and Engineering. His research interests include nanorobotics, nanoelectromechanical systems, mechatronics, and nanobiomedical devices.



Mildred S. Dresselhaus (deceased)

Chapter B.9

Professor Mildred Dresselhaus (1930-2017) received her PhD in Physics from the University of Chicago in 1958. She joined the MIT faculty in 1967. She was active in research across broad areas of solid state physics. Her research activities focused on carbon nanomaterials, transition metal dichalcogenides, and phosphorene. She received several major international awards in physics and 37 honorary degrees.

Andreas Ebner

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
andreas.ebner@jku.at



Chapter E.25

Professor Andreas Ebner is an Associate Professor in the Institute of Biophysics at the Johannes Kepler University Linz. He received his PhD in Technical Chemistry from Johannes Kepler University Linz. After a postdoctoral fellowship at the Comenius University Bratislava he habilitated in bionanotechnology in 2016. His expertise lies in biosensing techniques, including AFM force spectroscopy and recognition imaging on biological samples.

Toshiaki Enoki

Tokyo Institute of Technology
Dept. of Chemistry
Tokyo, Japan
tenoki@chem.titech.ac.jp



Chapter B.10

Toshiaki Enoki has received his PhD from Kyoto University in 1974. He has worked at the Institute for Molecular Science in Japan and at Tokyo Institute of Technology. He is an honorary member of Ioffe Institute, Russia, and a Visiting Fellow of the Toyota Physical and Chemical Research Institute. His research in physical chemistry focuses on carbon-based nanostructures.



Haig Alexander Eskandarian

Swiss Federal Institute of Technology Lausanne (EPFL)
Dept. of Bioengineering
Lausanne, Switzerland
alex.eskandarian@epfl.ch

Chapter D.19

Dr Haig Alexander Eskandarian received his PhD from the Pasteur Institute in Paris, in 2013. He is a microbiologist working as an EMBO Long Term Post-Doctoral Fellow at the Swiss Federal Institute of Technology Lausanne (EPFL), in the microbiology lab of John D. McKinney and the bioengineering lab of Georg Fantner.



Georg E. Fantner

Swiss Federal Institute of Technology
Lausanne (EPFL)
Dept. of Bioengineering
Lausanne, Switzerland
georg.fantner@epfl.ch

Chapter D.19

Dr Georg Fantner received his PhD from the University of California Santa Barbara. After a postdoctoral position at MIT, he joined the faculty of the Swiss Federal Institute of Technology Lausanne (EPFL). His research focuses on the interface of micro/nanotechnology and biology. His group develops new time resolved scanning probe microscopy methods and related technologies, primarily for applications in life-science.

Philip X.-L. Feng

Case Western Reserve University
Dept. of Electrical Engineering &
Computer Science
Cleveland, USA
philip.feng@case.edu



Chapter C.13

Philip Feng is an Associate Professor for Electrical Engineering at Case Western Reserve University. His research is primarily focused on emerging nanoscale devices and integrated microsystems. He received a PhD from the California Institute of Technology. He is a recipient of the NSF CAREER Award and the NAE Grainger Foundation Frontiers of Engineering Award.

Emmanuel Flahaut

Centre National de la Recherche
Scientifique, Université de Toulouse,
Institut National Polytechnique de
Toulouse
Centre Interuniversitaire de Recherche
et d'Ingénierie des Matériaux (CIRIMAT),
UMR #5085
Toulouse, France
flahaut@chimie.ups-tlse.fr



Chapter B.8

Emmanuel Flahaut is a CNRS Senior Scientist and works at the CIRIMAT in Toulouse. After receiving a PhD in Materials Science from the University of Toulouse in 1999, he worked as a postdoctoral research fellow at ICL, University of Oxford. His research deals with nanocarbons (carbon nanotubes, graphene, and related materials), with a special focus on biomedical applications as well as potential health and environmental impacts.



Francesco Gentile

University Naples Federico II
Dept. of Electrical Engineering and
Information Technology
Naples, Italy
francesco.gentile77@gmail.com

Chapter D.21

Francesco Gentile is an Associate Professor at the University Federico II in Naples. He received his PhD in Biomedical Engineering from the University Magna Graecia of Catanzaro in 2008. Afterwards he worked at the Synchrotron Laboratories in Trieste, at the University of Texas-Houston, and at the Italian Institute of Technology, Genova. His research is within the field of biomedical nanotechnology.



Christoph Gerber

University of Basel
Dept. of Physics
Basel, Switzerland
christoph.gerber@unibas.ch

Chapter C.15

Christoph Gerber is a Professor in Physics at the University of Basel and an Emeritus Researcher in Nanoscale Science at IBM Rüschlikon. He has made major contributions to scanning tunneling microscopy and is co-inventor of the atomic force microscope, for which he was awarded the Kavli Prize in Nanoscience in 2016. His recent work is focused on biochemical cantilever array sensors based on AFM technology.

Thilo Glatzel

University of Basel
Dept.of Physics
Basel, Switzerland
thilo.glatzel@unibas.ch



Chapter F.29

Dr Thilo Glatzel received his PhD from the Free University of Berlin in 2003. His dissertation was focused on the nanoscale analysis of interfaces and surfaces of chalcopyrite thin film solar cells. Now, as a Senior Researcher at the University of Basel, his work focuses on high-resolution scanning probe microscopy of molecules on insulating and semiconducting surfaces.

Enrico Gnecco

Friedrich Schiller University Jena
Otto Schott Institute of Materials
Research
Jena, Germany
enrico.gnecco@uni-jena.de



Chapter F.29

Enrico Gnecco is Professor of Physics at Friedrich Schiller University Jena. He received his PhD from the University of Genova, and worked as Postdoctoral Researcher at the University of Basel and as independent Group Leader at IMDEA Nanoscience Madrid. His research focuses on nanoscale friction, scanning probe-based manipulation of metal clusters and nanocrystals, nanomechanisms of abrasive wear in polymers and glasses, and the influence of ultrasonic vibrations on friction.



Miroslav Haluška

ETH Zürich
Dept. of Mechanical & Process Engineering
Zurich, Switzerland
haluska@micro.mavt.ethz.ch

Chapter C.14

Dr Miroslav Haluška received his degree from the Slovak Technical University in Bratislava and his PhD in Physics from the University of Vienna. He was a Postdoctoral Fellow and Visiting Researcher at the Max Planck Institute, Wake Forest University, and Eindhoven University of Technology. In 2009, he joined the Micro and Nanosystems Group at ETH Zurich. His research includes the synthesis of SWCNTs, and their characterization and integration into nanosensors.



Judith A. Harrison

U.S. Naval Academy
Dept. of Chemistry
Annapolis, USA
jah@usna.edu

Chapter F.30

Judith A. Harrison is a Professor of Chemistry at the US Naval Academy. She received her degrees from St. Anselm College and the University of New Hampshire. Before joining the faculty of the Naval Academy, she was a Postdoctoral Associate at the Naval Research Laboratory. Her research focuses on the theoretical examination of nanometer-scale processes, such as indentation, friction, wear, and tribochemistry of hydrocarbon systems.

Martin Hegner

Trinity College Dublin
School of Physics, CRANN
Dublin, Ireland
martin.hegner@tcd.ie



Chapter C.15

Professor Martin Hegner received his PhD from the Swiss Federal Institute of Technology in Zurich. Since 2007 he has been a Professor in the School of Physics at Trinity College Dublin and Principal Investigator at the CRANN Nanocenter. Martin Hegner's scientific interests are focused on interdisciplinary research in the fields of single biomolecule manipulation, biophysics, diagnostics, and development and application of biological sensing devices.

Thomas Helbling

greenTEG AG
Zurich, Switzerland
thomas.helbling@greenteg.com



Chapter C.14

Dr Thomas Helbling received his MSc in Electrical Engineering from the ETH Zurich with a specialization in MEMS technology. In 2009, he received his PhD in Mechanical Engineering from ETH Zurich for the use of carbon nanotubes as strain gauge and gas sensing material. Since 2010 Thomas has been with greenTEG, a Swiss spin-off company from ETH Zurich that produces and sells thermal sensor solutions.



Seong-Jun Heo

Lam Research Corp.
Fremont, USA
heogyver@gmail.com

Chapter F.30

Dr Seong-Jun Heo is a Process Engineer in the Conductor Etch Division at Lam Research Corp. He received his degrees in Materials Engineering from Hanyang University in Seoul. He worked as a Research Engineer in the Semiconductor Division of Samsung Electronics, before pursuing his PhD at the University of Florida. He is interested in semiconductor processes as well as computational simulation related to nanotechnology.



Barbara Herr Harthorn

University of California Santa Barbara
Dept. of Anthropology
Santa Barbara, USA
barbara.harthorn@ucsb.edu

Chapter J.44

Professor Barbara Herr Harthorn received her PhD in Anthropology from the UCLA. She is Professor of Anthropology and the Director of the NSF Center for Nanotechnology in Society at the University of California at Santa Barbara, as well as an Executive Committee member of the UC Center for Environmental Implications of Nanotechnology at UCLA. Her research focuses on multi-stakeholder risk perception and public deliberation on societal aspects of new technologies.

Christofer Hierold

ETH Zürich
Dept. of Mechanical & Process Engineering
Zurich, Switzerland
hierold@micro.mavt.ethz.ch



Chapter C.14

Professor Christofer Hierold has been Professor of Micro and Nanosystems at ETH Zurich since 2002, after having worked with Siemens AG and Infineon Technologies AG. At ETH Zurich, his research is focused on the evaluation of new materials for MEMS, on advanced microsystems, and on nanotransducers, mainly carbon nanotube sensors. Christofer Hierold is a member of the Swiss Academy of Engineering Sciences.

Peter Hinterdorfer

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
peter.hinterdorfer@jku.at



Chapter E.25

Peter Hinterdorfer is Head of the Department of Applied Experimental Biophysics at the Institute of Biophysics of Johannes Kepler University Linz. In 2001 he was appointed Associate Professor and in 2010 Professor at the Institute of Biophysics at Johannes Kepler University Linz. He is a recognized expert in single-molecule force spectroscopy and high-resolution topography imaging of biological samples and inventor of simultaneous topography and recognition imaging.



Maria Iliut

University of Manchester
School of Materials
Manchester, UK
maria.iliut@manchester.ac.uk

Chapter B.12

Dr Maria Iliut received her PhD in Physics from Babes-Bolyai University. She now works as a Research Associate at the University of Manchester. Her research focuses on the production, properties, and applications of graphene and other two-dimensional materials.



Douglas L. Irving

North Carolina State University
Dept. of Materials Science & Engineering
Raleigh, USA
dlirving@ncsu.edu

Chapter F.30

Douglas L. Irving is an Associate Professor of Materials Science and Engineering at North Carolina State University. He received his degree in physics from Furman University and his degrees in materials science and engineering from the University of Florida. His research focuses on the study of materials in extreme environments using theoretical and computational approaches that range from multi-scale to electronic structure methods.

Jacob N. Israelachvili

University of California Santa Barbara
Dept. of Chemical Engineering & Materials Dept.
Santa Barbara, USA
jacob@engineering.ucsb.edu



Chapter F.28

Jacob Israelachvili earned his PhD in 1971 at the University of Cambridge. He held positions in the Department of Applied Mathematics at the Australian National University before joining the faculty at the University of California, Santa Barbara in 1986 as Professor of Chemical Engineering and Materials. He is Fellow of the Royal Society of London and member of the US National Academy of Engineering and the National Academy of Sciences.

Esmaiel Jabbari

University of South Carolina
Dept. of Chemical Engineering
Columbia, USA
jabbari@cec.sc.edu



Chapter B.11

Esmaiel Jabbari is a Full Professor of Chemical and Biomedical Engineering and Director of the Biomaterials and Tissue Engineering Laboratory at the University of South Carolina. He received his PhD in Chemical Engineering from Purdue University and is a Fellow of AIMBE. His research focuses on the development of multi-cellular tissue models for applications in regenerative medicine and cancer therapy.



Anne Jourdain

IMEC
Leuven, Belgium
jourdain@imec.be

Chapter I.40

Dr Anne Jourdain received the MSc degree in Opto-Microelectronics from the Ecole Nationale Supérieure d'Ingénieurs de Caën. She received her PhD from the University Joseph Fourier of Grenoble. In 1999, she joined the Interuniversity Microelectronics Center (IMEC) in Leuven, where she was involved in the MEMS processing development. In 2000, she joined the RF-MEMS team and is currently in charge of the packaging of RF-MEMS devices.



Harold Kahn

Charlottesville, USA
hxkahn@gmail.com

Chapter I.41

Dr Harold Kahn received his BS in Metallurgical Engineering from Lafayette College and his PhD in Electronic Materials from the Massachusetts Institute of Technology. He retired as Research Associate Professor in Materials Science and Engineering at Case Western Reserve University in Cleveland, where his research focused on wafer-level mechanical testing of MEMS devices.

Gopakumar Kamalakshakurup

University of California Irvine
Dept. of Biomedical Engineering
Irvine, USA
gkamalak@uci.edu



Chapter C.17

Dr Gopakumar Kamalakshakurup received his PhD from Wayne State University. He worked at BioRad Laboratories, Lawrence Berkeley National Labs prior to joining the University of California Irvine as a Postdoctoral Fellow. His research includes developing microfluidic platforms for low cost diagnostics. His work on optofluidic tweezers has resulted in a US patent.

Roger D. Kamm

Massachusetts Institute of Technology
Depts. of Mechanical Engineering
& Biological Engineering
Cambridge, USA
rdkamm@mit.edu



Chapter F.31

Roger Kamm is the Cecil and Ida Green Distinguished Professor of Biological and Mechanical Engineering. He received his PhD in Mechanical Engineering from MIT, where he has stayed for his entire career. Roger Kamm's research focuses on the application of fluid and solid mechanics to better understand essential biological and physiological phenomena.



Josef A. Käs

University of Leipzig
Peter Debye Institute
Leipzig, Germany
jkaes@uni-leipzig.de

Chapter D.20

Professor Josef A. Käs is Head of the Soft Matter Physics Division at the University of Leipzig. He received his PhD from the Technical University of Munich and did postdoctoral work at Harvard Medical School. His research interests include all fields of cell biophysics with a particular focus on cellular biomechanics probed by optical traps, as well as soft matter physics.



Jongbaeg Kim

Yonsei University
School of Mechanical Engineering
Seoul, Republic of Korea
kimjb@yonsei.ac.kr

Chapter I.43

Jongbaeg Kim received his degrees in mechanical engineering from Yonsei University, the University of Texas Austin, and the University of California, Berkeley. After working with DiCon Fiberoptics, Inc., where he designed and developed high-performance optical MEMS components for telecommunication applications, he joined Yonsei University. He is Assistant Professor in the School of Mechanical Engineering. His research interests are modeling, design and fabrication of microsystems, and integrated nanostructures on MEMS.

Marcin Kisiel

University of Basel
Dept. of Physics
Basel, Switzerland
marcin.kisiel@unibas.ch



Chapter F.29

Dr Marcin Kisiel received his PhD from Marie Curie-Sklodowska University of Lublin. Currently, he is a Postdoctoral Researcher at the University of Basel. His research interest focuses on friction and energy dissipation at the nanoscale under ultra-high vacuum conditions and at low temperatures.

Kerstin Koch

Rhine-Waal University of Applied Science
Dept. of Life Science, Biology & Nanobiotechnology
Kleve, Germany
koch@hochschule-rhein-waal.de



Chapter H.36

Professor Kerstin Koch received her Diploma, PhD and Habilitation in Biology from the University of Bonn. Her main topics of research include molecular self-assembly, molecular architecture of plant surface lipids, biological surfaces, origin and functions of micro and nanostructures, and superhydrophobic and superhydrophilic biomimetic surfaces. Since 2009 she has been Professor of Biology and Nanobiotechnology at the Rhine-Waal University of Applied Science, Kleve.



Melanie Koehler

Université Catholique de Louvain
Institute of Life Sciences
Louvain-la-Neuve, Belgium
melanie.kohler@uclouvain.be

Chapter E.25

Melanie Köhler received her PhD in Biophysics from the Johannes Kepler University of Linz. Her research focuses on single-molecule force spectroscopy, high-resolution imaging, and mapping of biological samples. Currently, she is a Postdoctoral Fellow in the Nanobiophysics Group at the Université Catholique de Louvain.



Anders Kristensen

Technical University of Denmark
DTU Nanotech
Kongens Lyngby, Denmark
anders.kristensen@nanotech.dtu.dk

Chapter A.5

Professor Anders Kristensen received his PhD in Physics from the University of Copenhagen in 1994. He joined DTU Nanotech in 2001, where he investigates optofluidics – the integration of micro/nano photonics and fluidics for novel sensing and actuation schemes. Professor Kristensen has headed national and European research programs on advancing polymer nano and micro fabrication technology for industrial production. He has pioneered nano-imprint lithography in Denmark.

Jan Lammerding

Cornell University
Meinig School of Biomedical Engineering
Ithaca, USA
jan.lammerding@cornell.edu



Chapter F.31

Jan Lammerding is Associate Professor in the Meinig School of Biomedical Engineering and the Weill Institute for Cell and Molecular Biology at Cornell University. He received degrees from Dartmouth College, RWTH Aachen University, and the Massachusetts Institute of Technology. His research focuses on developing novel experimental techniques to investigate the interplay between cellular mechanics and function, with a particular emphasis on the cell nucleus and its response to mechanical forces.

Constanze Lamprecht

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
constanze.lamprecht@gmx.de



Chapter E.25

Dr Constanze Lamprecht received her PhD in Biophysics from the Johannes Kepler University Linz. She worked as Postdoctoral Fellow at the University of Waterloo and the Christian-Albrechts-University of Kiel. She is currently affiliated again with the Institute of Biophysics at the Johannes Kepler University, where she uses AFM and fluorescence microscopy to study cell transformations that lead to the initiation and progression of cancer.



Hans Peter Lang

University of Basel
Dept. of Physics
Basel, Switzerland
hans-peter.lang@unibas.ch

Chapter C.15

Dr Hans Peter Lang received a PhD in Physics from the University of Basel in 1994. As Postdoctoral Fellow, he directed research on low temperature scanning tunneling microscopy in Basel, followed by work in the emerging field of nanomechanical cantilever array sensors at IBM Rüschlikon. At Basel University, he focuses on biochemical applications of cantilever array sensors.



Christophe Laurent

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
Toulouse, France
laurent@chimie.ups-tlse.fr

Chapter B.8

Christophe Laurent is a Professor of Materials Chemistry at the University of Toulouse and Director of the Interuniversity Center of Materials Research and Engineering (CIRIMAT). For over 20 years, he has dedicated his research to carbon nanotubes (CNTs), notably to the synthesis and characterization of double-walled CNTs and the preparation and properties of CNT-ceramic and CNT-metal nanocomposites in the form of powders, foams, and dense materials.

Abraham Lee

University of California Irvine
Dept. of Biomedical Engineering
Irvine, USA
aplee@uci.edu



Chapter C.17

Professor Abraham P. Lee is the William J. Link Professor and Chair of the Biomedical Engineering Department at the University of California Irvine. He is Director of the NSF I/UCRC Center for Advanced Design and Manufacturing of Integrated Microfluidics. Professor Lee is an Associate Editor for the Lab on a Chip journal and Fellow of the AIMBE and ASME.

Aeju Lee

Kumamoto University
International Research Organization for Advanced Science and Technology
Kumamoto, Japan
aeju-lee@kumamoto-u.ac.jp



Chapter D.21

Dr Aeju Lee received her PhD in Medicine from Korea University in 2013. She worked at the Korea Institute of Science and Technology and the Italian Institute of Technology, before joining Kumamoto University as an Associate Professor. Her research focuses on developing activatable NIR fluorescence imaging sensors, multimodality imaging, and therapy systems for early disease detection and therapy.



Dong Woog Lee

Ulsan National Institute of Science & Technology
School of Energy & Chemical Engineering
Ulsan, Republic of Korea
dongwoog.lee@unist.ac.kr

Chapter F.28

Dr Dong Woog Lee received his PhD from the University of California, Santa Barbara in 2014 and then did postdoctoral research at the same university. He is currently an Assistant Professor at Ulsan National Institute of Science and Technology (UNIST). His current research focuses on the adhesion and friction of bio and biomimetic materials. He received the Doh Wonsuk Memorial Award in 2013.



Stephen C. Lee

The Ohio State University
Dept. of Biomedical Engineering
Columbus, USA
lee.1996@osu.edu

Chapter D.22

Stephen C. Lee is a pioneer in the field of semi-biological nanodevices, having published the first monograph devoted to the topic in 1998. His interests are in enabling technologies for the incorporation of functional proteins and nucleic acids into nanodevices, particularly for application in oncology and cardiovascular disease. He is currently Associate Professor of Cellular and Molecular Biochemistry, Chemical Engineering and Biomedical Engineering at the Ohio State University.

Liwei Lin

University of California Berkeley
Dept. of Mechanical Engineering
Berkely, USA
lwlin@me.berkeley.edu



Chapter I.43

Liwei Lin is Professor in the Mechanical Engineering Department and Co-Director of the Berkeley Sensor and Actuator Center at the University of California Berkeley. His research interests are design, modeling, and fabrication of micro/nano structures, micro/nano sensors, and micro/nano actuators, and mechanical issues in micro/nano systems, including heat transfer, solid/fluid mechanics and dynamics. He led the effort to establish the MEMS division in ASME and served as its founding Chairman.

Matthias Mail

University of Bonn
Nees-Institute for Biodiversity of Plants
Bonn, Germany
matthias.mail@uni-bonn.de



Chapter H.36

Matthias Mail is a graduate physicist from the Karlsruhe Institute of Technology and doctoral student of Wilhelm Barthlott. He leads the working group on functional aspects of plant surfaces at the Institute of Crop Science and Resource Conservation at the University of Bonn. His main interests are structures, functions, and biomimetic applications of plant surfaces.



Othmar Marti

Ulm University
Institute of Experimental Physics
Ulm, Germany
othmar.marti@uni-ulm.de

Chapter E.23

Professor Othmar Marti is a specialist for scanning probe microscopy and optical force measurements. His current research is focused on life science (cell mechanics, cytoskeleton systems), optics, plasmonics, and combinations thereof. He is the Director of the Institute of Experimental Physics and a Chair for Experimental Physics at Ulm University.



Shinji Matsui

University of Hyogo
Laboratory of Advanced Science & Technology for Industry
Hyogo, Japan
matsui@lasti.u-hyogo.ac.jp

Chapter A.4

Dr Shinji Matsui is Professor at the University of Hyogo. He obtained his PhD degree in Electrical Engineering from Osaka University in 1981. Prior to joining the University of Hyogo, he worked at NEC Corporation. His significant work included the demonstration of electron-beam induced deposition and atomic-beam holography. His current research is focused on three-dimensional nanofabrication by focused-ion-beam chemical-vapor-deposition and nanoimprinting at room temperature.



Etienne Menard

HeliosLite
Le Bourget du Lac, France
etienne.menard@helioslite.com

Chapter A.6

Dr Etienne Menard received a PhD in Chemistry from the University *Pierre et Marie Curie* and the Department of Material Science and Engineering at the University of Illinois at Urbana-Champaign. He is the lead inventor of Semprius' core transfer printing technology. He is currently the Chief Technical Officer and co-founder of the company HéliosLite, which develops innovative solar trackers for high concentration photovoltaic systems and standard flatplate photovoltaic modules.



David Mesguich

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
Toulouse, France
mesguich@chimie.ups-tlse.fr

Chapter B.8

David Mesguich is an Assistant Professor for Materials Chemistry at University Paul Sabatier Toulouse III. He is part of the Nanocomposites and Carbon Nanotubes team in the Interuniversity Center of Materials Research and Engineering (CIRIMAT). Here he works on nanocarbon synthesis by catalytic chemical vapor Deposition and preparation of metal matrix nanocomposites and carbides, both as powders and bulk materials sintered by SPS.



Vincent Meunier

Rensselaer Polytechnic Institute
Dept. of Physics, Applied Physics, & Astronomy
Troy, USA
meuniv@rpi.edu

Chapter B.9

Professor Vincent Meunier is the Head of the Physics, Applied Physics, and Astronomy Department at Rensselaer Polytechnic Institute, where he holds the Kodosky Constellation Chair since 2010. He earned his PhD from the University of Namur. He has published approximately 200 papers in peer-reviewed journals and is a Fellow of the American Physical Society.



Ernst Meyer

University of Basel
Dept. of Physics
Basel, Switzerland
ernst.meyer@unibas.ch

Chapter F.29

Ernst Meyer is a Professor of Physics at the University of Basel. He is interested in friction force and dynamic force microscopy with true atomic resolution. He is active in the field of sensors based upon micromechanics and magnetic spin resonance detection with force microscopy. He is member of the Swiss and American Physical Society, of the Editorial Board of Tribology Letters, and co-editor of books on atomic force microscopy.



Tom D. Milster

University of Arizona
College of Optical Sciences
Tucson, USA
milster@optics.arizona.edu

Chapter E.26

Professor Tom D. Milster graduated with a degree in electrical engineering from the University of Missouri-Rolla and received a PhD in Optical Sciences from the University of Arizona. He worked for IBM Corporation before joining the University of Arizona in 1989. Professor Milster teaches physical optics and has active research programs in microscopy, lithography, and data storage.

Hiroaki Misawa

Hokkaido University
Research Institute of Electronic Science
Sapporo, Japan
misawa@es.hokudai.ac.jp



Chapter A.4

Hiroaki Misawa is a Professor at Hokkaido University. He received his PhD in Chemistry from the University of Tsukuba in 1984. Prior to joining Hokkaido University, he worked at the University of Tsukuba, Microphotoconversion project (ERATO) of JST, and the University of Tokushima. Since 2015, he has also been a Professor at National Chiao Tung University, Taiwan. He studies plasmonic chemistry.



Robert Modliński

EPCOS AG
Munich, Germany
robert.modlinski@epcos.com

Chapter I.40

Dr Robert Modliński received his degrees in electronics and microsystems from Wrocław University of Technology and KU Leuven. His main research activities are NEMS/MEMS and CMOS-MEMS integrated products with a strong focus on competitiveness, reliability and price, testing, and characterization. At EPCOS, a TDK Group Company, he focuses on failure mechanisms and degradation models to predict the remaining useful life of MEMS devices and RF front-end modules for mobile applications.



Seyedsina Moeinzadeh

University of South Carolina
Dept. of Chemical Engineering
Columbia, USA
moeinzad@email.sc.edu

Chapter B.11

Dr Seyedsina Moeinzadeh is a Postdoctoral Fellow at the University of South Carolina. He received his PhD in Chemical Engineering from the University of South Carolina for his research on the development of degradable nanostructured hydrogels for musculoskeletal tissue engineering. His primary research interests include tissue engineering, biomaterials, nanomaterials, polymers, and nano/microfabrication.

Mohammad R. K. Mofrad

University of California Berkeley
Depts. of Bioengineering & Mechanical Engineering
Berkeley, USA
mofrad@berkeley.edu



Chapter F.31

Dr Mohammad R.K. Mofrad is a Professor of Bioengineering and Mechanical Engineering at the University of California Berkeley. His research focuses on the development of molecular and multiscale models of cell mechanobiology, with the ultimate aim to shed light on human diseases. He received the National Science Foundation CAREER award in 2010 and the American Heart Association Innovative Research Grant Award in 2016. He is Fellow of AIMBE.

Marc Monthioux

Centre National de la Recherche Scientifique, Université de Toulouse
Centre d'Elaboration de Matériaux et d'Etudes Structurales (CEMES), UPR #8011
Toulouse, France
marc.monthioux@cemes.fr



Chapter B.8

Marc Monthioux has been working on carbon materials for over 35 years, including carbon nanoforms since 1998. He is Director of Research at the French National Center for Scientific Research in Toulouse, and served as Editor for the journal Carbon for 13 years. He is former Chair of the French Carbon Society, former Chair of the European Carbon Association, and was Chair of the World Conference on Carbon in 2009.



Markus Morgenstern

RWTH Aachen University
II. Institute of Physics B & JARA-FIT
Aachen, Germany
mmorgens@physik.rwth-aachen.de

Chapter E.24

Markus Morgenstern earned his PhD from the Institute of Interface Research and Vacuum Physics of the Forschungszentrum Jülich (1996). In 2002 he completed his habilitation at the University of Hamburg with the subject scanning tunneling spectroscopy on semiconductor systems and nanostructures. In 2004 he became Full Professor at RWTH Aachen University. From 2007–2014, he was Scientific Director of JARA-FIT, the Jülich–Aachen Research Alliance – Fundamentals of Future Information Technology.



Bradley J. Nelson

Swiss Federal Institute of Technology
Institute for Robotics & Intelligent Systems, ETH Zentrum, CLA H 1.1
Zurich, Switzerland
bnelson@ethz.ch

Chapter C.18

Brad Nelson is Professor and Director of the Institute of Robotics and Intelligent Systems at ETH Zurich. He received a PhD in Robotics from Carnegie Mellon University. In 2002, he became a Professor at ETH Zurich. His primary research direction lies in extending robotics research into emerging areas of science and engineering, with a special emphasis in the area of micro and nanorobotics.



T. Kieran Nunan

Kieran Nunan Consulting
Carlisle, USA
tknunan25@gmail.com

Chapter I.42

T. Kieran Nunan received his degree in electronics from the University of Limerick Ireland, while working with Analog Devices Inc. Kieran's career with ADI has landed him multiple roles on both sides of the Atlantic. He was the Engineering Group Leader in Diffusion CVD and TF manufacturing and also helps process development roles in BICMOS and MEMS, including managing the MEMS development group.



Pascal D. Odermatt

Swiss Federal Institute of Technology
Lausanne (EPFL)
Dept. of Bioengineering
Lausanne, Switzerland
pascal.odermatt@epfl.ch

Chapter D.19

Dr Pascal D. Odermatt received his PhD from the Swiss Federal Institute of Technology Lausanne in 2016. He is currently working at the University of California San Francisco and Stanford University as a Postdoctoral Scholar. His research focuses on studying functional and mechanical aspects of physiological processes at the nanoscale on a tissue and single-cell level.



Yoojin Oh

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
yoo_jin.oh@jku.at

Chapter E.25

Dr Yoojin Oh received her PhD from Ewha Womans University in Seoul. She is currently working at the Institute of Biophysics, Johannes Kepler University Linz. Her main focus of research is the investigation of bacteria and their strategies to adhere to abiotic and biotic surfaces, and their capability of forming biofilms using different atomic force microscopy-based methods.



Anna Lisa Palange

Italian Institute of Technology
Laboratory of Nanotechnology for Precision Medicine
Genova, Italy
annalisa.palange@iit.it

Chapter D.21

Anna Lisa Palange received her PhD from the University Magna Graecia of Catanzaro in 2015. She is now a Postdoctoral Researcher in the Laboratory of Nanotechnology for Precision Medicine. Her actual interests deal with the synthesis, characterization, and in vitro and in vivo application of discoidal polymeric nanoconstructs for imaging and therapeutic studies.



Rémy Pawlak

University of Basel
Dept. of Physics
Basel, Switzerland
remy.pawlak@unibas.ch

Chapter F.29

Dr Rémy Pawlak received his PhD from the Aix-Marseille University in 2010. He is currently a Postdoctoral Fellow at the University of Basel. His research focuses on low-temperature investigations of single molecules and atoms by means of atomic force microscopy and scanning tunneling microscopy. He is particularly interested in high-resolution imaging and force spectroscopy.



Alain Peigney

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
Toulouse, France
peigney@chimie.ups-tlse.fr

Chapter B.8

Professor Alain Peigney is a Ceramic Engineer with a PhD in Physical Chemistry. He is an Associate Professor of Materials Chemistry at the University Paul Sabatier Toulouse III. His research encompasses the synthesis, sintering, and microstructural characterization of ceramics and ceramic matrix nanocomposites. Since 1994, he has been concentrating on the synthesis of single and double-walled carbon nanotubes and the preparation of nanocomposites containing carbon nanotubes.



Gerardo Perozziello

University Magna Graecia of Catanzaro
Dept. of Experimental and Clinical Medicine
Catanzaro, Italy
gerardo.perozziello@unicz.it

Chapter D.21

Dr Gerardo Perozziello works in the nanotechnology laboratory of the Department of Experimental and Clinical Medicine of the University Magna Graecia of Catanzaro, after receiving degrees in Mechanical Engineering and Micro- and Nanotechnology. His research interests and activities range from microfluidics to micro and nanofabrication technologies applied to biomedical issues.



Sandra Posch

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
sandra.posch@jku.at

Chapter E.25

Sandra Posch received her PhD in Biophysics from the Johannes Kepler University of Linz. As a Postdoctoral Fellow, she studies the blood protein von Willebrand factor by AFM and SPR. In 2012, she visited the Institute of Biomedical Technology in Tampere to learn standard techniques of molecular biology. In 2013, she received the Upper Austria Research Award.

Robert Puers

KU Leuven
ESAT/MICAS
Leuven, Belgium
robert.puers@kuleuven.be



Chapter I.40

Professor Robert Puers is a Full Professor at KU Leuven.. In 1986, he became Director of the Clean Room Facilities for Silicon and Hybrid Circuit Technology at the ESAT-MICAS Laboratories of the same university. He was a pioneer in the European research efforts in silicon micromachined sensors, MEMS, and packaging techniques, for biomedical implantable systems as well as for industrial devices.

Oded Rabin

University of Maryland
Dept. of Materials Science & Engineering
College Park, USA
oded@umd.edu



Chapter B.9

Oded Rabin is an Associate Professor of Materials Science and Engineering at the University of Maryland. He earned degrees in chemistry from the Technion – Israel Institute of Technology and the Weizmann Institute of Science, and received a PhD in Physical Chemistry from the Massachusetts Institute of Technology. His research interests include electrical and thermal transport in low dimensional systems, plasmonic nanostructures for spectroscopy, and nanoparticles as biomedical contrast agents.



Françisco M. Raymo

University of Miami
Dept. of Chemistry
Coral Gables, USA
fraymo@miami.edu

Chapter A.2

Françisco M. Raymo is a Professor in the Department of Chemistry at the University of Miami. His research interests lie at the interface of chemistry and biology. In particular, he develops photo-switchable fluorescent probes to image biological samples with nano-scaled resolution and to monitor dynamic processes.



Manitra Razafinimanana

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire des Plasmas et de Conversion de l'Energie (LAPLACE), UMR #5002
Toulouse, France
razafinimanana@laplace.univ-tlse.fr

Chapter B.8

Manitra Razafinimanana is a Professor in the Laboratory on Plasma and Conversion of Energy (LAPLACE – UMR 5213) at the University Paul Sabatier Toulouse III. He has worked on plasma diagnostics, transport coefficients, and calculation of thermodynamical properties. His research at LAPLACE focuses both on arc–material interaction and synthesis of nanomaterials by arc plasma.

Pouya Rezai

York University
Dept. of Mechanical Engineering
Toronto, Canada
pouya.rezai@lassonde.yorku.ca,
prezai@yorku.ca



Chapter C.16

Professor Pouya Rezai works in the field of bio-microfluidics. He was a Postdoctoral Fellow at the Public Health Agency of Canada before joining York University as an Assistant Professor in 2013. He received his PhD from McMaster University. His research interests include advancing microfabrication techniques to develop Lab-on-a-Chip for studying interactions between bio-substances and fluids in microenvironments.

Gemma Rius

Barcelona Microelectronics Institute
(IMB-CNM-CSIC)
NEMS & Nanofabrication Group
Bellaterra, Spain
gemma.rius@imb-cnm.csic.es



Chapter A.3

Dr Gemma Rius graduated from UAB. From 2002 to 2008, she was a Nanolithography Engineer and PhD student at the Barcelona Microelectronics Institute (IMB-CNM, CSIC). After postdoctoral work at the Toyota Technological Institute at Tohoku University and an Assistant Professorship at Nagoya Institute of Technology, she became a Beatriu de Pinós Postdoctoral Fellow at IMB-CNM and Principal Investigator for the SPS-NATO project RAWINTS, mainly working in applications of nanostructured materials.



John A. Rogers

Northwestern University
Dept. of Materials Science & Engineering
Evanston, USA
jrogers@northwestern.edu

Chapter A.6

Professor John A. Rogers received a PhD in Physical Chemistry from MIT. He was a Technical Staff Member at Bell Laboratories and served as Director of the Condensed Matter Physics Research Department there. He joined Northwestern University in 2016 as the Louis Simpson and Kimberly Querrey Professor of Materials Science and Engineering, Biomedical Engineering and Medicine, where he is the founding Director of the newly endowed Center on Bio-Integrated Electronics.



Cosmin I. Roman

ETH Zürich
Dept. of Mechanical & Process Engineering
Zurich, Switzerland
cosmin.roman@micro.mavt.ethz.ch

Chapter C.14

Dr Cosmin Roman received his MSc in Electrical Engineering from the Polytechnic University of Bucharest, and his PhD in Micro/Nanoelectronics from the National Polytechnic Institute of Grenoble for his work on modeling carbon nanotube sensors. Since 2006, he has been a Senior Researcher at ETH Zurich. His current research interests include micro and nanotransducers for energy-efficient sensor systems.

Marina Ruths

University of Massachusetts Lowell
Dept. of Chemistry
Lowell, USA
marina_ruths@uml.edu



Chapter F.28

Marina Ruths received her PhD from the University of California, Santa Barbara in 1996. She then did postdoctoral research at the University of Illinois at Urbana-Champaign, at the Max-Planck-Institute for Polymer Research in Mainz, and at Åbo Akademi University. Her current research includes friction, adhesion, and nanorheology studies of polymer, liquid crystal, and surfactant systems.

Hélder A. Santos

University of Helsinki
Division of Pharmaceutical Chemistry and Technology
Helsinki, Finland
helder.santos@helsinki.fi



Chapter D.21

Dr Hélder A. Santos obtained his degree in chemical engineering at the Helsinki University of Technology. Currently, he is an Adjunct Professor in Pharmaceutical Nanotechnology at the University of Helsinki. Dr Santos has published more than 160 scientific publications. He also serves as journal editor and is on the Editorial Board of several international journals.



Shintaro Sato

Fujitsu Laboratories Ltd.
Devices & Materials Laboratories
Atsugi, Kanagawa, Japan
sato.shintaro@jp.fujitsu.com

Chapter B.10

Shintaro Sato received his PhD from the University of Minnesota. He joined Fujitsu Limited in 2001. From 2010–2014 he was a group leader at the National Institute of Advanced Industrial Science and Technology. As a Research Manager at Fujitsu Laboratories Ltd. he now researches and develops nanoelectronics devises using graphene and carbon nanotubes.



Helmut Schift

Paul Scherrer Institute
Laboratory for Micro- & Nanotechnology
Villigen PSI, Switzerland
helmut.schift@psi.ch

Chapter A.5

Dr Helmut Schift is Head of the Polymer Nanotechnology Group in the Laboratory for Micro and Nanotechnology (LMN) at the Paul Scherrer Institute in Villigen. He received his PhD from the Institute of Microtechnology Mainz. His work includes all aspects of replication technology, from stamp manufacturing to polymer nanorheology. He has more than 20 years of experience in the development of nanoimprint lithography as an alternative nanopatterning method.

Jörg Schnauß

University of Leipzig
Peter Debye Institute
Leipzig, Germany
joerg.schnauss@uni-leipzig.de



Chapter D.20

Dr Jörg Schnauß received his PhD in Physics from the University of Leipzig. Subsequently, he joined the Division of Soft Matter Physics of the University of Leipzig and the Fraunhofer Institute for Cell Therapy and Immunology as a Postdoctoral Researcher. His research focuses on biological and biomimetic, semiflexible polymers. He transfers concepts to the cellular and subcellular level to explore new therapeutic approaches.

Alexander Schwarz

University of Hamburg
Institute of Nanostructure and Solid State Physics
Hamburg, Germany
aschwarz@physnet.uni-hamburg.de



Chapter E.24

Dr Alexander Schwarz belongs to the scientific staff of the Center of Microstructure Research at the Institute of Applied Physics at the University of Hamburg. He has over 20 years of experience in his field. He is a Senior Scientist and works on atomic-force microscopy with atomic resolution as well as related magnetically sensitive techniques, i.e., magnetic force microscopy and magnetic exchange force microscopy.



Udo D. Schwarz

Yale University
Dept. of Mechanical Engineering & Materials Science
New Haven, USA
udo.schwarz@yale.edu

Chapter E.24

Dr Udo D. Schwarz received his PhD from the University of Basel. Subsequently he moved to the University of Hamburg, specializing in low-temperature scanning force microscopy and nanotribology. After 1 year at the Lawrence Berkeley National Laboratory, he accepted a position as Associate Professor at Yale University and was promoted to Full Professor in 2009. Since 2012, he has been Chair of the Department of Mechanical Engeering and Materials Science.



P. Ravi Selvaganapathy

McMaster University
Dept. of Mechanical Engineering
Hamilton, Canada
selvaga@mcmaster.ca

Chapter C.16

P. Ravi Selvaganapathy is a Professor in Mechanical Engineering and the Canada Research Chair in Biomicrofluidics at McMaster University. He received his PhD from the University of Michigan in 2002 and worked at the Sandia National Laboratories. His research interests are in the development of microfluidic devices for drug discovery, drug delivery, diagnostics, and artificial organs.

Philippe Serp

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire de Chimie de Coordination (LCC), UPR #8241
Toulouse, France
philippe.serp@ensiacet.fr



Chapter B.8

Philippe Serp has been Professor of Inorganic Chemistry at Toulouse University since 2005. His research interests include the preparation and use of nanostructured catalytic materials, a field in which he has published over 150 papers. He holds 20 patents and received Division Awards of the French Chemical Society in Catalysis (2004) and Industrial Chemistry (2012).

Sameer S. Shroff

Carnegie Mellon University
Dept. of Mechanical Engineering
Pittsburgh, USA
sshroff@alumni.cmu.edu



Chapter I.39

Dr Sameer S. Shroff received his degree in mechanical engineering from the University of Pittsburgh. He worked in the Micro and Nanomechanics Division of Professor de Boer at Carnegie Mellon University and received his PhD in 2015, where he was supported by the NSF Graduate Research Fellowship Program. He is currently employed by the Bechtel Marine Propulsion Corporation.



Mark G. da Silva

Analog Devices Inc.
High Performance Sensors
Wilmington, USA
mark.dasilva@analog.com

Chapter I.42

Dr Mark G. da Silva received his PhD from The Johns Hopkins University at Baltimore. He worked in academia at the University of Maryland and then in industry at Coventor, and Analog Devices in the US. Over the past 20 years, he has worked extensively in MEMS product development and currently manages development of breakthrough MEMS inertial sensor products for high performance applications.



Susan B. Sinnott

The Pennsylvania State University
Dept. of Materials Science & Engineering
University Park, USA
sbs5563@psu.edu

Chapter F.30

Dr Susan B. Sinnott is Professor and Department Head of Materials Science and Engineering at Pennsylvania State University. She received her PhD in Physical Chemistry from Iowa State University and was on the faculty at the University of Kentucky and the University of Florida. Her research focuses on the development and application of computational methods to design and develop new materials and to reveal structure–property relationships.

David M. Smith

Fraunhofer Institute for Cell Therapy & Immunology (IZI)
DNA Nanodevices Group
Leipzig, Germany
david.smith@izi.fraunhofer.de



Chapter D.20

Dr David M. Smith received his PhD for studies on phase transitions in active polymer networks from the University of Leipzig. After a postdoctoral at Ludwig Maximilian University of Munich focusing on DNA nanotechnology, he joined the Fraunhofer Institute for Cell Therapy and Immunology as the Head of the DNA Nanodevices Unit. His research explores new biomedical applications of DNA nanostructures and to study fundamental aspects of soft matter physics.

W. Merlijn van Spengen

Falco Systems
Amsterdam, The Netherlands
w.m.vanspengen@tudelft.nl



Chapter I.40

Dr W. Merlijn van Spengen is the Director of Falco Systems, a company that designs and manufactures technology leading high voltage amplifiers for company R&D, institutes, and universities worldwide. His part-time university research at the Technical University Delft blends physics, chemistry, electronics, and micromechanics to understand the reliability of micromachines. He received the prestigious Dutch Veni and Vidi grants.



Jürgen Strasser

Center for Advanced Bioanalysis GmbH
Linz, Austria
juergen.strasser@cbl.at

Chapter E.25

Jürgen Strasser studied medical and pharmaceutical biotechnology as well as biophysics, and is currently a PhD student at the Johannes Kepler University and the Center for Advanced Bioanalysis GmbH in Linz. His work involves the detailed biophysical investigation of immune responses in the human body using single-molecule force spectroscopy, high-speed atomic force microscopy and quartz crystal microbalance.



Yu-Chuan Su

National Tsing Hua University
Dept. of Engineering & System Science
Hsinchu, Taiwan
ycsu@ess.nthu.edu.tw

Chapter I.43

Yu-Chuan Su received his degree in power mechanical engineering from the National Tsing Hua University. He received his PhD in Mechanical Engineering from the University of California Berkeley. In 2004, he joined the National Tsing Hua University and is currently a Professor in the Department of Engineering and System Science. His research interests are design and fabrication of polymer-based nano and microfluidic systems for biomedical and energy related applications.

Quan Sun

Hokkaido University
Research Institute of Electronic Science
Sapporo, Japan
quansun@es.hokudai.ac.jp



Chapter A.4

Quan Sun is an Assistant Professor at Hokkaido University. He received his PhD in Physics from Peking University in 2006. Prior to working at Hokkaido University, he worked at Laval University, Canada, as a postdoctoral researcher. His main research interests include femtosecond laser micro/nanofabrication and nanophotonics.

Flavien Valensi

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Laboratoire des Plasmas et de Conversion de l'Energie (LAPLACE), UMR #5002
Toulouse, France
valensi@laplace.univ-tlse.fr



Chapter B.8

Flavien Valensi is an Associate Professor in the Laboratory of Plasma and Conversion of (LAPLACE – UMR 5213) in Toulouse. He received a PhD in Plasma Physics from Orleans University in 2007. Since 2009 his research at LAPLACE has been dealing with the experimental study of continuous and transient arcs, with applications such as synthesis of nanoparticles (carbon, metal) by arc plasma.



Derek Vallejo

University of California Irvine
Dept. of Biomedical Engineering
Irvine, USA
dhvallej@gmail.com

Chapter C.17

Dr Derek Vallejo received his PhD in 2016 from the University of California Irvine (UCI) under the guidance of Dr Abraham Lee. His thesis focused on the microfluidic production of biocompatible and storable vesicles and multisomes. Currently, Derek is a Postdoctoral Scholar in Dr John Chaput's laboratory at UCI, developing a microfluidic system for rapid directed evolution of XNA polymerases.



Aravind Vijayaraghavan

University of Manchester
School of Materials
Manchester, UK
aravind@manchester.ac.uk

Chapter B.12

Dr Aravind Vijayaraghavan received his PhD in Materials Engineering from Rensselaer Polytechnic Institute in 2006. He was an Alexander von Humboldt Fellow at the Karlsruhe Institute of Technology, before he started to work as a Lecturer in Nanomaterials at the University of Manchester. His research focuses on the production, properties, and applications of graphene and other two-dimensional materials.

Alicia Weibel

Centre National de la Recherche Scientifique, Université de Toulouse, Institut National Polytechnique de Toulouse
Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux (CIRIMAT), UMR #5085
Toulouse, France
weibel@chimie.ups-tlse.fr



Chapter B.8

Dr Alicia Weibel received her PhD in Materials Science in 2004. Since 2006, she has been an Assistant Professor at the University Paul Sabatier Toulouse III. Her research activities involve the synthesis, spark plasma sintering, and microstructural and mechanical characterization of ceramics and nanocomposites containing carbon nanotubes.

Kurt Winkelmann

Florida Institute of Technology
Dept. of Chemistry
Melbourne, USA
kwinkel@fit.edu



Chapter J.47

Kurt Winkelmann is an Associate Professor in the Chemistry Department at the Florida Institute of Technology. His research interests include kinetics of nanoparticle-catalyzed photochemical reactions, toxicology of engineered and naturally occurring nanoparticles, and chemical education research of student learning and attitudes in the teaching laboratory.



Darrin J. Young

University of Utah
Dept. of Electrical & Computer Engineering
Salt Lake City, USA
darrin.young@utah.edu

Chapter C.13

Darrin J. Young received his degrees and PhD from UC Berkeley. He pioneered RF MEMS high-Q tunable passive devices for wireless communications. He worked at the Electrical Engineering and Computer Science Department at Case Western Reserve University, before joining the Electrical and Computer Engineering Department of the University of Utah as USTAR Associate Professor in 2009. His research interests include MEMS design, fabrication, low power ASIC design, and microsystem integration.



Rong Zhu

Johannes Kepler University Linz
Institute of Biophysics
Linz, Austria
rong.zhu@jku.at

Chapter E.25

Dr Rong Zhu earned received his PhD in Biomedical Engineering from Southeast University in China. Currently he is working in Professor Hinterdorfer's group at the Institute of Biophysics of the Johannes Kepler University Linz, focusing on nanopharmacology and molecular recognition.



Babak Ziaie

Purdue University
School of Electrical & Computer Engineering
West Lafayette, USA
bziaie@purdue.edu;
ziaiebabak@gmail.com

Chapter A.3

Babak Ziaie received his PhD in Electrical Engineering from the University of Michigan in 1994. From 1995 to 1999 he was a postdoctoral researcher at the Center for Integrated Microsystems (CIMS) of the University of Michigan. He subsequently joined the Electrical and Computer Engineering Department of the University of Minnesota as an Assistant Professor. Since 2005, he has been with the School of Electrical and Computer Engineering of Purdue University, where he is currently a Professor. His research interests are related to BioMEMS and include implantable microsystems, micromachined interfaces with the central nervous system, and ultra-sensitive sensors for biological applications. Dr Ziaie is the recipient of the NSF Career Award in Biomedical Engineering (2001) and McKnight Endowment Fund Award for Technological Innovations in Neuroscience (2002), and a member of the IEEE and the American Association for the Advancement of Science.



Christian A. Zorman

Case Western Reserve University
Dept. of Electrical Engineering & Computer Science
Cleveland, USA
caz@case.edu

Chapters A.7, C.13

Christian A. Zorman is a Professor for Electrical Engineering at Case Western Reserve University with secondary appointments in Biomedical Engineering and Mechanical Engineering. He has authored over 250 peer-reviewed technical publications on topics related to the development of novel, enabling materials and the requisite processing techniques for MEMS and NEMS with an emphasis on applications in challenging environments.

Detailed Contents

List of Abbreviations XXXIII

1 **Introduction to Nanotechnology**
Bharat Bhushan 1
1.1 Nanotechnology – Definitions and Examples 2
1.1.1 Nanomanufacturing 2
1.1.2 MEMS/NEMS 2
1.1.3 Convergence 4
1.1.4 Industrial Applications 4
1.2 History and Early Research Expenditures 6
1.3 Governance of the National Nanotechnology Initiative 7
1.4 Nanotechnology R&D Funding Since 2001 8
1.5 Worldwide R&D Investments and Output 11
1.5.1 R&D Investments 11
1.5.2 Output 11
1.6 Translation of Basic Research to Technology Commercialization 15
1.7 Nanoscience and Nanotechnology Education 16
1.7.1 Interdisciplinary Education 16
1.7.2 K–12 Education, Excitement, and Creativity 16
1.7.3 Vocational, Undergraduate, Graduate, and Continuing Education Programs 16
1.7.4 Educator Training 17
1.7.5 Informal Science Education 17
1.8 Summary and Outlook 17
1.9 Organization of the Handbook 18
References 18

Part A Micro/Nanofabrication Techniques

2 **Molecule-Based Devices**
Françisco M. Raymo 23
2.1 Chemical Approaches to Nanostructured Materials 24
2.1.1 From Molecular Building Blocks to Nanostructures 24
2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins 24
2.1.3 Chemical Synthesis of Artificial Nanostructures 26
2.1.4 From Structural Control to Designed Properties and Functions 27
2.2 Molecular Switches and Logic Gates 28
2.2.1 From Macroscopic to Molecular Switches 28
2.2.2 Digital Processing and Molecular Logic Gates 29
2.2.3 Molecular AND, NOT and OR Gates 29
2.2.4 Combinational Logic at the Molecular Level 31
2.2.5 Intermolecular Communication 32
2.3 Solid-State Devices 35
2.3.1 From Functional Solutions to Electroactive and Photoactive Solids 35

2.3.2 Langmuir–Blodgett Films 36
2.3.3 Self-Assembled Monolayers 40
2.3.4 Nanogaps and Nanowires 43
2.4 Conclusions and Outlook 46
References 47

3 **Introduction to Micro-/Nanofabrication**
Gemma Rius, Antoni Baldi, Babak Ziaie, Massood Z. Atashbar 51
3.1 Basic Microfabrication Techniques 52
3.1.1 Lithography 52
3.1.2 Thin-Film Deposition and Doping 53
3.1.3 Etching and Substrate Removal 58
3.1.4 Substrate Bonding 62
3.2 MEMS Fabrication Techniques 62
3.2.1 Bulk Micromachining 63
3.2.2 Surface Micromachining 66
3.2.3 High-Aspect-Ratio Micromachining 70
3.3 Nanofabrication Techniques 74
3.3.1 E-Beam Nanofabrication 75
3.3.2 Epitaxy and Strain Engineering 75
3.3.3 Scanning Probe Microscope Techniques 75
3.3.4 Self-Assembly and Template Manufacturing 78
3.4 Summary and Conclusions 82
References 82

4 **3-D Nanostructure Fabrication by Focused-Ion Beam, Electron- and Laser Beam**
Shinji Matsui, Hiroaki Misawa, Quan Sun 87
4.1 Various 3-D Nanostructure Fabrication Techniques 88
4.2 3-D Nanostructure Fabrication by Focused Ion Beam 88
4.2.1 Focused-Ion-Beam-Induced Chemical Vapor Deposition 88
4.2.2 3-D Fabrication Process 89
4.2.3 Mechanical Properties 91
4.2.4 Free-Space Nanowiring 93
4.2.5 Electrical Properties 95
4.2.6 Brilliant Blue Observation from a *Morpho* Butterfly Scale Quasistructure 96
4.3 3-D Nanostructure Fabrication by Electron Beam 97
4.3.1 Electron-Beam-Induced Chemical Vapor Deposition 97
4.3.2 3-D Nanofabrication by Electron-Beam-Induced CVD 102
4.4 3-D Nanostructure Fabrication by Laser 103
4.4.1 3-D Nanostructure Fabrication by Femtosecond Laser 103
4.4.2 3-D Nanodevices: Fabrication of Photonic Crystals 105
References 109

5 **Nanoimprint Lithography**
Helmut Schift, Anders Kristensen 113
5.1 Emerging Nanopatterning Methods 115
5.1.1 Next-Generation Lithography 116
5.1.2 Nanoprint and Nanoimprint Lithography 117

5.2 Nanoimprint Process 119
5.2.1 Limits of Molding 120
5.2.2 Squeeze Flow of Thin Films 120
5.2.3 Residual Layer Thickness Homogeneity 122
5.2.4 Demolding 122
5.2.5 Curing of Resists and UV-Assisted NIL 123
5.2.6 Mix-and-Match Methods and Combination with Directed Self-Assembly 124
5.2.7 Multilayer and Multilevel Systems 124
5.2.8 Reversal NIL and Layer Transfer 125
5.3 Tools and Materials for Nanoimprint 125
5.3.1 Resist Materials 126
5.3.2 Stamps 127
5.3.3 Stamp Fabrication and Tooling 128
5.3.4 Anti-adhesive Coatings 128
5.3.5 Machines 129
5.4 Applications 132
5.4.1 Types of Nanoimprint Applications 132
5.4.2 Semiconductor Memory Chips 132
5.4.3 Bit-Patterned (Magnetic) Media (BPM) for Hard Disk Drives 133
5.4.4 Subwavelength Metal-Strip Gratings and Transparent Electrodes for Displays 133
5.4.5 High Brightness Light-Emitting Diodes 134
5.4.6 Polymer Optics 135
5.4.7 Biological Applications 136
5.5 Conclusion and Outlook 137
References 138

6 **Stamping Techniques for Micro- and Nanofabrication**
John A. Rogers, Etienne Menard 143
6.1 High-Resolution Stamps 144
6.2 Microcontact Printing 146
6.3 Nanotransfer Printing 148
6.4 Applications 152
6.4.1 Unconventional Electronic Systems 152
6.4.2 Lasers and Waveguide Structures 157
6.5 Conclusions 159
References 159

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems**
Christian A. Zorman 163
7.1 Silicon and its Commonly-Used Derivatives 163
7.1.1 Single-Crystal Silicon 164
7.1.2 Polycrystalline and Amorphous Silicon 165
7.1.3 Porous Silicon 168
7.1.4 Silicon Dioxide 168
7.1.5 Silicon Nitride 169

7.2 Germanium-Based Materials 170
7.2.1 Polycrystalline Germanium 170
7.2.2 Polycrystalline Silicon Germanium 171
7.3 Metals 171
7.4 Semiconductors for Harsh Environment Applications 173
7.4.1 Silicon Carbide 173
7.4.2 Diamond 176
7.5 GaAs, InP and Related III-V Materials 178
7.6 Piezoelectric Materials 180
7.6.1 Lead Zirconate Titanate 180
7.6.2 Aluminum Nitride 181
7.6.3 Gallium Nitride 182
7.7 Polymer Materials 183
7.7.1 Polyimide 183
7.7.2 SU-8 183
7.7.3 Parylene 184
7.7.4 Liquid Crystal Polymer 184
7.8 Future Trends 185
References 185

Part B Nanomaterial and Nanostructures

8 **Carbon Nanotubes**
Marc Monthioux, Philippe Serp, Brigitte Caussat, Emmanuel Flahaut, Manitra Razafinimanana, Flavien Valensi, Christophe Laurent, Alain Peigney, David Mesguich, Alicia Weibel, Wolfgang Bacsa, Jean-Marc Broto 193
8.1 Structure of Carbon Nanotubes 194
8.1.1 Single-Wall Nanotubes 194
8.1.2 Multiwall Nanotubes 197
8.2 Synthesis of Carbon Nanotubes 199
8.2.1 Solid Carbon Source-Based Synthesis Techniques: The DC Electric Arc 199
8.2.2 Gaseous Carbon Source-Based Synthesis Techniques 202
8.2.3 Miscellaneous Techniques 207
8.2.4 Synthesis with Controlled Orientation 208
8.3 Growth Mechanisms of Carbon Nanotubes 210
8.3.1 Catalyst-Free Growth 210
8.3.2 Catalytically Activated Growth 210
8.4 Properties of Carbon Nanotubes 213
8.4.1 Overall Properties of SWNTs 213
8.4.2 Adsorption Properties 213
8.4.3 Electronic and Optical Properties 215
8.4.4 Mechanical Properties 216
8.4.5 Reactivity 216
8.5 Carbon Nanotube-Based Nano-Objects (Carbon Meta-Nanotubes) 217
8.5.1 Heteronanotubes 217
8.5.2 Filled Carbon Nanotubes 217

8.5.3 Functionalized Nanotubes 220
8.5.4 Coated/Decorated Nanotubes 221
8.6 Carbon-Nanotube-Containing Materials (Composites) 223
8.6.1 Metal Matrix Composites 223
8.6.2 Ceramic Matrix Composites 223
8.6.3 Polymer-Matrix Composites 224
8.6.4 Composites as Multifunctional Materials 225
8.7 Current Applications of Carbon Nanotubes (on the Market) 227
8.7.1 Carbon Nanotubes and Master Batches 227
8.7.2 Near-Field Microscopy Probes 227
8.7.3 Electron Emitter 229
8.7.4 Flexible and Touch-Screen Displays 229
8.7.5 Nonvolatile Random Access Memory 229
8.7.6 Light Absorbants 229
8.7.7 Automotive and Aeronautic Industry 230
8.7.8 High-Tech Goods and Clothes 230
8.7.9 Anodes for Li-Ion Batteries 230
8.7.10 Chemical Sensors 230
8.7.11 Catalyst Support 231
8.8 Toxicity and Environmental Impact of Carbon Nanotubes 231
8.9 Concluding Remarks 233
References 233

9 **Nanowires**
Mildred S. Dresselhaus, Marcie R. Black, Vincent Meunier, Oded Rabin 249
9.1 Synthesis 250
9.1.1 Template-Assisted Synthesis 250
9.1.2 VLS Method for Nanowire Synthesis 255
9.1.3 Etching Methods 257
9.1.4 Other Synthesis Methods 258
9.1.5 Nanowire Alignment and Superstructures of Nanowires 260
9.2 Characterization and Physical Properties of Nanowires 262
9.2.1 Structural Characterization 262
9.2.2 Mechanical Properties 266
9.2.3 Transport Properties 267
9.2.4 Optical Properties 277
9.3 Applications 282
9.3.1 Electrical Applications 282
9.3.2 Optical Applications 284
9.3.3 Energy Applications 286
9.3.4 Chemical and Biochemical Sensing Devices 288
9.3.5 Magnetic Applications 289
9.4 Concluding Remarks 290
References 290

10 **Nanoribbons**
Toshiaki Enoki, Shintaro Sato 303
10.1 Graphene Nanoribbons 303
10.2 Electronic and Magnetic Properties 305
10.3 Characterizations 310

10.4 Syntheses: Top-Down Methods and Bottom-Up Methods 317
10.4.1 Graphene Nanoribbon Formation by Top-Down Approaches 317
10.4.2 Graphene Nanoribbon Formation by Bottom-Up Approaches 319
10.4.3 Graphene Nanoribbon Formation by Other Approaches 321
10.5 Potential Applications 324
10.5.1 Application to Transistors 324
10.5.2 Interconnect Application 326
10.5.3 Other Potential Applications 328
10.6 Conclusions 330
References 331

11 **Nanoparticles and Their Applications**
Seyedsina Moeinzadeh, Esmaiel Jabbari 335
11.1 Overview 335
11.2 NPs Classification and Synthesis 336
11.2.1 NPs Classification 336
11.2.2 NPs Synthesis 336
11.3 Properties of NPs 348
11.3.1 Size of NPs 348
11.3.2 Shape of NPs 349
11.3.3 Magnetic Properties of NPs 349
11.3.4 Mechanical Properties of NPs 350
11.3.5 Electrical Properties of NPs 351
11.4 Applications of NPs 351
11.5 Summary 353
References 353

12 **Graphene**
Aravind Vijayaraghavan, Maria Iliut 363
12.1 Methods of Production 365
12.1.1 Micromechanical Exfoliation of Graphite (the Scotch Tape Technique) 365
12.1.2 Chemical Vapor Deposition 366
12.1.3 Decomposition of Carbides 366
12.1.4 Exfoliation in a Solvent 367
12.1.5 Synthetic Organic Chemistry Route 367
12.1.6 Unzipping of Carbon Nanotubes 368
12.1.7 Derivatives of Graphene 368
12.2 Properties 369
12.2.1 Structure and Physical Properties 369
12.2.2 Mechanical Properties 369
12.2.3 Electronic Properties 370
12.2.4 Optical Properties 371
12.2.5 Thermal and Thermoelectric Properties 371
12.2.6 Chemical Properties 372
12.2.7 Properties of Graphene Derivatives 372
12.3 Characterization 372
12.3.1 Optical Microscopy 372
12.3.2 Transmission Electron Microscopy 372

12.3.3 Scanning Probe Techniques 373
12.3.4 Angle-Resolved Photoemission Spectroscopy (ARPES) 374
12.3.5 Raman Spectroscopy 374
12.3.6 Electrical Characterization 375
12.4 Applications 376
12.4.1 Structural and Electrical Composites 377
12.4.2 Transparent Conductive Films 377
12.4.3 Sensors 379
12.4.4 Electronic Applications 379
12.4.5 Spintronics 382
12.4.6 Photonics and Optoelectronics 382
12.4.7 Filters and Barriers 382
12.4.8 Supercapacitors 384
12.4.9 Biomedical Applications 384
12.5 Conclusions and Outlook 385
References 386

Part C MEMS/NEMS

13 **MEMS/NEMS Devices and Applications**
Philip X.-L. Feng, Darrin J. Young, Christian A. Zorman 395
13.1 MEMS Devices and Applications 397
13.1.1 Pressure Sensor 397
13.1.2 Inertial Sensor 400
13.1.3 Optical MEMS 404
13.1.4 RF MEMS 408
13.2 NEMS Devices and Applications 415
13.2.1 Fabrication Techniques for NEMS 415
13.2.2 NEMS Measurement and Signal Transduction 419
13.2.3 NEMS Devices and Emerging Applications 419
13.3 Challenges and Perspectives 425
References 426

14 **Single-Walled Carbon Nanotube Sensor Concepts**
Cosmin I. Roman, Thomas Helbling, Miroslav Haluška, Christofer Hierold . 431
14.1 Sensor Design Considerations 432
14.1.1 CNT Properties for Sensing 432
14.1.2 Carbon Nanotube FET Structures 436
14.1.3 Sensor Characterization 438
14.2 Sensor Fabrication: SWNT Synthesis and Integration 439
14.2.1 Methods for SWNT Production 440
14.2.2 Strategies for SWNT Assembly into Devices 441
14.2.3 Process Contaminations and Sensor Performance 444
14.3 Summary of State-of-the-Art, Applications Examples 445
14.3.1 Chemical and Biochemical Sensors 445
14.3.2 Piezoresistive Sensors 447
14.3.3 Resonant Sensors 449
14.4 Concluding Remarks 451
References 451

15 **Nanomechanical Cantilever Array Sensors**
Hans Peter Lang, Martin Hegner, Christoph Gerber 457
15.1 Technique 458
15.1.1 Cantilevers 458
15.1.2 History of Cantilever Sensors 458
15.2 Cantilever Array Sensors 459
15.2.1 Concept 459
15.2.2 Compressive and Tensile Stress 459
15.2.3 Disadvantages of Single Microcantilevers 459
15.2.4 Reference and Sensor Cantilevers in an Array 460
15.3 Modes of Operation 460
15.3.1 Static mode 460
15.3.2 Dynamic Mode 462
15.3.3 Heat Mode 463
15.3.4 Further Operation Modes 464
15.4 Microfabrication 464
15.5 Measurement Setup 465
15.5.1 Measurements in Gaseous or Liquid Environment 465
15.5.2 Readout Principles 466
15.5.3 Miniaturized Piezoresistive Arrays 468
15.5.4 Membrane Surface Stress Sensors 468
15.6 Functionalization Techniques 468
15.6.1 General Strategy 468
15.6.2 Functionalization Methods 469
15.6.3 Microfluidics 469
15.6.4 Array of Dimension-Matched Capillaries 469
15.6.5 Inkjet Spotting 469
15.7 Applications 470
15.7.1 Chemical Detection 470
15.7.2 Biochemical Environment 472
15.7.3 Microcantilever Sensors to Measure Physical Properties 475
15.7.4 Recent Developments 476
15.8 Conclusions and Outlook 476
References 477

16 **Microfluidic Devices and Their Applications**
Aditya Aryasomayajula, Pouriya Bayat, Pouya Rezai, P. Ravi Selvaganapathy 487
16.1 Preface 487
16.1.1 Why Miniaturize? 488
16.1.2 Advantages of Microfluidics and Nanofluidics 488
16.2 Historical Developments 489
16.3 Dominant Phenomenon and Micro–Nanofluidic Design 491
16.3.1 Laminar Flow 492
16.3.2 Diffusion in Microfluidics 492
16.3.3 Rapid Heating–Cooling 492
16.3.4 Surface Tension 493
16.3.5 Electrokinetics 494
16.3.6 Geometric Control of Two Phase Flow 495

16.3.7 Nanoscale Phenomena 495
16.4 Fabrication of Micro–Nanofluidic Devices 496
16.4.1 Silicon and Glass 496
16.4.2 Polymers 497
16.4.3 Paper 499
16.4.4 Threads 501
16.4.5 Hydrogels 502
16.5 Applications 504
16.5.1 Commercial Applications and Their Working Mechanism 504
16.5.2 Research Applications 519
16.6 Outlook and Future Directions 521
16.6.1 Additive Manufacturing 521
16.6.2 Artificial Organs 521
16.6.3 Biopharmaceutical Production 521
16.6.4 Droplets in Food and Cosmetics 522
16.6.5 Continuous Health Monitoring 522
References 522

17 **Microfluidic Micro/Nano Droplets**
Gopakumar Kamalakshakurup, Derek Vallejo, Abraham Lee 537
17.1 Introduction to Micro/Nano Droplet Microfluidic Technologies 537
17.1.1 Droplet Generation 538
17.1.2 Droplet Merging 541
17.2 Overview of Current Trends in Droplet Microfluidic Technologies 546
17.3 Fundamental Designs and Techniques for Microfluidic Generation of Droplets 547
17.3.1 Passive Droplet Generation Techniques 548
17.4 Microfluidic Micro-/Nanodroplet Applications 550
17.4.1 Chemical Reactions in Droplets 550
17.4.2 Biomolecule Synthesis 552
17.4.3 Drug Discovery 553
17.5 Conclusion 556
References 556

18 **Nanorobotics**
Bradley J. Nelson, Lixin Dong 559
18.1 Overview of Nanorobotics 560
18.2 Actuation at Nanoscales 561
18.2.1 Electrostatics 562
18.2.2 Electromagnetics 562
18.2.3 Piezoelectrics 562
18.2.4 Other Techniques 563
18.3 Nanorobotic Manipulation Systems 563
18.3.1 Overview 563
18.3.2 Nanorobotic Manipulation Systems 565
18.4 Nanorobotic Assembly 568
18.4.1 Overview 568
18.4.2 Carbon Nanotubes 569
18.4.3 Nanocoils 574
18.5 Applications 576

18.5.1 Robotic Biomanipulation 576
18.5.2 Nanorobotic Devices 578
References 580

Part D BioMEMS/NEMS

19 **Applications of MEMS to Cell Biology**
Georg E. Fantner, Pascal D. Odermatt, Haig Alexander Eskandarian 587
19.1 Biological Background 588
19.2 High Resolution Microscopy Methods for Live Cell Imaging 592
19.3 Measuring Mechanical Properties on Living Cells 593
19.3.1 Using MEMS Cantilevers to Measure Cell Mechanical Properties 593
19.3.2 Cantilever Calibration 594
19.3.3 Instrument Calibration 594
19.3.4 Extracting Materials Properties from Cell-Indentation Measurements 595
19.4 Cancer Detection Using Nanomechanical Sensors 596
19.5 Measurement of Cell Adhesion Using Cantilever Based Force Sensor 599
19.5.1 Tip Functionalization Methods 599
19.5.2 Bacterial Adhesion 600
19.5.3 Cell–Substrate Interaction (Biomaterials/Tissue Engineering) 601
19.5.4 Cell–Cell Interaction 601
19.6 Manipulation of Cells Using MEMS Cantilevers 602
19.7 Mass Measurements of Single Cells Using Resonant Cantilevers 603
19.7.1 Resonant Mechanical Sensors 603
19.7.2 Cellular Growth 604
19.8 Characterization of Antibiotic Action Using MEMS Devices 606
19.8.1 Characterizing the Impact of Antibiotics 606
19.8.2 Detecting Fluctuations in the Buoyant Mass of Single Bacteria 606
19.8.3 Detecting the Dynamic Motion of Bacteria in Fluid 606
19.8.4 Antibiotic Effect on the Bacterial Cell Surface 607
19.9 Conclusions and Outlook 609
References 609

20 **Contact-free Mechanical Manipulation of Biological Materials**
Jörg Schnauß, Josef A. Käs, David M. Smith 617
20.1 Interaction of Laser Light with Biological Material 618
20.1.1 Conservation of Momentum – Gradient and Scattering Force 618
20.1.2 Quasiparallel Laser Light – Dual-Beam Laser Trap 619
20.1.3 Focused Laser Light – Optical Tweezers 620
20.2 Optical and Magnetic Tweezers for Single-Molecule Manipulations 621
20.2.1 Optical Tweezers 622
20.2.2 Magnetic Tweezers 623

20.2.3 Mechanical Measurements of Biomolecular Ensembles ... 624
20.3 Optical and Electric Forces for the Manipulation of Whole Cells ... 629
20.3.1 The Optical Stretcher ... 629
20.3.2 Cell Deformation via Optical Tweezers ... 634
20.4 Hydrodynamic Shear Forces for the Manipulation of Single Cells ... 636
20.4.1 Flow-Induced Shear Deformation of Cells ... 636
20.4.2 High-Throughput Mechanical Phenotyping of Cell Populations in Real Time ... 637
20.5 Conclusion and Outlook ... 638
References ... 638

21 **Nano-Particles for Biomedical Applications**
Paolo Decuzzi, Alessandro Coclite, Aeju Lee, Anna Lisa Palange, Daniele Di Mascolo, Ciro Chiappini, Hélder A. Santos, Maria Laura Coluccio, Gerardo Perozziello, Patrizio Candeloro, Enzo Di Fabrizio, Francesco Gentile ... 643
21.1 Overview ... 644
21.2 Rational Design of Nanoconstructs ... 646
21.2.1 A Hierarchical Multiscale Approach for Vascular Transport of Nanoconstructs ... 647
21.3 Multifunctional Polymeric Nanoconstructs ... 650
21.3.1 Spherical Polymeric Nanoconstructs (SPNs) for Combination Therapy and Biomedical Imaging ... 650
21.3.2 Discoidal Polymeric Nanoconstructs (DPNs) for Maximizing Accumulation Within Tumor Vasculature ... 650
21.4 Sensing and Drug Delivery with Porous Silicon Nanomaterials ... 655
21.4.1 Fabrication ... 656
21.4.2 Surface Chemistry ... 657
21.4.3 Biointerface ... 658
21.4.4 Drug Delivery ... 659
21.4.5 Sensing ... 661
21.5 Nanomedicine Synthesis by Microfluidics Technology ... 662
21.5.1 Why Microfluidics for the Synthesis of Nanomedicines? ... 662
21.5.2 Microfluidic Production of Nanocarriers Using PDMS Microfluidic Devices ... 664
21.5.3 Microfluidic Production of Nanocarriers Using Glass Capillary Microfluidic Devices ... 665
21.5.4 Summary ... 668
21.6 Electroless Formation of Metal Nanoparticle Aggregates ... 669
21.6.1 DLA Methods Can Predict Electroless Formation of Metal Nanoparticles ... 673
21.7 2-D and 3-D Optical Nanostructures ... 677
21.7.1 Integration of Optical Nanostructures in Microfluidic Devices ... 677
21.7.2 Three-Dimensional Optical Nanostructure ... 679
21.7.3 Two-Dimensional Optical Nanostructure ... 679
References ... 681

22 **Biological Molecules in Therapeutic Nanodevices**
Stephen C. Lee, Bharat Bhushan ... 693

22.1 Definitions and Scope 694
22.1.1 Design Issues 695
22.1.2 Identification of Biomolecular Components 695
22.1.3 Design Paradigms 697
22.1.4 Utility and Scope of Therapeutic Nanodevices 698
22.2 Assembly Approaches 700
22.2.1 Low-Throughput Construction Methods 701
22.2.2 Supramolecular Chemistry and Self-Assembly 702
22.2.3 Chemoselective Conjugation 708
22.2.4 Unnatural Amino Acids to Support Chemoselective Conjugation of Biologically Produced Proteins 709
22.3 Sensing Devices 710
22.3.1 Planar FET Protein Sensors 710
22.3.2 Biotechnology Approaches to the *Fundamental Limitations* of Planar ImmunoFETs 712
22.3.3 Nanotribology of Protein-Sensing Interfaces on Micromachined Surfaces 713
22.4 Concluding Remarks: Barriers to Practice 717
22.4.1 You Do Not Know What You Do Not Know: The Consequences of Certainty 717
22.4.2 Are Proteins and Molecules Legitimately Part of Nanotechnology? 718
References 718

Part E Nanometrology

23 **Scanning Probe Microscopy – Principle of Operation, Instrumentation and Probes**
Bharat Bhushan, Othmar Marti 725
23.1 Scanning Tunneling Microscope 727
23.1.1 The STM Design of Binnig et al. 727
23.1.2 Commercial STMs 728
23.1.3 STM Probe Construction 730
23.2 Atomic Force Microscope 731
23.2.1 The AFM Design of Binnig et al. 733
23.2.2 Commercial AFM 733
23.2.3 AFM Probe Construction 738
23.2.4 Friction Measurement Methods 742
23.2.5 Normal Force and Friction Force Calibrations of Cantilever Beams 745
23.3 AFM Instrumentation and Analyses 747
23.3.1 The Mechanics of Cantilevers 747
23.3.2 Instrumentation and Analyses of Detection Systems for Cantilever Deflections 749
23.3.3 Combinations for 3-D-Force Measurements 757
23.3.4 Scanning and Control Systems 758
23.4 Conclusion 763
References 763

24 **Low-Temperature Scanning Probe Microscopy**
Mehmet Z. Baykara, Markus Morgenstern, Alexander Schwarz, Udo D. Schwarz 769
24.1 Microscope Operation at Low Temperatures 771
24.1.1 Drift 771
24.1.2 Noise 771
24.1.3 Stability 771
24.1.4 Piezo Relaxation and Hysteresis 771
24.2 Instrumentation 772
24.3 Scanning Tunneling Microscopy and Spectroscopy 773
24.3.1 Atomic Manipulation 774
24.3.2 High-Resolution Spectroscopy 775
24.3.3 Imaging Electronic Wave Functions 779
24.3.4 Imaging Spin Polarization: Nanomagnetism 786
24.4 Scanning Force Microscopy and Spectroscopy 788
24.4.1 Atomic-Scale and Intramolecular Imaging 789
24.4.2 Force Spectroscopy 790
24.4.3 Atomic and Molecular Manipulation 793
24.4.4 Kelvin Probe Force Microscopy 794
24.4.5 Magnetic Force Microscopy 795
24.4.6 Magnetic Exchange Force Microscopy and Spectroscopy 797
24.4.7 Magnetic Resonance Force Microscopy 798
24.5 Summary 799
References 799

25 **Biomedical Sensing with the Atomic Force Microscope**
Constanze Lamprecht, Jürgen Strasser, Melanie Koehler, Sandra Posch, Yoojin Oh, Rong Zhu, Lilia A. Chtcheglova, Andreas Ebner, Peter Hinterdorfer 809
25.1 Topographical Imaging of Biological Samples 810
25.1.1 Basic Imaging Modes for Biological Application 811
25.1.2 Sample Preparation and Imaging Conditions 812
25.1.3 Imaging Artifacts 812
25.2 Single-Molecule Force Spectroscopy (SMFS) 814
25.2.1 Basic Principle of SMFS 814
25.2.2 Theory of Single-Molecule Binding and Force Spectroscopy 816
25.2.3 Data Evaluation and Determination of Interaction Rate Constants 818
25.2.4 AFM Tip Chemistry 819
25.3 Simultaneous Topography and Recognition Imaging (TREC) 821
25.3.1 Principle of TREC 821
25.3.2 Parameter Optimization in TREC 821
25.3.3 Application of TREC 822
25.4 AFM Biomedical Sensing – Examples 827
25.4.1 Isolated Membranes 828
25.4.2 Mammalian Cells 830
25.4.3 Bacteria 832
25.5 Perspectives and Concluding Remarks 836
References 837

26 **Superresolution Microscopy**
Tom D. Milster 845
26.1 Overview 845
26.1.1 Review of Classical Resolution 845
26.1.2 Point Spread Function (PSF) and Frequency Response 847
26.2 Scanning Aperture Techniques 849
26.3 4-Pi Microscopy 853
26.4 Enhancement/Depletion Techniques 854
26.5 Photoactivated Localization 856
26.6 Structured Illumination 858
26.7 Light-Sheet Microscopy 862
26.8 Comparison of Techniques for Live-Cell Imaging 863
26.9 Summary 864
References 864

Part F Bio/Nanotribology and Bio/Nanomechanics

27 **Nanotribology, Nanomechanics and Materials Characterization**
Bharat Bhushan 869
27.1 Description of AFM/FFM and Various Measurement Techniques 871
27.1.1 Surface Roughness and Friction Force Measurements 872
27.1.2 Adhesion Measurements 875
27.1.3 Scratching, Wear, and Fabrication/Machining 875
27.1.4 Surface Potential Measurements 876
27.1.5 In situ Characterization of Local Deformation Studies 876
27.1.6 Nanoindentation Measurements 877
27.1.7 Localized Surface Elasticity and Viscoelasticity Mapping 878
27.1.8 Boundary Lubrication Measurements 881
27.1.9 AFM Tip Wear 881
27.2 Surface Imaging, Friction, and Adhesion 882
27.2.1 Atomic-Scale Imaging and Friction 882
27.2.2 Microscale Friction 884
27.2.3 Directionality Effect on Microfriction 887
27.2.4 Surface Roughness-Independent Microscale Friction 891
27.2.5 Velocity Dependence on Micro/Nanoscale Friction 893
27.2.6 Nanoscale Friction and Wear Mapping 898
27.2.7 Adhesion and Friction in Wet Environments 899
27.2.8 Separation Distance Dependence of Meniscus and van der Waals Forces 901
27.2.9 Scale Dependence in Friction 902
27.3 Micro/Nanoscale Wear and Scratching, Local Deformation, and Nanofabrication/Nanomachining 906
27.3.1 Nanoscale Wear 906
27.3.2 Microscale Scratching 906
27.3.3 Microscale Wear 907
27.3.4 In situ Characterization of Local Deformation 911
27.3.5 Nanofabrication/Nanomachining 914
27.4 Indentation 915

27.4.1 Picoindentation 915
27.4.2 Nanoindentation 915
27.4.3 Localized Surface Elasticity and Viscoelasticity Mapping .. 917
27.5 Boundary Lubrication 919
27.5.1 Perfluoropolyether Lubricants 919
27.5.2 Self-Assembled Monolayers 924
27.5.3 Liquid Film Thickness Measurements 926
27.6 Conclusion 928
References 929

28 **Surface Forces and Nanorheology of Molecularly Thin Films**
Dong Woog Lee, Marina Ruths, Jacob N. Israelachvili 935
28.1 Types of Surface Forces 936
28.2 Methods Used to Study Surface Forces 938
28.2.1 Force Laws 938
28.2.2 Adhesion Forces 938
28.2.3 The SFA 940
28.3 Normal Forces Between Dry (Unlubricated) Surfaces 941
28.3.1 Van Der Waals Forces in Vacuum and Inert Vapors 941
28.4 Normal Forces Between Surfaces in Liquids 942
28.4.1 Van Der Waals Forces in Liquids 942
28.4.2 Electrostatic Forces 943
28.4.3 Solvation and Structural Forces 945
28.4.4 Hydration and Hydrophobic Forces 947
28.4.5 Polymer-Mediated Forces 949
28.5 Adhesion and Capillary Forces 951
28.5.1 Capillary Forces 951
28.5.2 Adhesion Mechanics 952
28.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch 952
28.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis 954
28.6 Introduction: Different Modes of Friction 956
28.7 Relationship Between Adhesion and Friction 957
28.7.1 Amontons' Law 957
28.7.2 Adhesion Force and Load Contribution to Interfacial Friction 958
28.7.3 Examples of Experimentally Observed Friction of Dry Surfaces 962
28.7.4 Transition from Interfacial to Normal Friction with Wear . 965
28.8 Liquid Lubricated Surfaces 966
28.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime 966
28.8.2 Friction of Intermediate Thickness Films 967
28.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology 969
28.9 Effects of Nanoscale Texture on Friction 976
28.9.1 Role of the Shape of Confined Molecules 976
28.9.2 Effects of Surface Structure 978
References 978

29 **Atomic Scale Friction Phenomena**
Enrico Gnecco, Rémy Pawlak, Marcin Kisiel, Thilo Glatzel, Ernst Meyer 987
29.1 Friction Force Microscopy in Selected Environments 988
29.1.1 Force Calibration 988
29.1.2 Friction Force Microscopy in Ultra High Vacuum 990
29.1.3 Energy Dissipation in Noncontact Atomic Force Microscopy 992
29.1.4 Friction Force Microscopy in Water 993
29.2 The Prandtl–Tomlinson Model 993
29.2.1 One-Dimensional Prandtl–Tomlinson Model 993
29.2.2 Two-Dimensional Prandtl–Tomlinson Model 994
29.2.3 Friction Between Atomically Flat Surfaces 995
29.3 Friction Experiments on the Atomic Scale 996
29.4 Thermal Effects on Atomic Friction 998
29.4.1 The Prandtl–Tomlinson Model at Finite Temperature 998
29.4.2 Velocity and Temperature Dependence of Atomic-Scale Friction 999
29.5 Friction on the Nanometer Scale 999
29.5.1 Continuum Mechanics of Single Asperities 1000
29.5.2 Dependence of Friction on Load 1000
29.5.3 Estimation of the Contact Area 1001
29.6 Wear on the Atomic Scale 1002
29.7 Noncontact Friction 1003
29.8 Single Molecule Friction 1005
29.9 Conclusion 1008
References 1008

30 **Computer Simulations of Nanometer-Scale Indentation and Friction**
Susan B. Sinnott, Seong-Jun Heo, Donald W. Brenner, Judith A. Harrison, Douglas L. Irving 1013
30.1 Computational Details 1014
30.1.1 Energies and Forces 1014
30.1.2 Important Approximations 1016
30.2 Indentation 1019
30.2.1 Surfaces 1019
30.2.2 Thin Films 1028
30.3 Friction and Lubrication 1033
30.3.1 Bare Surfaces 1033
30.3.2 Decorated Surfaces 1039
30.3.3 Thin Films 1041
30.4 Conclusions 1058
References 1058

31 **Cellular Nanomechanics**
Roger D. Kamm, Jan Lammerding, Mohammad R. K. Mofrad 1069
31.1 Overview 1069
31.1.1 The Importance of Cell Mechanics in Biology and Medicine 1069
31.1.2 Examples Drawn from Biology and Pathophysiology 1070

31.2 Structural Components of a Cell 1071
31.2.1 Membranes 1071
31.2.2 Cytoskeleton 1072
31.2.3 Nucleus 1075
31.2.4 Cell Contractility and Motor Proteins 1076
31.2.5 Adhesion Complexes 1077
31.3 Experimental Methods 1077
31.3.1 Methods of Force Application 1077
31.3.2 Rheological Properties 1081
31.3.3 Active Force Generation 1081
31.3.4 Biological Responses 1082
31.3.5 Nonlinear Effects 1082
31.3.6 Homogeneity and Anisotropy 1082
31.4 Theoretical and Computational Descriptions 1082
31.4.1 Continuum Models 1083
31.4.2 Biopolymer Models 1084
31.4.3 Cellular Solids 1085
31.4.4 Tensegrity 1086
31.5 Mechanics of Subcellular Structures 1086
31.5.1 Cell–Cell and Cell–Matrix Adhesions 1086
31.5.2 Cell Membranes 1088
31.5.3 Cell Nuclei 1090
31.5.4 Mechanosensing Proteins 1091
31.6 Current Understanding and Future Needs 1095
References 1096

32 **Nanomechanical Properties of Nanostructures and Scale Effects**
Bharat Bhushan 1101
32.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures 1103
32.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters 1103
32.1.2 Bending Tests of Nanostructures Using an AFM 1104
32.1.3 Bending Tests of Micro/Nanostructures Using a Nanoindenter 1108
32.2 Experimental Results and Discussion 1109
32.2.1 Indentation and Scratch Tests of Various Ceramic and Metals Using Micro/Nanoindenter 1109
32.2.2 Bending Tests of Ceramic Nanobeams Using an AFM 1113
32.2.3 Bending Tests of Metallic Microbeams Using a Nanoindenter 1116
32.2.4 Indentation and Scratch Tests of Polymeric Microbeams Using a Nanoindenter 1117
32.2.5 Bending Tests of Polymeric Microbeams Using a Nanoindenter 1121
32.3 Finite Element Analysis of Nanostructures with Roughness and Scratches 1125
32.3.1 Stress Distribution in a Smooth Nanobeam 1126

32.3.2 Effect of Roughness in the Longitudinal Direction 1126
32.3.3 Effect of Roughness in the Transverse Direction and Scratches 1130
32.3.4 Effect on Stresses and Displacements for Materials Which are Elastic, Elastic–Plastic or Elastic–Perfectly Plastic 1131
32.4 Summary 1131
32.A Appendix: Fabrication Procedure for the Double-Anchored and Cantilever Beams 1132
References 1134

Part G Molecularly-Thick Films for Lubrication

33 **Nanotribology of Ultrathin and Hard Amorphous Carbon Films**
Bharat Bhushan 1141
33.1 Description of Commonly Used Deposition Techniques 1145
33.1.1 Filtered Cathodic Arc Deposition Technique 1145
33.1.2 Ion-Beam Deposition Technique 1146
33.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique 1148
33.1.4 Sputtering Deposition Technique 1148
33.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique 1148
33.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties 1149
33.2.1 EELS and Raman Spectroscopy 1149
33.2.2 Hydrogen Concentrations 1150
33.2.3 Physical Properties 1153
33.2.4 Summary 1154
33.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques 1154
33.3.1 Micromechanical Characterization 1154
33.3.2 Microscratch and Microwear Studies 1161
33.3.3 Macroscale Tribological Characterization 1168
33.3.4 Coating Continuity Analysis 1173
33.4 Conclusion 1174
References 1175

34 **Self-Assembled Monolayers for Nanotribology and Surface Protection**
Bharat Bhushan 1179
34.1 Background 1179
34.1.1 Need for Hydrophobic Surfaces for Nanotribology 1180
34.1.2 Surface Films for Nanotribology and Surface Protection .. 1180
34.1.3 Scope of the Chapter 1182
34.2 A Primer on Organic Chemistry 1182
34.2.1 Electronegativity/Polarity 1183
34.2.2 Classification and Structure of Organic Compounds 1183
34.2.3 Polar and Nonpolar Groups 1186

34.3 Self-Assembled Monolayers: Substrates, Spacer Chains, and End Groups in the Molecular Chains 1186
34.4 Contact Angle and Nanotribological Properties of SAMs 1189
34.4.1 Measurement Techniques 1191
34.4.2 Hexadecane Thiol and Biphenyl Thiol SAMs on Au(111) .. 1192
34.4.3 Perfluoroalkylsilane and Alkylsilane SAMs on Si(100), and Perfluoroalkylphosphonate and Alkylphosphonate SAMS on Al 1198
34.4.4 Chemical Degradation and Environmental Studies 1205
34.5 Conclusion 1209
References 1210

35 **Nanoscale Boundary Lubrication Studies**
Bharat Bhushan 1215
35.1 Nanodeformation, Molecular Conformation, Spreading, Nanotribological and Electrical Studies, and Environmental Effects of Commonly Used PFPE Lubricant Films 1216
35.1.1 Nanodeformation, Molecular Conformation, and Spreading 1217
35.1.2 Nanotribological Studies and Environmental Effects 1219
35.2 Nanotribological, Electrical, and Chemical Degradations Studies and Environmental Effects in Novel PFPE Lubricant Films 1233
35.2.1 Nanotribological Studies 1234
35.2.2 Wear Detection by Surface Potential Measurements 1234
35.2.3 Wear Detection by Electrical Resistance Measurements of Z-TETRAOL and the Effect of Cycling 1237
35.2.4 Chemical Degradation and Environmental Studies 1239
35.3 Nanotribological and Electrical Studies of Ionic Liquid Films 1242
35.3.1 Monocationic Liquid Films 1244
35.3.2 Dicationic Ionic Liquid Films 1249
35.4 Conclusion 1257
References 1258

Part H Biomimetics and Bioinspired Surfaces

36 **Plant Surfaces: Structures and Functions for Biomimetic Applications**
Wilhelm Barthlott, Matthias Mail, Bharat Bhushan, Kerstin Koch 1265
36.1 500 Million Years of Evolution for Innovative Technologies 1266
36.2 Chemistry of Plant Surfaces 1270
36.2.1 Chemical Composition of Wax 1271
36.2.2 Chemical Heterogeneities 1274
36.2.3 Crystallinity 1275
36.3 Structuring of Plant Surfaces: Hierarchical Architecture Between Nano- and Macrostructures 1279
36.3.1 The Cuticle 1279
36.3.2 Hierarchical Sculpturing 1279
36.3.3 First Sculptural Level 1280

36.3.4 Second Sculptural Level 1281
36.3.5 Third Sculptural Level 1283
36.3.6 Fourth Sculptural Level 1284
36.4 Physical Basis of Surface Wetting 1285
36.4.1 Static Wetting Processes 1285
36.4.2 Dynamic Interactions 1286
36.5 Superhydrophilic and Superhydrophobic Plant Surfaces 1286
36.6 Functional Diversity of Plant Surfaces 1289
36.6.1 Mechanical Properties 1289
36.6.2 Attachment 1289
36.6.3 Reflection, Absorption, and Transmission of Spectral Radiation 1289
36.6.4 Reduction of Water Loss 1291
36.6.5 Superhydrophilicity 1291
36.6.6 Superhydrophobicity 1291
36.6.7 Anti-adhesive *Slippery* Surfaces and Aquaplaning 1293
36.7 Biomimetic Application 1294
36.7.1 Mechanical Properties 1294
36.7.2 Attachment 1294
36.7.3 Reflection, Transmission, and Absorption of Spectral Radiation 1294
36.7.4 Reduction of Water Loss 1294
36.7.5 Superhydrophilicity 1294
36.7.6 Superhydrophobicity 1295
36.7.7 Other Applications 1296
36.8 Living Prototypes: Evolution of Plant Surfaces and Biodiversity 1297
36.8.1 Water Plants 1297
36.8.2 Land Plants 1297
36.9 Conclusions 1298
References 1298

37 **Bioinspired Nanostructured Antibiofouling and Anti-inorganic Surfaces**
Bharat Bhushan 1307
37.1 Fields Susceptible to Fouling 1307
37.2 Biofouling and Inorganic Fouling Formation Mechanisms 1310
37.2.1 Biofouling Formation 1310
37.2.2 Inorganic Fouling Formation 1310
37.2.3 Surface Factors 1311
37.3 Antifouling Strategies from Living Nature 1313
37.4 Antifouling: Current Prevention and Cleaning Techniques 1315
37.4.1 Prevention Techniques 1315
37.4.2 Self-Cleaning Surfaces and Cleaning Techniques 1318
37.5 Bioinspired Rice Leaf Surfaces for Antifouling 1318
37.5.1 Fabrication of Micropatterned Samples 1318
37.5.2 Antibiofouling Measurements 1319
37.5.3 Anti-inorganic Fouling Measurements 1320
37.5.4 Results and Discussion 1320
37.5.5 Antibiofouling and Anti-inorganic Fouling Mechanisms 1323
37.6 Closure 1324
References 1324

Part I Micro/Nanodevice Reliability

38 **MEMS/NEMS and BioMEMS/BioNEMS: Tribology, Mechanics, Materials and Devices**
Bharat Bhushan 1331
38.1 MEMS/NEMS Basics 1332
38.1.1 Introduction to MEMS 1334
38.1.2 Introduction to NEMS 1335
38.1.3 Introduction to BioMEMS/BioNEMS 1336
38.2 Nanotribology and Nanomechanics Issues in MEMS/NEMS and BioMEMS/BioNEMS 1336
38.2.1 MEMS 1337
38.2.2 NEMS 1342
38.2.3 BioMEMS 1344
38.2.4 BioNEMS 1348
38.2.5 Microfabrication Processes 1351
38.2.6 Tribological Needs 1351
38.3 Nanotribology and Nanomechanics Studies of Silicon and Related Materials 1352
38.3.1 Virgin and Treated/Coated Silicon Samples 1352
38.3.2 Nanotribological and Nanomechanical Properties of Polysilicon Films and SiC Films 1357
38.4 Lubrication Studies for MEMS/NEMS 1359
38.4.1 Perfluoropolyether Lubricants 1359
38.4.2 Self-Assembled Monolayers (SAMs) 1361
38.4.3 Hard Diamond-Like Carbon (DLC) Coatings 1365
38.5 Nanoscale Friction, Wear, and Mechanical Behavior of Nano-Objects 1365
38.5.1 Single Nano-Object Friction 1368
38.5.2 Multiple Nano-Object Contact 1369
38.5.3 Summary 1375
38.6 Nanotribological Studies of Biological Molecules on Polystyrene and Silicon Surfaces and Coated Polymer Surfaces .. 1376
38.6.1 Nanoscale Adhesion, Friction, and Wear of Protein Layers on Polystyrene Surfaces 1376
38.6.2 Adhesion, Friction, and Wear of Biomolecules on Si-Based Surfaces 1382
38.6.3 Adhesion of Coated Polymer Surfaces 1389
38.7 Trajectory of Submicron Particles for Therapeutics and Diagnostics . 1390
38.8 Component-Level Studies 1392
38.8.1 Surface Roughness Studies of Micromotor Components ... 1392
38.8.2 Adhesion Measurements of Microstructures 1393
38.8.3 Microtriboapparatus for Adhesion, Friction, and Wear of Microcomponents 1394
38.8.4 Static Friction Force (Stiction) Measurements in MEMS 1397
38.8.5 Mechanisms Associated with Observed Stiction Phenomena in Digital Micromirror Devices (DMDs) and Nanomechanical Characterization 1400
38.9 Conclusion 1403
38.A Appendix: Micro/Nanofabrication Techniques 1404
References 1407

39 **Friction and Wear in Micro- and Nanomachines**
Maarten P. de Boer, Sameer S. Shroff, Frank W. DelRio, W. Robert Ashurst . 1417
39.1 From Single- to Multiple-Asperity Friction 1419
39.1.1 Micromachined Test Structures 1420
39.1.2 Monolayer Lubrication in Nano- and MEMS Tribology 1422
39.2 Rate-State Friction 1422
39.2.1 MEMS Rate-State Friction Test Platform with Decades of Range-in-Control Parameters 1423
39.2.2 Rate-State Friction Measurements 1425
39.3 Putting MEMS Friction to Use: Making Motors and Friction Instruments 1429
39.3.1 Microengines, Pumps and Generators 1429
39.3.2 In-situ TEM for Friction Investigations 1429
39.4 Wear and Tribopolymer Evolution in Micro- and Nanoswitches 1430
39.5 Concluding Remarks 1432
References 1432

40 **Failure Mechanisms in MEMS/NEMS Devices**
W. Merlijn van Spengen, Robert Modliński, Robert Puers, Anne Jourdain .. 1437
40.1 Failure Modes and Failure Mechanisms 1438
40.2 Stiction and Charge-Related Failure Mechanisms 1439
40.2.1 Stiction Due to Surface Forces 1439
40.2.2 Stiction Due to Electrostatic Attraction 1442
40.3 Creep, Fatigue, Wear, and Packaging-Related Failures 1445
40.3.1 Creep 1445
40.3.2 Fatigue 1448
40.3.3 Wear 1450
40.3.4 Packaging 1451
40.4 Conclusions 1454
References 1454

41 **Mechanical Properties of Micromachined Structures**
Harold Kahn 1459
41.1 Measuring Mechanical Properties of Films on Substrates 1459
41.1.1 Residual Stress Measurements 1459
41.1.2 Mechanical Measurements Using Nanoindentation 1460
41.2 Micromachined Structures for Measuring Mechanical Properties 1460
41.2.1 Passive Structures 1461
41.2.2 Active Structures 1464
41.3 Measurements of Mechanical Properties 1471
41.3.1 Mechanical Properties of Silicon and Polysilicon 1471
41.3.2 Mechanical Properties of Other Materials 1474
41.4 Summary 1475
References 1476

42 **High Volume Manufacturing and Field Stability of MEMS Products**
T. Kieran Nunan, Mark G. da Silva 1481
42.1 High-Volume Manufacturing Strategy 1482
42.1.1 Foundry versus Internal Fab 1482
42.1.2 The ADI Experience 1482

42.1.3 Foundry Manufacturing 1486
42.1.4 High-Volume Inertial MEMS Manufacturing Worldwide ... 1487
42.2 Robust Design for Volume Manufacturing 1491
42.2.1 MEMS Design for Manufacturability (DfM) 1491
42.3 Stable Field Performance 1496
42.3.1 Electric Field Stability 1496
42.3.2 Mechanical Field Stability 1497
42.4 Internet of Things (IoT) 1498
42.4.1 Heterogeneous Integration 1499
42.5 Conclusions and Outlook 1503
References 1503

43 **Packaging and Reliability Issues in Micro/Nano Systems**
Yu-Chuan Su, Jongbaeg Kim, Yu-Ting Cheng, Mu Chiao, Liwei Lin 1505
43.1 Introduction to MEMS Packaging 1505
43.1.1 MEMS Packaging Fundamentals 1506
43.1.2 Contemporary MEMS Packaging Approaches 1508
43.1.3 Bonding Processes for MEMS Packaging Applications 1508
43.2 Hermetic and Vacuum Packaging 1511
43.2.1 Integrated Micromachining Processes 1511
43.2.2 Postpackaging Processes 1513
43.2.3 Localized Heating and Bonding Processes 1515
43.3 Emerging Packaging Approaches 1518
43.3.1 Wafer-Level Packaging 1518
43.3.2 3-D Packaging 1521
43.3.3 Polymer-MEMS Packaging 1524
43.4 Thermal Issues and Packaging Reliability 1526
43.4.1 Thermal Issues in Packaging 1526
43.4.2 Packaging Reliability 1528
43.4.3 Long-Term and Accelerated MEMS Packaging Tests 1530
43.5 Future Trends and Summary 1533
References 1534

Part J Nanotechnology and Society and Education

44 **Nanotechnologies in Societal Context**
Barbara Herr Harthorn 1543
44.1 Assessing Technological Progress in Societal Terms 1545
44.1.1 Approaches to nanoELSI/nanoELSA 1545
44.1.2 Responsible Development and Responsible Innovation .. 1547
44.2 Nanotechnologies and Upstream Societal Engagement 1547
44.3 Ethics 1549
44.4 Governance, Law and Regulation 1550
44.4.1 Standards 1551
44.4.2 EHS/OHS 1551
44.4.3 Voluntary Approaches/Soft Law 1551
44.4.4 Responsible Development/Responsible Innovation 1552
44.5 Public Perceptions and Participation in Decision-Making 1552

44.6 Integrating the Societal with the Technical 1553
44.7 Concluding Remarks 1554
References 1554

45 **Environment, Health and Safety Issues in Nanotechnology**
Rui Chen, Chunying Chen 1559
45.1 Impacts of the Development of Nanotechnology on Environment, Health, and Safety Issues 1559
45.2 Current Progress of the Most Important Nanomaterials 1562
45.2.1 Carbon Nanotubes 1562
45.2.2 Silver Nanoparticles 1568
45.2.3 Zinc Oxide Nanoparticles 1568
45.3 Physicochemical Characteristics of Nanoparticles that Determine the Toxicity Impacts 1569
45.3.1 Particle Size and Surface Area 1570
45.3.2 Nanostructure and Shape 1571
45.3.3 Agglomeration and Aggregation 1572
45.3.4 Chemical Composition, Purity, and Impurities 1573
45.3.5 Coating, Surface Modification, and Surface Charges 1573
45.4 Novel Techniques and Biomarker Development in Nanotoxicology 1574
45.4.1 The Development of Biomarkers for Evaluation of EHS Impacts 1574
45.4.2 Novel Techniques Used for ADME Study in Nanotoxicology 1576
45.5 Conclusion and Perspectives 1580
References 1580

46 **Nanoscience and Nanotechnology Convergence**
William S. Bainbridge 1587
46.1 Background 1587
46.1.1 The Convergence Social Movement 1587
46.1.2 The Convergence Conferences 1588
46.1.3 Parallel Activities in Europe 1591
46.2 Convergence with Society 1591
46.2.1 Convergence Platforms 1592
46.2.2 Handbook Overview of Convergence 1594
46.3 Concepts and Methods 1595
46.3.1 Five General Principles 1596
46.3.2 The Information Technology Infrastructure 1597
46.4 Convergence in Education 1598
46.4.1 Innovations in Formal Education 1598
46.4.2 Integration of Learning with Life 1599
46.5 Conclusions 1600
References 1600

47 **Global Perspectives of Nanotechnology Education**
Kurt Winkelmann, Bharat Bhushan 1603
47.1 Growth and Trends of Nanotechnology Education 1605
47.1.1 History of Nanotechnology Education 1605
47.1.2 Essential Nanotechnology Topics 1607

47.2 Primary and Secondary Education 1608
47.2.1 Nanotechnology Field Trip for Primary School Children in Taiwan 1608
47.2.2 Secondary School Nanotechnology Education in Italy 1608
47.2.3 Nanotechnology School Programs in Germany 1609
47.2.4 Evaluation of Secondary School Nanotechnology Education in Europe 1609
47.3 Vocational Education Training 1610
47.3.1 Potential Areas of Growth for Vocational Education Training 1610
47.3.2 Vocational Education Training in North America 1611
47.4 Undergraduate Education 1611
47.4.1 Different Models of Undergraduate Education in Australia 1611
47.4.2 Undergraduate Nanoscience Education in Switzerland ... 1612
47.4.3 Research Experience for US Undergraduate Students 1613
47.4.4 Nanotechnology Education in South Africa 1613
47.5 Graduate Education 1614
47.5.1 Nanoscience Graduate Program in Switzerland 1614
47.5.2 Nanoscale Science and Technology Master's Program in Great Britain 1615
47.5.3 Laboratory Instrumentation Course in Taiwan 1615
47.5.4 Collaborative, Online Graduate Program in Israel and Europe 1616
47.6 Teacher Professional Development 1616
47.6.1 Teacher Training in Israel 1617
47.6.2 Teacher Training in South Korea 1617
47.6.3 Online Education for Teachers 1617
47.7 Informal Education 1618
47.7.1 Science Festival Activity in Israel 1618
47.7.2 NanoAventura in Brazil 1619
47.8 Summary and Outlook 1620
References 1622

About the Authors 1625
Detailed Contents 1645
Subject Index 1671

Subject Index

(3,6,9,12,15-pentaoxapentadecane-1,15-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethane-sulfonyl)imide] (BHPET) 1244
(3-aminopropyl)-dimethyl-ethoxysilane (APDMES) 714, 1387
(3-aminopropyl)-triethoxysilane (APTES) 713, 819
(pentane-1,5-diyl)bis(3-hydroxyethyl-1*H*-imidazolium-1-yl) di[bis(trifluoromethane-sulfonyl)imide] (BHPT) 1244
1,1′-biphenyl-4-thiol (BPT) 924, 1191
1,1′-biphenyl-4-thiol crosslinked (BPTC) 924, 1191
1,2-dichloroethane (DCE) 321
1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) 474
1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) 650
1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[amino(polyethyleneglycol)-2000] (carboxylic acid) (DSPE-PEG) 650
1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) 650
1-D localization effects 272
1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) 708
1*H*,1*H*,2*H*,2*H*-perfluorodecyltrichlorosilane (FDTS) 1441
2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid (HEPES) 830
2-D
– FKT model 995
– histogram technique 1001
– optical nanostructure 677
3-aminopropyltriethoxysilane (APTES) 713
3-D optical nanostructure 677
3-D pattern generator system 90
3-D STED PSF 854
3-mercaptopropyltrimethoxysilane (MPTMS) 150
4-Pi microscopy 853
5-fluorouracil (5-FU) 348
10,10-dibromo-9,9-bianthryl (DBBA) 319

A

Abrikosov lattice 796
absorption, distribution, metabolism and excretion (ADME) 1564
acceleration factor (AF) 1531
accelerometer 401, 1334, 1442, 1480
acoustic streaming 511
acousto-optical
– deflector (AOD) 622
– modulator (AOM) 622
actin 625
– binding protein (ABP) 1074, 1085
– microfilament 1072
activation energy 1448
active
– linearization 760
– merging 541
– sorting 544
– structure 1464
actuation 458
– advanced 1444
– force 1394
– voltage 1442
actuator 397
additive manufacturing 521
adenosine triphosphate (ATP) 625, 825, 830, 1073
adhesion 725, 1071, 1148, 1220, 1265, 1291, 1351, 1393, 1420, 1424
– cell–cell 1086
– complex 1077
– energy hysteresis 960
– force 938, 1421
– hysteresis 954
– in microstructure 1397
– linkage 1093
– measurement 875
– molecules 601
– nonspecific 599
– of biomolecular film 1387
– of biomolecule 1382
– promoter 146
– proteoglycan (AP) 600
– receptor-mediated 1086
– tip–surface 1019
– transient 1087
adhesive
– force 735, 938, 1000, 1033, 1395, 1421
– interaction surface 1019
– wear 1429, 1450
adiabatic compression 679
AFM/STM 732
agglomeration 1373, 1572
aggregation 1572
– diffusion-limited (DLA) 674
air
– -induced oscillation 747
– layer 1286
– mass (AM) 286
– -retaining Salvinia effect 1295
Airy disk 845
Al Cooper pair box (CPB) 422
Al-Ge eutectic bond 1490
alkanethiolate 146
– on gold 146
– on palladium 146
– on silver 146
allosteric effect 830
allotrope, two-dimensional 363
alumina 250
amino acid (AA) 705
aminofunctionalization 819
aminopropyldimethylethoxysilane (APDMES) 714, 1387
aminopropyltriethoxysilane (APTES) 713, 819, 1383
ammonium fluoride 1485

Amontons' law 957, 966, 1000, 1419
amorphous carbon (a-C) 88, 1143, 1365
– chemical structure 1149
– coating 1149
amplitude modulation (AM) 731, 795
– scanning force microscopy (AM-SFM) 788
angle of twist formula 754
angle-resolved photoemission spectroscopy (ARPES) 316, 374
anodic etching 657
anodization 250
anomalous 376
anti-adhesion performance 1395
anti-adhesive property 128
antibiofouling 1307, 1319
antibody 626, 1382
antiferromagnetic 306, 328
antifouling 1315
antigen-antibody recognition 462
anti-inorganic fouling 1320
antireflection grating 136
antistiction coating 1442, 1483
antistiction layer (ASL) 128
APDMES 714, 1387
apertureless scanning near-field optical microscopy (ASNOM) 852
apparent contact area 1419
application-specific integrated circuit (ASIC) 1483, 1518
APTES 713, 819
– -coated silicon 713
Ar ring 1184
arc discharge 1145
architectural design phase 1492
area expansion modulus 1089
arginine–glycine–aspartic acid (RGD) 1087
armchair
– edge 304
– -edged GNR 319
– -edged nanographene 305
– nanoribbon (AGNR) 304
– -type nanotube 195
aromatic
– ring 1184
– sextet 305
– stability 305, 308, 317
aromaticity 304
asperity 956
– contact lifetime 1426
– contact model 1420
association 814
– rate constant 818
ASU 1544
athermal flow stress 1448
atmospheric-pressure chemical vapor deposition (APCVD) 1103
atom transfer radical polymerization (ATRP) 341
atom-by-atom attrition 1002
atomic
– layer deposition (ALD) 58, 222, 255
– mass unit (AMU) 697
– resolution 730
atomic force acoustic microscopy (AFAM) 726, 878
atomic force microscopy (AFM) 4, 75, 88, 134, 176, 208, 265, 303, 321, 365, 458, 563, 587, 592, 617, 652, 693, 701, 725, 731, 770, 809, 869, 871, 952, 991, 1013, 1077, 1102, 1154, 1179, 1189, 1215, 1273, 1335, 1353, 1422, 1438, 1450, 1564, 1614
– adhesion 1192, 1200
– cantilever 571, 747, 810
– carbon nanotube tip 1028
– commercial 733
– control electronics 761
– dynamic 872
– feedback loop 761
– for UHV application 991
– imaging 810
– instrumentation 747
– manufacturer 733
– multimode 737
– noncontact dynamic (NC-AFM) 992
– probe construction 738
– static mode 753
– surface height image 1401
– test 1163
– tip wear 881
– wear measurement 1192, 1204
atomic friction 1013, 1035, 1037
– experiment 996
– thermal effect 998
atomic-scale
– force measurement 746
– imaging 727, 882
– stick–slip 1038
ATP 625, 825, 830, 1073
Au nanoparticles, mechanical properties 1367
Au pattern 150
Auger electron spectroscopy (AES) 98, 1173
automatic nanoassembly 574
average transmission probability 437

B

backlight illumination 136
bacmid system 697, 699
bacteria 589, 834
– *E. coli* 810
bacteriocidal 606
bacteriostatic 606
ball
– grid array (BGA) 1507
– milling 336, 348
– mining 353
ballistic
– phonon transport 275
– transistor 324
– transport 267
balloon angioplasty 148
bamboo multiwall nanotube 198
band structure, subband 268
bandgap 321, 324
– tunability 279
Bardeen–Cooper–Schrieffer (BCS) 786, 1004
base-growth 210
baseline drift 460
beam
– bending 1122
– Bessel 863
– plane 744
– S-shaped 1439
– theory 1085
beam deflection 1085
– FFM 989
– optical 467
– technique 459
beam–substrate interfacial energy 1394
bearing area 1440
Bell model 1088
Bell–Evans model 817
bending
– quasistatic test 1115
– stiffness 748, 1072, 1089
– strength 1105, 1113

benzocyclobutene (BCB) 1451
Berkovich tip 1162
Bessel beam 863
BHPET 1244
BHPT 1244
bias voltage 728
biased enhanced nucleation (BEN) 177
bi-dimensional optical nanostructure 679
bifurcation 1427
– prediction 1428
– steady sliding 1423
bimetallic effect 459
binding immunoglobulin protein (BiP) 1575
binding probability (BP) 816, 832
binding site, transient 831
bio
– active module 705
– adhesion 1310, 1337, 1347, 1351, 1389
– capsule 1348
– chemical sensor 445, 446
– chip 1336
– decoration 1269
– degradable 658
– device 1389
– diversity 1298
– film 600
– imaging 335, 353
– marker 597, 645
– mimetics 1269
– molecule 624
– nanotechnology 559, 1612
– polymer 625, 1085
– printing 520
– sensing 335, 351
– sensor 606, 661, 710, 1336, 1348, 1382
biocompatibility
– graphene 384
– graphene oxide 384
biofouling 1307, 1349
– formation 1310
– growth rate 1310
– surface factor 1311
biological
– NP 344, 353
– role model 1265, 1270
– surface 1269, 1279
biological molecules, nanotribological studies 1376
biomedical
– applications 627
– device 148, 1101
– sensing 810
biomedicine 644
bioMEMS (biological or biomedical microelectromechanical system) 3, 1102, 1117, 1336, 1344, 1482
– biosensor application 1382
– polymer 1103, 1348
biomimetic 1270
– application 1265, 1286, 1294
– composite material 1294
– surface technology 1270
biomolecular component 695
biomolecule synthesis 552
bioNEMS (biological nanoelectromechanical system) 3, 1102, 1336, 1348
bionics 1265, 1269
biotin 289
– surface roughness 1388
bipartite lattice, two-dimensional 307
biphenyldimethylchlorosilane (BDCS) 925, 1195
bipolar complementary metal oxide semiconductor (BiCMOS) 64
bipyridinium (BIPY) 31
bismuth 271
– -antimony-alloy nanowire 277
– nanowire 255, 270, 280
bit-patterned media (BPM) 132, 133
black silicon 257, 286
Bloch
– state 780
– wave 780
block copolymer (BCP) 74, 702
block-like debris 1170
blood
– platelet 810
– -pressure sensor chip 1336
body-on-chip system 519
bond
– energy 816
– -order potential 1015
– strength 816
bonding
– adhesive 1510
– aluminum/silicon-to-glass 1516
– anodic 1509
– covalent 1015
– epoxy 1510
– eutectic 1510
– fusion 1490, 1509
– interfacial 1519
– localized 1510, 1515, 1524
– Pyrex glass–aluminum–silicon 1527
– quartz–aluminum–silicon 1527
– silicon 1508
– solder 1510
bone morphogenic protein-2 (BMP-2) 346
boron ion implantation 730
borophosphosilicate glass (BPSG) 1512
bottom effect cone correction (BECC) 596
bottom-up 349, 578, 1615
– approach 319, 336
– chemical synthesis 415
– fabrication 1407
– method 305, 1332
boundary
– film 1215
– layer 962
– lubrication 919, 957, 969
– lubrication measurement 881
– scattering 272
bovine serum albumin (BSA) 345, 471, 472, 1376, 1383, 1564
BPTC 924, 1191
branched polyethyleneimine (BPEI) 1568
brilliant blue observation 96
broken coating chip 1163
Brownian 1488
– motion 674
– noise 1483
bubble pump 510
buckling 1156, 1160
– stress 1160
Buckyball 1142
buffered HF (BHF) 59, 1512
buffered oxide etch (BOE) 1484
bulk
– diamond 1143
– graphitic carbon 1172
– micromachining 63, 1352
– phonon mode 1016
– single crystal 1110
bullous pemphigoid antigen 1 (BPAG1) 1074
buoyancy 1392
buoyant mass 605

C

C_{60} 1000, 1142
– film 729, 1052
– ultralow friction 1053
Cajal body 1075
calorimeter, miniaturized 460
cancer 596, 633
cantilever 289, 458, 466, 739, 1077, 1395
– array 460, 1335
– deflection 732, 988
– deflection calculation 743
– effective mass 989
– flexible 732
– free standing 1439
– hollow 602
– mass 731
– material 739
– mechanics 747
– microbeam 1116
– motion 731
– mount 735
– passivated reference 459
– piezoelectric 466
– polymer 1134
– resonance behavior 749
– stainless steel 1161
– stiffness 737, 748
– thickness 988
– triangular 739, 748, 754
– untwisted 735
– V-shaped 740, 989
– wear 715
cantilever beam
– array (CBA) 1393
– fabrication procedure 1132
– friction force calibration 745
– lateral bending 1124
cantilever-in-cantilever (CIC) 475
capacitance detection 755
capacitive
– detection 760
– detector 731
– pressure sensor 398
– -type silicon 1334
capacitor 409
– nanowire 287
capillary
– condensation 951, 1439, 1441
– force 1440, 1442
– valve 506
capped nanotube 1030
– tip 1030
capping 1483
carbon 1141
– amorphous (a-C) 88, 1143, 1365
– -based molecular device 305
– crystalline 1141
– magnetron sputtered 1148
– meta-nanotube 217
– nanofilament 199
– nanohorn (CNH) 1366
– unhydrogenated coating 1153
carbon nanotube (CNT) 193, 210, 256, 323, 324, 363, 431, 559, 568, 1051, 1294, 1335, 1350, 1561
– application 227
– catalyst-free growth 210
– catalytically activated growth 210
– electronic property 215
– enviornmental impact 231
– field-effect transistor (CNFET) 431
– filled 217
– functionalization 222
– helical 568
– mechanical property 216
– optical property 215
– oxidation 220
– oxidized 221
– property for sensing 432
– reactivity 216
– resonator sensing capability 450
– semiconducting single-walled 432
– separation 440
– sidewall functionalization 221
– single-wall (SWNT) 3, 146, 194, 431, 569, 702, 741, 780, 1332, 1350
– tip 741
– toxicity 231
– unzipping 368
carnivorous plants 1279
carrier
– density 271
– gas 1151
– mean free path 267, 271
– mobility 271
– plate exchange 991
Casimir force 91, 937
casting 144
catalysis 335, 353
catalyst 258, 262
– support 231
catalytic chemical vapor deposition (CCVD) 202, 440
catenane 39
cathodic arc carbon 1155
CdS nanowire 284
CdSe nanorod 286
cell 636, 645
– -adhesion molecule (CAM) 592
– anisotropy 1082
– -based assay 520
– computational description 1082
– contractility 1076
– deformation 633, 636
– growth 591, 604
– homogeneity 1082
– injection 577
– –matrix adhesion 1086
– mechanics 633, 1069
– mechanics model 1085
– membrane 1088
– morphology 604
– nucleus 1075, 1090
– poking 1083
– red blood cell membrane 1089
– rheological property 631, 1081
– strain 1091
– viscoelastic property 1081
– wall 590
– Young's modulus 1083
cellular
– compression 1090
– force application 1078
– force sensor 577
– mechanosensor 1095
– solid 1085
– trichomes 1286
ceramic
– matrix composite 223
– nanobeam 1113
chaotic advection 512
characterization
– chemical 1149
– electrical 375
– nanowire 262
– structural 262, 264
charge
– -coupled device (CCD) 88, 622, 1319, 1421
– density wave (CDW) 782, 1005
– discharge model (CDM) 1496
– exchange interaction 937
– injection 1442
charged carboxylated polystyrene (COOH-PS) 1574
chemical
– bonding force 731

– characterization 1149
– composition 1573
– degradation 1192, 1205, 1233
– force microscopy (CFM) 227
– mechanical polishing (CMP) 169, 353, 1405, 1487
– sensor 230, 445
chemical vapor deposition (CVD) 55, 87, 88, 97, 129, 202, 251, 255, 323, 366, 742, 1142, 1338, 1463
– atmospheric pressure (APCVD) 1103
– catalytic (CCVD) 202, 440
– graphene 366
– hot filament (HFCVD) 176
– low pressure (LPCVD) 55, 165, 1109, 1352, 1404, 1471, 1511
– metalorganic (MOCVD) 56, 179, 251
– STEM electron-beam-induced 100
chemically assisted ion beam etching (CAIBE) 417
chemiluminescence 1609
chemoselective conjugation 708, 709
chemotherapy 645, 706
chiral angle 196
chirped-pulsed amplifier (CPA) 105
CHO-K1 830
chord theorem 951
circular dichroism spectroscopy (CS) 1576
circularly permuted (CP) 708
– biofunctional 708
Clar's aromatic sextet rule 304, 311
clay nanoparticle 1118
climb 1445
climbing plants 1289
coated polymer surface, adhesion 1389
coating 1573
– continuity 1173
– damage 1170
– failure 1167
– friction and wear behavior 1019
– friction coefficient 1032
– hard 1145, 1365
– mass density 1150
– microstructure 1149
– properties 1145
– substrate interface 1166
– thickness 1157, 1160
cobblestone model 958
coefficient of friction (COF) 957, 964, 977, 990, 1000, 1034, 1161, 1170
– lubricant 1033, 1057
– relationship 743
coefficient of thermal expansion (CTE) 223, 463, 1527
coercivity 289
coflowing 548
collagen 828
collective intelligence system 1597
colloidal
– particle 335
– probe 600
– probe experiment 601
colony forming unit (CFU) 606
coloration signal 1290
compact disc, capillary 114
complementarity determining region (CDR) 709
complementary metal-oxide-semiconductor (CMOS) 58, 115, 132, 167, 257, 324, 398, 431, 471, 542, 1341, 1480, 1508, 1520
– bipolar (BiCMOS) 64
complex pattern 148
composition of matter 1619
compression-decompression cycle 962
compressive 459
– forces nanotube 1054
– stress 1160
compute pattern generator (CPG) 93
computed tomography (CT) 1577
conductance 288
– quantized 267, 268
– thermal 275
conductivity
– differential 775
– nanowire 273
– thermal 273
confinement 937
confocal laser scanning microscopy (CLSM) 837, 1577
consilience 1590
constant
– current mode 728
– force mode 761
– height mode 728
– *NVE* 1018
contact
– area estimation 1001
– conductance 1002
– mode 811
– mode photolithography 154
– printing (CP) 143, 157
– stiffness 1158
– time 149, 1286
contact angle (CA) 1269, 1275, 1285, 1322
– hysteresis (CAH) 1285, 1322
– nanotribological property 1189
contaminating particle 1292
contamination 732, 1451
continuous stiffness measurement (CSM) 1104, 1158
continuous wave (CW) 855
continuum
– mechanic of single asperity 1000
– model 996
– regime 966
– theory 944
contracting cell 1078
contrast
– agent 645
– in photon emission 857
control
– system 758, 761
– unit (CU) 728
controlled
– geometry (CG) 730
– orientation 208
convergence–divergence cycle 1596
converging technology 1588, 1590, 1591
cooling, rapid 492
copper
– Cu(111) 1003
– hexadecafluorophthalocyanine 154
– surface adhesion 1034
– surface wear 1034
– tip 1034
coprecipitation 336, 349, 353
Coriolis acceleration 403
corona splash 1286
cost of good (COG) 699
Coulomb
– interaction potential 1016
– law 1423
– law of friction 998
covalent bonding, modeling of material 1015
μCP (microcontact printing) 146, 156

crack 1156
crater formation, surface 1021
C-reactive protein (CRP) 474
creep 1118, 1445
– model 972
critical
– concentration 1003
– micelle concentration (CMC) 702
– normal load 1365
– shear stress 957
– thinking 1611
– velocity for stick–slip 975
Crk-associated substrate (CAS) 1092
cross
– -domain language 1597
– -flow 548
– hair alignment 153
– linked 123
– -linking protein 1081
– sectioning 1454
crosstalk 762
crystal
– birefringent 752
– growth direction 253
– morphology 1277
– orientation 253
– preferred orientation 253
– structure 733
crystallinity 1275
cumulative failure function 1529
curcumin (CURC) 650
curing agent 147
curved surface 147
cuticle 1279
cuticular fold 1281
cyclic
– fatigue 1158
– olefin copolymer (COC) 498, 547
cyclotron radius 272
cylindrical
– mirror analyzer (CMA) 98
– roller stamp 152
– shape 647
cytometry 636
cytoplasm 589
cytoskeletal matrix 1072
cytoskeleton (CSK) 590, 1072, 1087
– computational description 1082
– constituent 1072
– nonlinear response 1082
cytosol 588
cytosolic side (c-side) 830

D

damage mechanism 1163
Damascus steel 1604
D-band 312
DBBA 319
DCE 321
DDMS (dimethyldichlorosilane) 1441
de Broglie wavelength 268
debris 1170
Debye
– frequency 1017
– length 943
deep reactive-ion etching (DRIE) 61, 165, 258, 513, 577, 1489, 1507
deep ultraviolet (DUV) 117
defect site 149
deflection 814
– sensitivity 595
deformable mirror display (DMD) 68
degree of freedom (DOF) 564
dehydration reaction 148
delamination 1153, 1160
delivery
– particle 1390
– system 660
demolding 122
dendrimer 336
dendritic structure 673
density
– functional theory (DFT) 214, 316
– gradient ultracentrifugation (DGU) 367, 440
– of defect 152, 154
density of states (DOS) 268, 277, 433, 776
– phonon 273
deoxyadenosine triphosphate (dATP) 516
deoxycytidine triphosphate (dCTP) 516
deoxyguanosine triphosphate (dGTP) 516
deoxyribonucleic acid (DNA) 25, 77, 81, 132, 136, 487, 587, 623, 626, 627, 697, 1075, 1562
– complementary (cDNA) 518
– double stranded 707
– hybridization 462, 473
– origami 627
– sensor 289
– tetrahedra 821
deoxythymidine triphosphate (dTTP) 516
depletion
– attraction 937
– force 950
– stabilization 937
deposition
– rate 1146
– sealing 1519
– technique 1145
Derjaguin–Landau–Verwey–Overbeek (DLVO) 945
Derjaguin–Muller–Toporov (DMT) 595, 952, 1000, 1419
design
– for reliability 1437
– of experiments (DOE) 1486
– rule 748
– rule check (DRC) 1494
desolvation 345
destructive fabrication 571
detection system 732, 749
deterministic lateral displacement (DLD) 544
dewetting 1360
DI TESP 874
diagnostic 645, 1382
– tools 637
dialysis 338, 349, 353
diamond 1141
– bulk 1143
– coating 1153
– film 1143
– friction 1036
– nanoindentation 1027
– tip 1162, 1167, 1353
diamond-like carbon (DLC) 89, 894, 1032, 1141, 1181, 1355, 1359, 1365, 1451
– coating 1032, 1143, 1149, 1365
– coating microstructure 1160
– film 897
– pillar 91
diblock copolymer 252
dibromoterfluorene (DBTF) 1006
dichlorodiphenyltrichloroethane (DDT) 472
dichlorvos (DDV) 501
dideoxynucleotide triphosphate (ddNTP) 516
die per wafer (DPW) 1495

dielectric 1443
– charging 1442
– function 278
dielectrophoresis (DEP) 513, 541, 634
dielectrophoretic force 261
differential conductivity 775
differential scanning calorimetry (DSC) 267
diffusion-limited aggregation (DLA) 674
digital
– feedback 761
– light processing (DLP) 1335, 1482
– micromirror device (DMD) 5, 404, 1181, 1331, 1400, 1437, 1446
– signal processor (DSP) 728
diisopropyl methylphosphonate (DIMP) 471, 472
dilatancy 968
dimethyl sulfoxide (DMSO) 349, 819
dimethylacetamide (DMAc) 349
dimethyldichlorosilane (DDMS) 1441
dimethylformamide (DMF) 349
dimethylmethylphosphonate (DMMP) 472
diode 268
– light-emitting 282
– nanowire 282
dip-pen nanolithography (DPN) 77, 701
Dirac
– cone 307
– fermion 304
– point 307
direct
– current/radio frequency (DC/RF) 1144
– gap material 435
– laser writing (DLW) 103
– numerical simulation (DNS) 647
– stochastic optical reconstruction microscopy (dSTORM) 858
– write electron beam/focused ion beam lithography 144
directed self-assembly (DSA) 74, 317
discoidal
– nanoparticle 657
– polymeric nanoconstruct (DPN) 650, 655
– shape 647
discrete track recording (DTR) 133
disease 645
– infectious 606
dislocation
– defect accumulation 1449
– glide 1447
– motion 957
– nanoindentation 1022
– stick–slip 1038
dispersion interaction 937
displacement sensor 1394
display 229
– paperlike 152
dissociation 814
– path 819
– rate constant 818
distance distribution function 1440
distortion 759
distributed
– Bragg reflector (DBR) 157
– feedback (DFB) 157
– laser resonator 157
dithiol monolayer 151
DMA 367
DNA 25, 77, 81, 132, 136, 487, 587, 623, 626, 627, 697, 1075, 1562
– complementary (cDNA) 518
– double stranded 707
– hybridization 462, 473
– origami 627
– sensor 289
– tetrahedra 821
docetaxel (DTXL) 650
domain pattern 795
donor impurity concentration 278
dopant 54
DOPC 474
DOTA 650
double
– -layer interaction 937, 943
– -wall nanotube (DWNT) 201, 440
doxorubicin (Dox) 346
DP 1198, 1363
DPPC 650
drag 1265
– -reducing shark skin 1270
– reduction 1265
drift
– correction 792
– thermal 459
driving frequency 750
droplet
– application 550
– chemical reaction 550
– detection technique 542
– digital polymerase chain reaction (ddPCR) 546
– generation 538
– merging 541
– microfluidic 537
– sorting 543
– splitting 541
Drosophila melanogaster dopamine transporter (dDAT) 830
drug
– binding 830
– discovery 553
– screening 519
– target 606
drug delivery 335, 353, 662, 1102
– device 1336
– graphene 384
– vehicle 704
dry etching 60
dry sliding friction 1033
DSPE-PEG 650
dual-in-line packaging (DIP) 1513
Dupré equation 955
dye penetrant 1453
dynamic
– (atomic) force microscopy 872
– force spectroscopy (DFS) 817
– interaction 937
– light scattering (DLS) 668
– mode 460, 462, 731

E

E. coli bacteria 810, 834
Earth-scale platform 1593
EDC 708
edge 303
– armchair 305
– disorder 147
– geometry 305
– state 307, 316, 321, 328
– zigzag 305
effective
– medium theory 278
– PSF 853
– shear stress 1002
– spring constant 835, 994
– stiffness 1083

EHS 1551
eigenfrequency 462
elastic
– contact 1419
– continuum 1083
– deformation 1000
– modulus 1103, 1118, 1154
– tip–surface interaction 1027
elastin-like polypeptide (ELP) 346
elastohydrodynamic lubrication (EHL) 957, 966
elastomer stamp 148
electric
– -arc reactor 200
– field gradient microscopy (EFM) 265
– force gradient 737
electrical
– contact resistance (ECR) 1430
– double-layer capacitor (EDLC) 384
– feedthrough 1520
electrochemical
– AFM 738
– nanowire growth 261
– pump 510
– STM 730
electrochemical deposition (ECD) 251, 253
– pulsed 254
– selective 260
electrohydrodynamic
– mixing 513
– pump 510
electrokinetic 494
– flow 1348
electroless 657
– deposition 146, 669
electroluminescence (EL) 282
electromagnetic
– energy 646
– force 941
– interference (EMI) 230
electron
– correlation 309, 310, 312, 321
– emitter 229
– energy loss near edge structure (ELNES) 315
– energy loss spectroscopy (EELS) 265, 1146, 1149
– Fermi wavelength 268
– holography 95
– transport 317
– tunneling 727
– unpaired 306
π-electron 304
electron beam 87, 97, 701
– assisted evaporation 469
– evaporation 149, 336
– induced deposition (EBID) 568
– lithography (EBL) 75, 118, 305, 317, 418, 441, 651, 1405
– nanofabrication 75
– resist 148
– three-dimensional nanostructure fabrication 97
electron cyclotron resonance (ECR) 1148
– CVD 1145, 1148, 1365
– CVD coating 1167
electron diffraction (ED) 1275
– low energy (LEED) 772
– selected area 263
electron-beam-induced CVD
– in situ observation 98
– three-dimensional nanofabrication 102
electronegativity 1183
electronic 335, 353
– ink 154
– newspaper 154
– noise 733
electroosmotic pump 510
electroplating 57, 1507
electrospraying 336, 353
electrostatic
– actuated valve 506
– actuation 1394
– discharge (ESD) 230, 1438, 1496
– double layer (EDL) 1391
– force 937, 943, 1442
– loading 1424
electrowetting
– micropump 511
– -on-dielectric (EWOD) 541
elliptical shape 647
elongation speed 1445
embedded atom method (EAM) 1015
embedding energy 1016
embossing technique 149
emulsion inversion point (EIP) 338
emulsion–solvent extraction 345
encapsulation 1506
– vacuum 1511, 1530
endoplasmic reticulum (ER) 1574
– stress 1574
endothelium barrier 645
energy
– barrier 814
– -dispersive x-ray spectroscopy (EDS) 264, 502
– dose 858
– gap 304, 308, 312, 317
– landscape 816
– MD simulation 1014
– -transfer mechanism 960
engineered
– nanomaterial (ENM) 1560
– nanoparticle (ENP) 1570
enhanced permeability and retention (EPR) 650
entangled state 957
entropic
– force 937
– repulsion 961
enzyme-linked immunosorbent assay (ELISA) 626, 832
epicuticular wax 1265, 1268, 1272
– crystal 1281
epidermal surface 1282
epidermis 1268, 1283
– cell 1268, 1283
epi-poly 1487
epitaxial growth 56, 261
epitaxy 256
– molecular beam (MBE) 56, 75, 179, 1407
epithelial–mesenchymal transition (EMT) 519, 1564
equilibrium dissociation constant 816
equity issues 1550
ER stress 1575
erythrocyte (EC) 1563
etching 144
– anisotropic 58
– dry 60
– isotropic 58
– metal-assisted chemical (MACE) 257, 656
– sacrificial 1520
– through wafer 1507
– wet 59
ethanolamine-hydrochloride 819
ethylenediamine pyrocatechol (EDP) 59
eukaryotes 589
Euler equation 748

evaporation
– electron beam-assisted 469
– thermal 469
exchange, carrier plate 991
exciton, diffusion length 286
exfoliation 367
– graphite 365
– high-shear mixing 367
– micromechanical 365
– solvent 367
– ultrasonication 367
exosome 345
external
– foundry 1483
– vibrations 729
extracellular
– matrix (ECM) 502, 592, 1077
– nanoparticle 344
– polymeric substance (EPS) 1310
– signal-regulated kinase (ERK) 1574
extreme ultraviolet (EUV) 53, 117

F

fabricated silicon 1390
fabrication technology 395
– destructive 571
faceted cavity 284
F-actin 1072
failure
– density function 1529
– mechanism 1172, 1437
– mode 1437
– modes and effects analysis (FMEA) 1496
fatigue 1158, 1448
– crack 1170
– cyclic 1158
– damage 1158
– measurement 1469
– strength 1107, 1114
– test 1161
FDTS 1441
feedback
– circuit 733
– loop 762, 990
– network 727
femtosecond laser (fs-laser) 87, 103
– nanofabrication 103
– photopolymerization 104
fenestration 650
ferritin 345
fetal bovine serum (FBS) 1389
few-layer graphene (FLG) 372
fiber
– matrix 1085
– optical interferometer 751
FIB-milled probe 741
fibrinogen (BFg) 1564
field
– cooling 796
– emission 283
field-effect transistor (FET) 172, 230, 269, 282, 325, 379, 432, 569, 693, 702, 1348
– carbon nanotube 436
– heterojunction 713
– immunoFET 710
– nanowire 269
– protein sensing 693, 710
– sensing channel 711
– spin 329
field-emission scanning electron microscopy (FESEM) 564, 1565
field-programmable logic arrays (FPGA) 283
filament
– cross-linker 1094
– elastic property 1074
– intermediate 1086
filled carbon nanotube 217
– in situ filling method 218
– molten-state filling method 218
– wet chemistry filling method 218
filled hairs 1283
film 53, 1028
– bulk acoustic resonator (FBAR) 1519
– C_{60} 729, 1052
– graphene 377
– nanoindentation thickness 1028
– nonconductive 1146
– surface interface viscosity 1045
– transparent conductive 377
filtered cathodic arc (FCA) 1145, 1146, 1365
– coating 1155
– deposition 1145
fimbriae 592
fine leak 1453
finite
– -difference time-domain (FDTD) 107
– element analysis (FEA) 1461
– element modeling (FEM) 1103, 1443
– -size effect 280
fish-hooked hairs 1289
flavonoid 1274, 1279
flexible
– cantilever 732
– electronic device 152
flip-chip technique 1507
floating catalyst method 205
flow
– -focusing configuration 540
– -focusing geometry 549
– laminar 492
– nanoliter 1345
fluid
– drag reduction 1293
– force microscopy (FluidFM) 602
fluidized bed (FB) 203
– CCVD reactor 204
fluorescence 29
– (or Förster) resonant energy transfer (FRET) 1080
– -activated droplet sorting (FADS) 544
– total internal reflection (TIRF) 832, 857
fluorescent photoactivated localization microscopy (fPALM) 857
fluorescent polystyrene particle 1573
fluorinated silane monolayer 149
fluorographene 369
focal adhesion kinase (FAK) 1092
focused ion beam (FIB) 87, 128, 671, 730, 1438
– cross section 1450
force
– attractive 937
– between surfaces in liquid 937
– between surfaces in vacuum 937
– calibration 988
– calibration plot (FCP) 1220
– Casimir 937
– colloidal 937
– detection 748
– determination 1080
– –distance curve 593, 737
– electrostatic 937
– entropic 937
– generation 624, 1081
– hydration 937
– hydrodynamic 937

– hydrophobic 937
– indentation 1027
– ion correlation 937
– law 938
– MD simulation 1014
– nonequilibrium 937
– oscillatory 937
– protrusion 937
– repulsive 937
– sensing tip 757
– sensor 732, 991, 1394
– solvation 937
– spectroscopy 599, 626, 761, 791
– spectrum 817
– structural 937
– undulation 937
foreign body giant cell (FBGC) 1563
forward recoil spectrometry (FRS) 1149
fouling 1307
foundry
– manufacturing 1486
– MEMS 1483
Fourier transform infrared spectroscopy 1253
four-quadrant photodetector 988
fractal 673
– dimension 676
fracture
– failure 1116
– Griffith theory 1116
– stress 1105
– surfaces 1115
– toughness 1103, 1109, 1114, 1154, 1157, 1164, 1463
free standing cantilever 1439
free-space nanowiring 93
– electrical property 95
freezing-melting transition 972, 975
Frenkel–Kontorova (FK) 1006
Frenkel–Kontorova–Tomlinson (FKT) 995
frequency
– measurement precision 757
– modulation (FM) 731, 795
– -modulation scanning force microscopy (FM-SFM) 771, 788
– response 847
frication time dependence 1428
friction 725, 987, 1013, 1033, 1220, 1470
– accelerated 1169
– and wear 1387
– attachment 941
– coefficient 957, 962, 968, 977, 1000, 1034, 1161, 1170
– dry sliding 1033
– interlayer 572
– load dependence 1000
– loop 994
– lubrication 1033, 1057
– measurement 1199
– measurement methods 742
– MEMS 1429
– metal surface 1034
– of biomolecule 1382
– of polysilicon 1474
– profile 1110
– rate-state 1422
– SAM 1046
– scale dependence 902
– surface 1033
– temperature dependence 999
– -torque model 1397
– velocity dependence 1428
friction force 716, 731, 998, 1044, 1161, 1392
– calibration 735
– curve 762
– magnitude 744
– map 1001
– measurement 872
– of Si(100) 1360
friction force microscopy (FFM) 725, 870, 959, 987, 1013, 1189, 1353, 1398, 1419, 1450
– in water 993
– tip 994
fuel cell 329
fullerene 702, 781, 1141
– frictional property 1052
– particle 1142
functionalized nanotube 220
fundamental resonant frequency 749
fused deposition modeling (FDM) 521
fusion bonding 1490

G

G protein 1070
GaAs 149
– nanowires 286
– stamp 149
– wafer 150
G-actin 1072
gamma-globulin (γ-Ig) 1564
GaN nanowire 262, 281, 288
gap 316
– stability 727
gas exchange 1293
gate control 269
gauge factor (GF) 449
G-band 312
Ge 170
– nanowire 257
geometry 548
– effect in nanocontact 999
getter 1489
giant magnetoresistive (GMR) 352, 1335
glancing angle deposition (GLAD) 105
gland 1279
glass 149
– capillary 469
– frit 1483
– nanochannel 250
– slide 147
– transition temperature 126, 957, 1117
glucose transporter 4 (GLUT4) 832
glutaraldehyde 812
glutathione peroxidase (GSH-Px) 1562
gold 730
– -coated 821
– eutectic 1489
– film 146
– nanomechanical properties 1367
– nanoparticle (GNP) 645, 1367
– nanowire 285
– pattern 150
grain
– boundary scattering 272
– nanoindentation boundary 1022
– size 147, 1143
graphene 303, 305, 363, 365
– application 376
– bilayer 370
– composite 377
– decomposition 366
– derivatives 368, 372
– doping 374
– few-layer (FLG) 372
– film 377
– gene transfection 385

– nanoribbon (GNR) 303, 366, 379, 1006
– nanostructure 303
– properties 369
– quantum dot 379
– transistor 379
– zigzag-edged 306
graphene oxide (GO) 367
– reduced 367
graphite 1141
– cathode 1145
– flake 1053
– intercalation compounds (GIC) 364
– sheet 1053
grating 157
– antireflection 136
green fluorescent protein (GFP) 832, 1090
Greenwood and Williamson (GW) 1420, 1440
Griffith fracture theory 1116
gross leak 1453
ground-state depletion (GSD) 855
guanosine triphosphate (GTP) 1074
gyroscope 403, 1480

H

Hagen–Poiseuille equation 501
hair 1276
Hall effect 375
– quantum 376
Hamaker constant 902, 939, 942, 1391
hard
– amorphous carbon 1145
– coating 1365
– law 1552
hardness 1103, 1118, 1149, 1154, 1354
harmonic oscillator, driven damped 462
heat 646
– curable prepolymer 144
– dissipation 1526
– mode 460
heating
– localized 1510, 1515
– rapid 492
– rapid cooling 492
helical nanotube 195
helium ion microscope 318
hemodynamic resistance 1392
HEPES 830
hermeticity 1442, 1451
herringbone texture 197
Hertz contact model 595
Hertz-plus-offset relation 1000
hetero-coagulation 343
heterodyne interferometer 750
heterogeneous CCVD 203
heterointegration 1500
heteronanotube 217
hexadecane thiol (HDT) 147, 924, 1191, 1395
hexadecyl trimethyl ammonium bromide (CTAB) 955, 1573
hexagonal boron nitride (hBN) 380
hexagonal honeycomb polysilicon (HEXSIL) 72
hexamethyldisilazane (HMDS) 52
hierarchical
– sculpturing 1279
– structure 1265
high aspect ratio 122
– combined poly- and single-crystal silicon (HARPSS) 73
– MEMS (HARMEMS) 1334, 1406
– tip (HART) 741
high content screening (HCS) 1573
high density interconnect (HDI) 1508
high density lipoprotein (HDL) 345
high electron mobility transistor (HEMT) 182
high school education 1607
high-angle annular dark-field-scanning transmission electron microscope (HAADF-STEM) 101
high-cycle fatigue 1449
high-definition television 132
higher harmonic 812
highest occupied molecular orbital (HOMO) 780
highly ordered monolayer 146
highly oriented pyrolytic graphite (HOPG) 81, 260, 315, 737, 882, 1277
high-modulus elastomer 146
high-pressure liquid chromatography (HPLC) 543
high-resolution
– electrode 154
– patterning 152
– printing 143, 147
– stamp 144
– transmission electron microscopy (HRTEM) 219, 263, 564
high-temperature
– MEMS/NEMS 1403
– superconductor (HTSC) 785
high-throughput 637
– screening (HTS) 520, 537, 1575
high-volume manufacturing (HVM) 137
hinge memory effect 1446
hollow
– cantilever 602
– nanoparticle 345
holographic optical tweezers (HOT) 635
homodyne interferometer 750
honeycomb lattice 303, 304
Hooke's law 814
horizontal coupling 729
hot embossing lithography (HEL) 117
– thermoplastic 117
hot filament chemical vapor deposition (HFCVD) 176
human
– body model (HBM) 1496
– –computer interaction 1594
– interleukin (hIL) 708
– -scale convergence 1592
– -scale platform 1594
– serum albumin (HSA) 345, 474
Human Genome Program 1336, 1549
hybrid
– continuum-atomistic thermostat 1018
– nanocarrier 664
– self-assembly-based chip-to-wafer bonding (HAS-CtW) 1502
hybridization 312, 518, 646
hydration force 947
hydrocarbon 1183
– precursor 1148
– unsaturated 1184
hydrodynamic
– radius 501
– shear forces 636
hydrogel 502
hydrogen
– concentration 1150
– content 1153

– evolution recovery (HER) 1344
– silsesquioxane (HSQ) 327
hydrogenated
– carbon 1149
– coating 1149
hydrogen-terminated diamond 1030, 1037
hydrophilic 1280
– anchor cell 1284
– coverage 1293
– SAM 1049
– surface 1285
– surface chemistry 1293
hydrophobic 1280
– coating 1441
– force 948
– interaction 703
– SAM 1049
– surface for nanotribology 1180
hydrophobicity 1285, 1363, 1397, 1608
hydrothermal synthesis 349
hyperbolic metamaterial 278
hyperconjugation 312
hysteresis 267, 759
– loop 733, 762

I

IBD (ion beam deposition) 1145
illumina sequence 517
illumination function 849
image
– processing software 763
– topography 731
imaging
– bandwidth 738
– therapy 661
– tool 263
iMEMS® 1483
immersed boundary method (IBM) 647
immersed finite element method (IFEM) 647
immunoFET 710
– limitation 712
– molecular nanocomponent 718
– planar 693, 712
– receptor 712
immunoglobuline G (IgG) 473, 695, 712
imprint
– lithography 261
– polymer 157
impurity 1573
in situ
– growth study 264
– sharpening of the tips 728
incoherent cutoff frequency 848
indentation 1104
– creep 1104, 1120
– depth 1154, 1157
– experiment 834
– hardness 725
– induced compression 1160
– rate 1025
independent component analysis (ICA) 472
indium tin oxide (ITO) 152, 229, 283, 378
inductively coupled plasma (ICP) 172
– mass spectrometry (ICP-MS) 1577
– optical emission spectroscopy (ICP-OES) 1577
inelastic
– regime 1102
– tunneling 778
inertial measurement unit (IMU) 1487
inertial MEMS 1482
– high-volume manufacturing 1487
inertial sensors 400
infrared (IR) 645
– spectroscopy 278
inkjet 504
– -spotting 469
inorganic fouling 1310
– surface factors 1311
inorganic NP 336, 342, 349, 353
inorganic–organic NP 343
inositol-requiring protein 1 (IRE1) 1575
InP nanowire 279, 286
in-plane
– mechanical strain 146
– stresses nanoindentation 1023
input/output (I/O) 1507
insect catching plants 1291
instrumentation 1605, 1611, 1614
integrated
– circuit (IC) 51, 163, 324, 353, 395, 1506
– comprehensive droplet digital detection (IC3D) 542
– polymer optic 132
– tip 739
interaction rate constant 814
interatomic
– attractive force 743
– force 732
– spring constant 732
intercalation 327
interconnect 326
interfacial
– defect 1161
– engineering 693
– friction 958
– stress 1164
interferometric detection sensitivity 752
interferometry 1438
interlayer friction 572
intermediate
– filament 1074
– or mixed lubrication 957
– thickness film 967
intermittent contact 811
intermolecular force 814
internet of things (IoT) 1430, 1498
intervalley transition 307, 313
intracellular
– NP 344
– pathway 1091
intracuticular wax 1268
intramolecular 814
– junction 574
– structure 790
intravalley
– scattering 313
– transition 308
intravital microscopy 655
intrinsic stress 1153
invariant natural killer T (iNKT) cell 826
InvenSense 1487
involuntary exposure 1550
ion
– channel 1095
– implantation 1356
– plating technique 1145
– pump (IP) 98
– source 1146
ion beam (IB) 1144, 1146
– deposition (IBD) 1145
– sputtered carbon 1146

ionic liquid film 1242
– dicationic 1249
ionizing radiation 1442
irreversible deformation 1445
isofunctional object 699
isogeometric analysis (IA) 647
isoleucine–lysine–valine–alanine–valine (IKVAV) 703
isopropanol alcohol (IPA) 1513
isopropyl-β-D-thiogalactoside (IPTG) 836
isospectral structure 781
isotopic tracing (IT) 1577
isotropic etching 147
I–V characteristics 277, 424

J

Jahn–Teller effect 782
jellium approximation 1016
jet-and-flash imprint lithography (JFIL) 118
Johnson–Kendall–Roberts (JKR) 595, 952, 1000, 1419
jump into contact 1392
jump-to-contact (JC) 789, 1019

K

Kelvin probe force microscopy (KPFM) 769, 794
kinesin 625
kinetic friction 994
knife-edge blocking 760
Kohn anomaly 374
Kondo effect 778
Kramers' diffusion model 818
kurtosis 1392

L

lab-on-chip (LoC) 132, 488, 538, 1336, 1345
– system 514
laminar flow 492
laminated n-channel transistor 154
Landau level 273
Langevin dynamics approach 1017
Langmuir–Blodgett (L–B) 260, 924, 1033, 1181, 1216, 1361
– film 36
– technique 260
Laplace
– force 1221
– pressure 951
large-scale integrated circuit (LSI) 324
laser 284, 618
– ablation 257, 336, 1406
– assisted growth 256
– -based vapor deposition 348
– deflection technique 733
– interference lithography 145
– nanowire 284
– three-dimensional nanostructure fabrication 103
– tracking microrheology (LTM) 635
– trap 619
– trapping 565
– tweezers Raman spectroscopy (LTRS) 635
lasing threshold 284
lateral
– bending 1124
– contact stiffness 987
– force 754, 988, 998
– force microscopy (LFM) 146, 725, 871, 988
– manipulation 793
– resolution 728, 731
– spring constant 738
– stiffness 761, 1002
laterally integrated nanogenerator (LING) 288
lattice
– Boltzmann method (LBM) 647
– every 959
– imaging 732
layer
– lubricant 1042
– of resist 144
layer-by-layer (LBL) 377
lead zirconate titanate (PZT) 57, 164, 180, 287, 562, 728, 1343
leaf margins 1284
leaf surface 1268, 1275, 1283
Lennard-Jones (LJ) 1016
– potential 902, 1016
leukocyte rolling 1087
Lieb's theorem 308
life cycle 1544, 1551
lifetime 816
– broadening 776
Lifshitz theory 942
lift mode 738
ligand 591
– -receptor interaction 810
light
– absorbant 229
– beam deflection galvanometer 752
– -emitting diode (LED) 132, 282–284, 1617
– -induced fluorescent (LIF) 542
– -sheet microscopy 862
limit-of-detection (LoD) 438
linear variable differential transformer (LVDT) 760
linearization, active 760
linker of nucleoskeleton and cytoskeleton (LINC) 1075
linker protein 1074
lipid 645
– bilayer 590, 1071
– vesicle 1089
liposome 1390
liquid 936, 1042
– -crystal display (LCD) 134
– crystal on silicon (LCoS) 134
– film thickness measurement 926
– lubricant 1042
lithium ion battery 230, 286
lithography 52, 115, 261, 336, 701, 1332, 1349
– hot embossing (HEL) 117
– imprint 261
– jet-and-flash imprint (JFIL) 118
– laser interference 145
– multiple beam interference 103
– nanoimprint (NIL) 113, 1406
– self-aligned imprint (SAIL) 124
– soft (SL) 114, 1406
– step-and-stamp imprinting (SSIL) 131
– STM 318
– substrate conformal imprint (SCIL) 130
lithography, electroforming and molding (LIGA) 70, 169, 509, 1101, 1337, 1404, 1446
– microsample 1446
– process 1406
– technique 1101
living plant material 1276
living prototype 1265
load
– -carrying capacity 1171
– dependence of friction 1000

– peak indentation 1157
– peak-to-peak 1104
load–displacement
– characteristic 1103
– curve 1105, 1157
– profile 1123
loading rate 814
local
– contact potential difference (LCPD) 769, 794
– deformation study 876
– density approximation (LDA) 308, 311, 319
– density of states (LDOS) 308, 311, 775
– stiffness 761
localization 276
– effects 272, 277
– microscopy 856
localized surface
– elasticity 878, 917
– plasmon (LSP) 672
locked nuclei acid (LNA) 476
logic gate 29, 283
London dispersion interaction 937, 942
longitudinal piezoresistive effect 755
loop-mediated isothermal amplification (LAMP) 515
lotus 1286
– effect 1265, 1270
– leaf 1608
low-cycle fatigue 1449
low-density lipoproteins (LDL) 345, 473
low-energy electron diffraction (LEED) 772
lowest observable effect concentration (LOEC) 1575
lowest unoccupied molecular orbital (LUMO) 780
low-noise measurement 991
low-pressure chemical vapor deposition (LPCVD) 55, 165, 1103, 1352, 1404, 1471, 1483, 1511
low-temperature (LT)
– AFM/STM 732
– oxide (LTO) 68, 169
– scanning tunneling microscopy (LT-STM) 773
lubricant
– film 1041, 1384
– layered structure 1057
– liquid 1359
– long-chain hydrocarbon 1043
– perfluoropolyether 1359
lubrication 335, 353, 1352
– monolayer 1422
– property ceramics 1034
– property polymer 1031
luminescent moss 1290

M

MAC mode 812
machine model (MM) 1496
machining 336
macromolecule 694
macroscale
– friction 1352
– tribological characterization 1168
magnetic 645
– application 289
– bead microrheology 1077
– data storage 1343
– exchange force microscopy (MExFM) 769, 797
– exchange force spectroscopy (MExFS) 798
– exchange interaction 798
– field 623
– force gradient 737
– force microscopy (MFM) 134, 265, 738, 769, 795
– logic architectures 290
– memory array 289
– moment 310, 312
– nanowire 279, 289
– NP 349, 646
– resonance force microscopy (MRFM) 769, 798
– resonance imaging (MRI) 352, 661, 718, 798, 1577
– shell cross-linked knedel-like nanoparticles (MSCK) 1344
– storage 290, 1144
– thin-film head 1172
– tweezers 623, 626
– twisting cytometry 1078
magnetic force microscopy (MFM)
– sensitivity 738
– vortex imaging 796
magnetically actuated valve 506
magnetization switching 798
magneto
– -optical Kerr effect 279
– resistance 272, 1144
– resistive random access memory (MRAM) 329
magnetohydrodynamic pump 510
magnetron sputtered carbon 1148
malondialdehyde (MDA) 1562
mammalian cell 810
manipulation
– of individual atom 725
– of whole cells 629
manufacturing 1611
marginal ray angle 846
mask, high resolution stamp 144
mass 462, 603
– flow controller (MFC) 98
– of cantilever 731
– sensors, beam-based 603
massless Dirac fermion 304, 307
material
– initiative 1593
– lubricant 1057
matrix electronic paper, active 154
matrix-assisted laser desorption ionization (MALDI) 553
– -time of flight mass spectrometer (MALDI-TOF) 554
Matthiessen's rule 272
Maugis–Dugdale model 1000
maximum likelihood estimator (MLE) 1531
maximum load, nanoindentation 1023
mCube 1487, 1490
mean
– -field theory 944
– free path 268
– time to failure (MTTF) 1441, 1531
measurement chamber 465
mechanical
– deflection 458
– field stability 1497
– micropump 509
– phenotyping 636
– properties of cells 633
– stability fullerene 1052
– stiffness 654
– stress 1445
mechanical property 1102, 1115, 1459
– experimental technique 1103

– measurement 1471
– of DLC coating 1161
mechanically controlled break junction (MCBJ) 267
mechanics of cantilever 747
mechanochemistry 573
mechanosensing 1094
– protein 1091
mechanosensitive channel of large conductance (MscL) 1095
mechanotransduction 1070, 1092
megacity initiative 1593
Meissner effect 796
melting temperature 1445
membrane 830, 1071
– deformation 1088
– stiffness 1088
– surface tension 1089
membrane-type surface stress sensor (MSS) 468
MEMS fabrication 62
– material properties 1495
– process control parameters 1495
MEMS/NEMS
– lubrication studies 1359
– mechanical property 1102
– microfabrication processes 1351
– nanotribology 1336
– radio frequency 1332
– tribological property 1352
meniscus force 902
mercaptopropyltrimethoxysilane (MPTMS) 150
mesoporous
– silicon particle 645
– structure 659
messenger ribonucleic acid (mRNA) 518
metal 171, 645
– -assisted chemical etching (MACE) 257, 656
– eutectic bond 1485
– evaporated (ME) 911, 1144
– matrix composite 223
– NP 352
– oxide NP 336, 337
– oxide surface 789
– particle (MP) 911, 1172
– particle (MP) tape 1172
– salt reduction 336, 353
– tip, deformable 1019
metal/insulator/metal (MIM) 155
metallic
– bonding, modeling of material 1015
– microbeam 1116
– nanowire 268
– NP 337, 351
– single-walled carbon nanotube (m-SWNT) 432
metalorganic chemical vapor deposition (MOCVD) 56, 179, 251, 255
metal–oxide–semiconductor (MOS) 88, 1442
metal–oxide–semiconductor field-effect transistor (MOSFET) 325, 713, 1348, 1430
metal-to-metal contact 1471
metastases 596
methotrexate (MTX) 668
methylmethanesulfonate (MMS) 1566
mica 737, 812, 996
micelle 349
– worm-like 349
micro total analysis system (μTAS) 488, 490, 1344
micro wine glass 89
micro/nanoindenters 1103
micro/nanoscale friction, velocity dependence 893
micro/nanostructures using bending tests 1108
microactuator 1339
microamplification device 515
microcantilever (MC) 459, 471, 1335, 1342
micro-computer tomography (micro-CT) 514
microcontact printing (μCP) 114, 117, 146, 1181
microdevice 1101
microelectromechanical system (MEMS) 2, 51, 163, 395, 441, 443, 561, 869, 1013, 1077, 1101, 1102, 1145, 1179, 1216, 1331, 1334, 1419, 1437, 1459, 1481, 1505
– electric field stability 1496
– field performance 1496
– foundry 1483
– friction 1419, 1429
– manufacturing strategy 1482
– nonsilicon 1334
– optical 404
– packaging 1480, 1505
– polymer 1524
– radio frequency (RF MEMS) 172
– resonator 413
– resonator-based oscillator 414
– static friction force 1397
– stiction 1397
– tunable capacitors 410
microemulsification 338
microemulsion 336, 348, 353
microengine 1338, 1429
microfabrication 146, 396, 464, 1351, 1406
microfilament 1072
microflow-focusing device (MFFD) 664
microfluid diffusion 492
microfluidic 260, 489, 662, 1345
– device 516
– fuel cell 521
– mixing 662
– network (μFN) 469
– paper-based analytical device (μPAD) 499
– sequence 516
– system 634
– technology 488
– trap 678
microfriction directionality effect 887
microimplant 1382
microindenter 1077
microinductor 575
micromachine 1102
micromachined
– inductor 411
– polysilicon beam 1393
– silicon 1332
– structure 1460
– surface 713
micromanipulation 565
micromechanical calorimetry 464
micromechanical characterization 1154
micromirror array 68, 407, 1335
micromotor 396, 1392, 1397
– component, surface roughness 1392
– lubricated 1398
micro-nanofluidic 504
– application 504
– design 491
– fabrication 496

microoptoelectromechanical systems (MOEMS) 3, 1332, 1522
micropatterned SAM 1194
micropipette 602
– aspiration 1077–1079
microplate platform 537
microplate stretcher 1077
micropump 509, 1346, 1347
microreactor 514
microscale
– friction 884, 1251, 1353, 1392
– scratching 875, 906
– wear 875, 907, 1353
microscope eigenfrequency 747
microscopic asperity 978
microscratch 1161
microsensor 397, 1352
microspark erosion 1406
microstructures, adhesion measurements 1393
microswitch 1430
microsystem technology (MST) 2, 1102, 1332
microtransformer 156
microtriboapparatus 1394
microtubule (MT) 625, 1074, 1086
– filament 1094
microvalve 505, 1348
microwave plasma CVD (MPCVD) 177
microwave-PECVD (MPECVD) 1149
microwear 1161, 1167
Mie regime 618
millimeter-scale device 1406
milling 336
millipede 1335
MIM capacitor 155
miniature device 562
miniemulsion 343
minimum inhibitory concentration (MIC) 606
misfit angle 995
mitogen-activated protein kinase (MAPK) 1082, 1574
mobile exhibit 1609
mobility 326
mode coupling grating 157
modified EAM (MEAM) 1016
modified simulated body fluid (mSBF) 344
modulus
– elastic 725, 1154
– of rigidity 748
molding 144
molecular
– adsorption 435
– beam epitaxy (MBE) 56, 75, 179, 1407
– cloning device 697
– conformation 1217
– dynamics (MD) 993, 1014, 1018
– engineering 693
– imaging 781
– motors 625
– recognition 821
– recognition phases (MRP) 472
– sieve 252
– spring model 1362
– switch 28
– weight 120
monocationic liquid film 1244
monolayer indentation 1030
monolithic elastomer device (MED) 542
Monte Carlo 1493
MoO_3 nanoparticle 1051
Moore's law 116
moral order 1549
More Moore approach 74
morpho butterfly scale quasistructure 96
MoSe nanowire 266
most probable unbinding force 816
most-probable rupture 818
motor
– neuron 705
– protein 1076, 1078, 1094
MPTMS 150
multiasperity measurement 1422
multicellular 1283
– base 1284
– hair 1279
– structure 1265, 1284
multichip module (MCM) 1508
multicomponent device 710
multifunctional surface 1265
multilayer
– device 155
– graphene (MLG) 327
– nanoparticle 657
– thin-film 1172
multilayered wax film 1276
multimedia 1609
multimode AFM 737
multimolecular layer 1359
multiple beam interference lithography 103
multiple path particle dosimetry (MPPD) 1573
multiple-asperity friction 1419
multipole interaction potential 1016
multistage strategy 645
multi-user MEMS process (MUMP) 1511
multiwalled (carbon) nanotube (MWNT) 197, 567, 742, 874, 1054, 1350, 1366, 1565
– bc 198
– bh 198
myocbacteria 607
myosin 625

N

N,*N*-dimethylamide (DMA) 367
N-acetyl-cysteine (NAC) 1576
NAND flash memory 132
nanoasperity junction 1430
– nanoassembly 574
NanoAventura 1619
nanobarcode 657
nanobiological device, design paradigm 695, 697
nanobiotechnology 1595
nanocapsule 340
nanocarrier 645
nanochannel 104, 1365, 1405
nanochemistry 1407
nanoclay 1118
nanocomposite material 40, 1051
nanocone 679
nanoconstruct 646
– accumulation 650
nanocrystal 1149, 1390
nanodeformation 1216
nanodevice 695
– supramolecular 693
NanoEIS 1620
nanoelectromechanical system (NEMS) 2, 62, 163, 395, 396, 415, 441, 559, 869, 1101, 1102, 1145, 1179, 1216, 1331, 1437, 1459, 1481, 1505, 1615
– fabrication 415
– radio frequency resonators 419
nanoelectronic 559
nanoELSI 1543
nanofabrication 88, 103, 146, 559, 644, 726, 914
nanofatigue 1158

nanofiber 346
nanofluidic
– channel 136
– device 1101, 1348
– silicon array 1348
nanogap 43
nanogel 340, 352
nanogenerator
– laterally integrated (LING) 288
– vertically integrated (VING) 287
nanographene 303
nanohardness 1153
nanohelix 23
nanoimprint lithography (NIL) 113, 1406
nanoindentation 915, 1019, 1154, 1354, 1460
– hardness measurement 875
– maximum load 1023
– MD simulation 1019
– measurement 877
– surface atom relaxation 1021
nanoindenter 1103, 1108
nanoinjection 660
nanolens array 672
nanoliposome 664
nanolithography 701, 1335
nanomachine 561
nanomachining 726
nanomanufacturing 2, 335, 353
nanomaterial, heterologous 707
nanomechanical characterization 1402
nanomechanics 1069, 1352
nanomedicine 645, 1619
– synthesis 662
nanometer 1013
– resolution 144
nanometer-scale
– device 1013, 1332
– electronic device 1051
– friction 999, 1013
– friction, ceramics 1034
– indentation 1013
– lubrication 1013
– material property 1019
nanomotor 563
nanoneedle 657
nano-objects
– friction 1365
– mechanical behavior 1365
nanooptoelectromechanical systems (NOEMS) 3, 1332
nanoparticle (NP) 41, 201, 221, 335, 643, 1051, 1366, 1559
– application 351
– Au 1367
– classification 336
– clay 1118
– discoidal 657
– electrical property 351
– extracellular 344
– hollow 345
– hybrid 336, 342, 353
– inorganic 336, 342, 353
– intracellular 344
– magnetic 349, 646, 1348
– magnetic property 349
– mechanical property 350
– metal 336, 352
– naturally occurring 344
– organic 338, 344, 353
– polymer 336, 352, 646
– property 348
– quasihemispherical 657
– shape 349
– size 348
– synthesis 336
– tip 1051
nanopharmacological 831
nanopillars 657
nanopore 1349
– sequence 517
nanopositioning 758
nanoprecipitation 338, 341, 349, 353, 663
nanorobotic
– assembly 568
– manipulation 560
– manipulator (NRM) 563
nanorobotics 559
nanorod (NR) 1366
– CdSe 286
nanorope 703
nanoscale
– adhesion 1352
– adhesion measurement 1377
– antibiotics 706
– device 1365
– distance, tunable 700
– friction measurement 1378
– friction wear mapping 898
– protein interface 693
– texture 976
– wear 906
– wear test 1380
nanoshuttle 1609
nanostructure 1276, 1571
– bending test 1104
– bi-dimensional optical 679
– fabrication, three-dimensional 88, 103
– finite element analysis 1125
– hybrid 707
– measurement of mechanical properties 1103
– roughness 1125
– scratch 1125
– scratch test 1103
– space-filling 702
– stress and deformation analysis 1101
nanoswitch 1430
nanosystem technology (NST) 2, 1102, 1332
nanotechnology 2
– material 695
– method 644
nanotherapeutic device
– barrier to practice 717
– construction method 701
– intellectual property 701
nanothermometer 267
nanotransfer printing (nTP) 148, 155
nanotribological properties of silicon 1355
nanotribology 713, 870, 1219, 1352, 1422, 1429
nanoTruck 1609
nanotube (NT) 23, 256, 336, 346, 349, 415, 572
– assembly 572
– bamboo multiwall 198
– -based sensor 1335
– bundle 1053
– capped 1030
– coated 221
– decorated 221
– double-wall (DWNT) 201, 440
– field-effect transistor (FET) 703
– functionalized 220
– helical 195
– junction 573
– zigzag-type 195
nanovesicle 346
nanowire (NW) 43, 63, 415, 657, 1335
– aligned 262, 263
– applications 290
– arrangement 262

– boundary scattering 272
– capacitor 287
– CdS 284
– characterization 262
– compositionally modulated 258
– core-shell 265
– crystallinity 252
– diameter distribution 256
– FET 269, 711
– film, aligned 261
– free-space 95
– gating 269
– junction 282
– magnetic 279, 289
– metal 253
– metallic 254
– morphology 258
– optical bandgap 278
– palladium 289
– pearl-shaped 262
– photodetector 285
– physical properties 262
– position control 260
– Raman spectra 281
– semiconducting 711
– sensor 288
– superlattice 253, 265, 277, 284
– synthesis 250
– thermal conductivity 274
– thermal stability 267
– transport property 271
– ZnO 261, 266, 284, 288
nanowire growth
– electrochemical 254, 260
– laser-assisted 256
– MACE 257
– oxide-enhanced 257
– plasma-assisted 256
– RIE 257
– solution phase 259
– stress-induced 260
– template 255
Navier–Stokes equation 647
n-decylphosphonate (DP) 1198, 1363
near-edge x-ray absorption fine structure (NEXAFS) 315
near-field
– microscopy 227
– scanning optical microscopy (NSOM) 279, 849
near-infrared (NIR) 385
– imaging 661
negative contact force 1048
network shear modulus 1086
neuroactive agent 705
neuroscience 1596
neurotransmitter sodium symporter (NSS) 830
neutron activation analysis (NAA) 1577
Newtonian
– flow 957
– viscosity 1084
next-generation lithography (NGL) 115
N-hydroxysuccinimide (NHS) 820
nitrile-tri-acetic acid (NTA) 820
nitroxide mediated polymerization (NMP) 341
n-octadecylmethyl (dimethylamino) silane (ODDMS) 1198, 1363
n-octadecylphosphonate (ODP) 1198, 1363
n-octyldimethyl (dimethylamino) silane (ODMS) 1198, 1363
noise
– electronic 733
– power 439
– source 754
Nomarski interferometer 752
nonacosan-10-ol 1271, 1278
nonbonding π-electron state 305, 306
nonconducting film 1146
nonconductive sample 729
noncontact
– dynamic force microscopy (NC-AFM) 992
– friction 1003
– imaging 731
noncovalent interaction 26
nonlinear
– enhancement factor 860
– illumination 860
– I–V behavior 277
non-mechanical pump 510
non-Newtonian flow 957
nonpolar end group 1359
nonsilicon MEMS 1334
nonspherical tip 1001
nonvolatile random access memory (NRAM) 229
normative 1548
Nosé–Hoover thermostat 1017
nTP (nanotransfer printing) 155
nuclear
– analytical and related technique 1578
– analytical technique (NAT) 1576
– magnetic resonance (NMR) 798, 1275
– pore complex (NPC) 811
– radiation 1442
– stiffness 1091
nucleation of nanowire 262
nucleic acid 24
nucleocytoskeletal coupling 1076
nucleus, mechanical property 1091
numerical aperture 846
Nyquist sampling theorem 857

0

octadecyltrichlorosilane (ODTS) 1050
octadecyltrichlorosilane (OTS) 80, 1189, 1422, 1441
octamethylcyclotetrasiloxane (OMCTS) 971
ODDMS 1198, 1363
ODMS 1198, 1363
ODP 1198, 1363
one-dimensional (1-D) 105, 215, 249, 363, 434, 564, 702, 782, 994, 1423
– quantum effect 267
on-off current ratio 326
optical
– absorption 280
– beam deflection 732
– cavity 284
– data switching 407
– deformability 630
– detector 731
– -electronic-optical switches (O-E-O) 407
– fiber 147
– force 618
– head 735
– interferometer 749
– MEMS 404
– microscope (OM) 564
– microswitch 1342
– properties 278
– ridge waveguide 148
– sectioning 862
– stretcher (OS) 629, 630
– switch 284

– transfer function (OTF) 848
– transition, phonon-assisted 280
– trap 618, 622, 633, 1077–1079
– tweezers (OT) 620, 1079
– tweezers Raman spectroscopy (OTRS) 635
optical lever 752
– angular sensitivity 754
– deflection method 761
– optimal sensitivity 754
optimal beam waist 754
optoelectronic
– component 156
– graphene 382
order
– long-range 957
– out-of-plane 957
organ model 519
organ on a chip 519
organic
– inverter circuit 154
– light-emitting device (OLED) 132
– NP 338, 344, 353
– system 646
– transistor 152, 154
organic chemistry 1182
– graphene 367
– synthetic 367
organic–inorganic hybrid NP 342
organisms
– multicellular 589
– unicellular 589
organometallic vapor phase epitaxy (OMVPE) 1407
organosilsesquioxane 150
ormocer 128
oscillating tip 734
oscillatory
– force 946
– simple shear 1084
osmotic force 950
Ostwald ripening 337
oxidation furnace 54
oxide enhanced growth 257
oxidized carbon nanotube functionalization 221

P

package design 1493
packaging
– 3-D 1521
– hermetic 1511
– IC 1506
– MEMS 1505
– polymer-MEMS 1524
– vacuum 1511
– wafer-level 1518
packed system friction 1046
paclitaxel (PTX) 664
palladium nanowire 289
paperlike display 152
papilla 1282
papillated epidermal cell 1290
papillose 1283
parasitic
– capacitance 1483
– charge 1442
particle size 1570
passivated reference cantilevers 459
passive
– linearization 759
– merging 541
– sorting 543
– structure 1461
pathogen 1279
pattern
– of ink 144
– transfer 149
patterned
– self-assembled monolayer (SAM) 146
– substrate 262
patterning 261
Pauli repulsion 902
paxillin (Pax) 1092
peak indentation load 1157
peak-to-peak load 1104
peapod 703
pearl-shaped nanowire 262
peel number 1441
Peierls instability 782
pendulum AFM 992
pentaerythritol tetranitrate (PETN) 472
peptide
– amphiphile (PA) 346, 703
– NP 346
– nucleic acid (PNA) 476
– toroid 705
peptidoglycan 590
perfluorodecanoic acid (PFDA) 1182, 1341
perfluorodecylphosphonate (PFDP) 1198, 1363
perfluorodecyltricholorosilane (PFTS) 1198, 1363
perfluorodecyltriethoxysilane (PFDTES) 1365, 1389
perfluoropolyether (PFPE) 919, 1215, 1359
perfluoropolyether lubricants 1359
periclinal wall 1283
periodic
– boundary condition (PBC) 676, 1016
– potential 994
perpendicular scan 744
persistence length 628
personal meaning mapping (PMM) 1619
Persson 1000
perylenetetracarboxylic dianhydride (PTCDA) 1006
petal 1282
pH 472
– sensor 288
pharmacokinetic 519
pharmacophore 703
phase
– -breaking length 272
– change micropump 510
– inversion 338
– inversion temperature (PIT) 338
– lock loop (PLL) 463, 992
– transformation, nanoindentation 1026
– transition model 973
phenanthrene 88
phonon 374
– confinement effect 280
– density of states 273
– mean free path 273
– -phonon scattering 274
phosphate buffered saline (PBS) 716, 1377, 1384
phosphatidylinositol-3-kinase (PI3K) 1082
phosphosilicate glass (PSG) 55, 68, 167, 1512
photoactivatable green fluorescent protein (PA-GFP) 857
photoactivatable protein fluorophores (PA-PF) 857
photoactivated localization microscopy (PALM) 593, 857
photodamage 858
photodetector
– four-quadrant 988

– nanowire 285
– quad 734
photoemission spectroscopy (PES) 316, 776
photoinitiator 123
photolithographic fabrication 1404
photolithography (PL) 53, 115, 134
– contact mode 154
– near-field conformal 146
– pattern 144
– projection mode 144
– proximity mode 144
photoluminescence (PL) 278, 344
– NSOM imaging 284
photomodification 103
photomultiplier tube (PMT) 542
photon flux rate 863
photon noise 856
photonic band gap (PBG) 105
photonic crystal (PhC) 88, 102, 132
photonics, graphene 382
photopolymerization 103, 126
photoresist 147, 148
– master 1133
photosolidification 103
photoswitchable dyes 858
photosynthesis 1289
photothermal spectroscopy 460, 464
photovoltaics 285, 329
physical vapor deposition (PVD) 56, 167, 222, 255, 336
physisorption 460
Pickering emulsion polymerization 343
picoindentation 915
picomotor 563
piezo
– ceramic material 759
– effect 758
– hysteresis 772
– nonlinearity 772
– relaxation 772
– scanner 762
– stack 760
– translator 762
– tube 729, 758
piezoelectric 287
– drive 727
– leg 991
– material 180
piezoelectrically actuated valve 507
piezoresistive
– cantilever 755
– coefficient 755
– detection 755
– longitudinal effect 755
– pressure sensor 397
– sensor 447
pigment particle 154
pile-up
– nanoindentation 1023
– surface 1027
– surface atom 1021
piranha etch solution 1113
planar immunoFET 693, 711
plant hair 1283
plant surface 1267
– structure 1265
– superhydrophilic 1286
– superhydrophobic 1286
plant systematics 1284
plasma
– assisted growth 256
– deposited coating 1441
– deposition 1441
– -enhanced chemical vapor deposition (PECVD) 55, 155, 167, 319, 739, 874, 1109, 1145, 1148, 1181, 1340, 1365, 1472
– etching 1405
– frequency 278
plasmon polariton (PP) 679
plasmonic nanostructured device 678
plastic
– circuit 152
– contact regime 901
– electronics 155
– large area circuit 153
plastic deformation 1041, 1160
– surface 1022
plastic/elastic model 1440
platelet-derived growth factor (PDGF) 629
platinum-iridium tip 730
p-n junction 283
– nanowire 269, 282
pneumatic actuated valve 508
point spread function (PSF) 847
polarity 1183
pole-tip recession (PTR) 1172
polishing 1143
poly(diallyldimethylammonium chloride) (PPDDAC) 1573
poly(dimethylsiloxane) (PDMS) 80, 127, 146, 149, 288, 366, 469, 490, 541, 651, 811, 967, 1080, 1101, 1103, 1133, 1181, 1318, 1365, 1406
poly(ether urethane) (PEUT) 471
poly(ethylene glycol) methacrylate (PEGMA) 342
poly(ethylene terephthalate) (PET) 150
poly(lactic-*co*-glycolic acid) (PLGA) 339, 650
poly(lactide fumarate) (PLAF) 349
poly(lactide-*co*-glycolide fumarate) (PLGF) 349
poly(methacrylic acid) (PMAA) 342, 473
poly(methyl methacrylate) (PMMA) 71, 102, 117, 135, 173, 224, 252, 287, 339, 372, 416, 441, 498, 1101, 1117, 1319, 1389, 1406, 1525
poly(*N*-isopropylacrylamide) (PNIPAm) 343
poly(propyl methacrylate) (PPMA) 1103, 1117
poly(propylene glycol) methacrylate (PPGMA) 342
poly(styrene-*b*-dimethylsiloxane) (PS-b-PDMS) 317
poly[(*m*-phenylenevinylene)-*co*-(2,5-dioctoxy-*p*-phenylenevinylene)] (PmPV) 225, 322
polyamidoamine (PAMAM) 700
polyaniline (PANI) 384
polybutadiene (PBD) 967
polycaprolactone (PCL) 339
polycyclic aromatic hydrocarbon (PAH) 303–305, 367
polyelectrolyte complexation 345, 348
polyether ether ketone (PEEK) 466
polyethylene (PE) 1051
polyethylene dioxythiophene (PEDOT) 225
polyethylene glycol (PEG) 338, 384, 471, 819, 1568
polyethylenimine (PEI) 346, 473
polyglycolic acid (PGA) 348
polyimide (PI) 183, 498
polylactic acid (PLA) 339, 664
polymer 645
– biocompatible 1101
– bioMEMS 1103, 1348
– brush 1043
– cantilever 1134
– chemistry 702
– filament 1084

– film 714
– hydrophilic 461
– hydrophobic 461
– ion-track–etched foil 250
– material 183
– mediated force 949
– microbeam 1133
– NP 352
– sensing layers 461
– transparent microcapsule 154
– unfilled 1123
polymerase chain reaction (PCR) 489, 552, 1346, 1526
– -capillary electrophoresis (PCR-CE) 508
polymeric
– domain 702
– nanocarrier 645
– nanoparticle 336, 344, 349, 646
– unit self-assembling 705
polymeric microbeam 1117
– bending 1121
– scratch tests 1117
polymer-matrix composite 224
polymersome 664
polymorphonuclear (PMN) 1563
polyoxymethylene (POM) 1337
polypeptide 25
polypropylene nonwoven fabric (PPNF) 502
polysaccharide NP 346
polysilicon (poly-Si) 496, 1459
– doped 1357
– film 1109, 1357
– film, nanotribological 1357
– hexagonal honeycomb (HEXSIL) 72
– mechanical property 1471
– undoped 1357
polystyrene (PS) 135, 339, 498, 967, 1103, 1117, 1573
polystyrene/nanoclay composite (PS/clay) 1117
polytetrafluoroethylene (PTFE) 963, 1051, 1181
polyvinyledene fluoride (PVDF) 562
pore 656
poroelastic solid 1083
porous silicon
– particle 659
– pillar 657
porphyrin 1007
position accuracy 762
position-sensitive detector (PSD) 465, 595, 734
power
– dissipation 992
– factor 275
– spectrum density function 677
– spectrum, MD simulation 1040
Prandtl–Tomlinson model 993
– finite temperature 998
– one-dimensional 993
– two-dimensional 994
precipitation 348
– hardening 1448
prepolymer 147
pressure
– and flow sensing 475
– injection technique 252
– osmotic 588, 589
– sensor 398
– touch-point 398
– turgor 589
prestress 1086
primary creep 1446
printed
– coil 156
– DFB resonator 157
probability density function (PDF) 816
process
– design kit (PDK) 1485
– gas 1148
processing contamination 444
projection mode photolithography 144
prokaryotes 589
promyelocytic leukemia body (PML) 1075
prostate-specific antigen (PSA) 474
protein 24
– biologically produced 709
– circular permuted 710
– coating 1382
– database 696
– kinase (PKA) 474
– kinase C (PKC) 1082
– kinase inhibitor (PKI) 474
– layer, nanoscale adhesion 1376
– NP 345
– protein interaction 1078
– sensing interface 713
– sensor 693
– unfolding 1078
pseudospin 307
p-type/intrinsic/n-type (PIN) 172
pull-in voltage 1443
pull-out voltage 1444
pulsed laser deposition (PLD) 58
pure water repellency 1295
purine nucleotide (PN) 830
PZT (lead zirconate titanate) 57, 164, 180, 287, 562, 728, 1343
– tube scanner 728, 733

Q

Q-(quality) factor 411, 462, 738, 748, 821, 1515, 1532
quad photodetector 734
quadrant
– detector 753
– photodiode (QPD) 622
quantitative structure–activity relationship (QSAR) 1569, 1575
quantum
– boxes (QB) 1335
– computer 329
– conductance 437
– confinement 268, 270, 278, 290
– dot (QD) 329, 351, 352, 379, 665, 1407, 1561
– electromechanics (QEM) 422
– Hall effect (QHE) 376
– size effect 271, 317
– wire (QWR) 1335
quartz
– crystal microbalance (QCM) 1014
– tuning fork 772
quasi-continuum model 1026
quasihemispherical nanoparticle 657
quasistatic bending tests 1115
quasistructure 97

R

radial cracking 1156
radio frequency (RF) 55, 409, 1332, 1438, 1471
– magnetron sputtering 1144
– MEMS (RF MEMS) 172, 408, 1334, 1443, 1444, 1449
– MEMS/NEMS 1332
– NEMS resonators 419
radius of curvature 461
radius of gyration 120, 967
Raleigh's method 747

Raman
– optical tweezers (OTRS) 635
– spectroscopy 281, 312, 644, 1149
ramping 816
random access memory (RAM) 229
rapid
– cooling 492
– thermal annealing (RTA) 166
– thermal processing (RTP) 1513
rapid expansion, supercritical solution 338, 341
ratchet mechanism 1036
rate-and-state friction (RSF) 1417, 1422
– measurement 1425
rational design 646
Rayleigh
– criterion 846
– instability 267
– regime 618
reaction force 1398
reactive
– ion etching (RIE) 60, 134, 165, 257, 324
– oxygen species (ROS) 1562
– spreading 147
readout electronics 753
real contact area 1419
real-time deformability cytometry (RT-DC) 637
receptor 591
– binding, Bell model 1088
– –ligand complex 1087
– -mediated adhesion 1086
– transmembrane 1092
recognition, molecular 821
rectangular cantilever 739
red blood cell (RBC) 635
redistribution layer (RDL) 1519
reference cantilever 460
reflection high-energy electron diffraction (RHEED) 58
reflexive 1547
regeneration 1276
relative humidity (RH) 899, 1227, 1338, 1532
relaxation time 957
release profile of the payload 659
reliability 1437
– packaging 1528
relief on a surface 144
remanence 289
remote detection system 753
repetitive mechanical stress 1449
repulsive tip–substrate 1024
residual stress 1153, 1160, 1459, 1471
– measurement 1459
resistance
– antibiotic 606
– negative differential 277
– temperature-dependence 270
resolution, vertical 731
resonance
– curve detection 738
– frequency 458, 462, 603
– Raman effect 281
resonant
– sensor 449
– tunneling 277
resonator, graphene 370
rest time effect 1222
reverse
– mapping 1597
– osmosis (RO) 1309, 1316
– -transcription PCR (RT-PCR) 554
reversible
– addition and fragmentation transfer (RAFT) 340
– saturable optical fluorescence transition (RESLOFT) 854
rheology 119, 631
ribonucleic acid (RNA) 588, 710, 1075
rice leaf surfaces 1318
rigidity 1086
robotics 559
robust design phase 1493
roll embossing 120
root mean square (RMS) 715, 878, 1143, 1199, 1421
rotary pump 509
rotaxane 38
rotor
– –hub interface 1392, 1393
– –stator interface 1337
rough surface 952, 978
roughness 1126, 1440
rubbing/sliding surface 1450
Ruderman–Kittel–Kasuya–Yoshida (RKKY) 786

S

sacrificial
– etching 1520
– layer 1349
Sad1p/UNC-84 (SUN) 1076
Sader method 594
Salvinia 1274, 1284
– effect 1265
– paradox 1279
Sanger sequencing 516
saturation irradiance 855
scan
– area 735
– direction 744
– frequency 735
– range 758
– rate 729
– size 736
– speed 759
scanning
– acoustic microscopy (SAM) 1453
– capacitance microscopy (SCM) 726, 876
– chemical potential microscopy (SCPM) 726
– electrochemical microscopy (SEcM) 726
– electron microscopy (SEM) 64, 75, 88, 125, 146, 166, 252, 262, 366, 396, 502, 564, 652, 726, 741, 908, 988, 1112, 1156, 1266, 1337, 1438, 1462, 1512
– electrostatic force microscopy (SEFM) 726
– field emission electron microscopy (FESEM) 564, 1565
– force acoustic microscopy (SFAM) 726
– force microscopy (SFM) 458, 726, 769, 788, 1049
– head 728
– ion conductance microscopy (SICM) 726
– Kelvin probe microscopy (SKPM) 726, 876
– magnetic microscopy (SMM) 726
– near-field optical microscopy (SNOM) 564, 726, 849
– probe lithography 145
– probe microscopy (SPM) 75, 227, 563, 726, 769, 1451, 1605
– -probe-induced oxidation 76
– speed 735
– spreading resistance microscopy (SSRM) 876, 1237
– system 758
– thermal microscopy (SThM) 265, 726

– transmission electron microscope (STEM) 88, 256
– tunneling microscopy (STM) 4, 75, 257, 265, 303, 373, 561, 725, 730, 769, 774, 870, 992, 1020, 1335
– tunneling microscopy (STM) lithography 318
– tunneling spectroscopy (STS) 265, 303, 374, 771, 774
scattering 273, 276
– forces 620
Schottky barrier (SB) 437, 438
scratch
– critical load 1171
– damage mechanism 1164
– depth 1034
– drive actuator (SDA) 1418
– force 1034
– induced damage 1161
– measurement 725
– resistance 1104, 1110, 1121, 1196
– test 1109, 1162
S-curve 1449
sealing 1506
– glas 1483
– hermetic 1510
second harmonic generation (SHG) 280
second phase inversion 338
secondary alcohol 1274
secondary battery 329
Seebeck coefficient 275
selected area electron diffraction (SAED) 263, 266
selective plane illumination microscopy (SPIM) 862
selenium 259
self-aligned imprint lithography (SAIL) 124
self-assembled monolayer (SAM) 40, 69, 117, 146, 442, 459, 471, 819, 924, 1029, 1179, 1186, 1216, 1315, 1359, 1361, 1425, 1441
– coated surface 1390
– coating 1441
– edge resolution 146
– head group 1186
– micropatterned 1194
– spacer chain 1186
– sudden failure 1365
– surface 1186
– terminal 1186
– tribological property 1362
self-assembled structure 1045
self-assembly 36, 78, 282, 283, 348, 349, 702, 1271, 1275
– chemical 78
– physical 78
self-assembly-based multi-chip-to-wafer bonding (SA-MCtW) 1502
self-cleaning
– lotus effect 1269, 1295
– surfaces 1315, 1318
self-healing 1288
self-replicating assembly 574
self-similar chain 672
semiconducting single-walled carbon nanotube (s-SWNT) 432
semiconductor 173
– nanowire, gating 275
semimetal-semiconductor transition 271, 276
semisynthetic nanodevice 708
sensing device 710
sensitivity 449, 755, 989
sensor 269, 458
– biological 288
– characterization 438
– chemical 288, 379
– DNA 289
– fabrication 439
– gas 379
– graphene 379
– hydrogen gas 289
– mass 379
– nanowire 288
– neuron 705
– performance 444
– pH 288
– release 1483
– thermal 274
sequencing 516
serotonin transporter (SERT) 830
shape 1571
– memory alloy (SMA) 507
shear 1445
– deformation 636
– flow 1077, 1079
– melting 957
– modulus 988, 1081, 1083
– strength 1000
– stress 1078
– -thickening (dilatant) fluid 966
– -thinning (pseudoplastic) fluid 966
shifted order 859
Shockley–Queisser limit 286
shot noise 754
shrinkage 106
signal-to-noise ratio (SNR) 438
silane
– chemistry 128
– film 715
– fluorinated monolayer 149
– polymer 714
silanization process 1384
silicon 149, 163, 732, 755, 1352
– anode 287
– APTES-coated 713
– -based surface 1382
– beam 1113
– black 257, 286
– cantilever 988
– coated 1352
– fabricated 1390
– high pressure phase 1026
– mechanical property 1471
– microimplant 1382
– nanobeam 1131
– nanotribological properties 1355
– nanowire 257, 263, 274, 287
– nitride (Si_3N_4) 732
– on-insulator (SOI) 577
– oxidation 54
– porous 657
– Si(100) 1167, 1353
– stamp 149
– thermal conductivity 274
– tip 990
– tribological performance 1352
– wafer 144, 147
– wear data 1357
silicon carbide SiC 173, 323
– bulk properties 1353
– decomposition 366
– film 1357
silicon nitride Si_3N_4 990
– tip 739, 1389
silicon oxide SiO_2 beam 1113
silicon-on-insulator (SOI) 164, 410, 577, 1509
– microelectromechanical systems (SOIMEMS) 1483
– substrate 416
silicon-on-sapphire (SoS) 56
siloxane polymer lattice 1387

silver nanoparticle (AgNP) 645, 646, 672, 1568
simple harmonic oscillator (SHO) 594
single
– cell force spectroscopy 602
– crystal, silicon cantilever 739
– electron transistor (SET) 422, 569
– friction 1419
– mode optical fiber 147
– molecule friction 1005
– molecule localization microscopy (SMLM) 593
– nano-object friction 1368
single-asperity 1000
– contact 900
– friction 1422
– measurement 1422
single-chain fragment variable (SCFv) 695, 712
single-crystal reactive etching and metallization (SCREAM) 65
single-molecule
– force spectroscopy (SMFS) 809
– manipulation 621
– real-time (SMRT) 517
single-particle tracking PALM (sptPALM) 858
single-photon emission computed tomography (SPECT) 1577
single-walled (carbon) nanotube (SWNT) 3, 146, 194, 431, 569, 702, 741, 780, 1332, 1350, 1564, 1566
– advantage 431
– application 431
– application example 445
– assembly 441
– biosensor 1350
– electronic property 432, 434
– integration 439
– mechanical property 434
– metallic (m-SWNT) 432
– optical property 435
– production 440
– sensor energy efficiency 447
– structure 432
site-directed coupling 820
size 644
– and scale 1608
– effect, one-dimensional (1-D) 268
– shape, surface and stiffness 647
skewness 1392
S-layer 822
sliding
– induced chemistry 1035
– of tip 988
– velocity 1192
– work 1041
slip 1445
slippery liquid-infused surfaces (SLIPS) 1296
small mothers against decapentaplegic (Smad) 1563
smooth nanobeam 1126
– stress distribution 1126
societal 1543
– context 1544
– impact 1617
– implication 1543, 1588
– -scale platform 1595
soft
– DPN 655
– law 1551
– lithography (SL) 114, 1406
– polymer stamp 148
– stiction 1401
– stuck micromirror 1402
– substrate 1031
solar cell 285
– efficiency 286
– inorganic-organic 285
– nanowire 286
sol-gel 336, 353
solid thin film 1056
solid–lubricant interaction 1042
solution hardening 1448
solution–liquid–solid (SLS) 257
sp^3
– bonded tip 1027
– bonding 1155
– carbon 1161
spark plasma sintering (SPS) 223
spatial resolution 88
spectral reflection 1269
spectroscopy
– angle-resolved photoemission 374
– Raman 281, 312, 374, 644, 1149
– scanning tunneling 374
spherical
– micelle 346
– NP 346
– polymeric nanoconstruct 650
spin
– casting 150
– coherence time 329
– FET 329
– MOSFET 329
– -polarized scanning tunneling microscopy (SP-STM) 786
– quantization 777
spin-on-glass (SOG) 169, 1521
spintronics 328
– device 305, 321
– graphene 382
splash phenomena 1286
spring
– -shaped nanowire 262
– sheet cantilever 741
– system 727
spring constant 594, 814, 989
– calculation 746
– lateral 738
– measurement 745
– vertical 738
sputtered coatings, physical property 1153
sputtering (SP) 336, 1144
– deposition 57, 1148
– power 1154
square spiral architecture 105
squeeze flow 120
S-shaped beam 1439
stamp
– Au coated PDMS 151
– Au/Ti coated 149
– composite 146
– depth 149
– fabrication 144
– mechanical property 146
– positioning 154
– rigid 148
standard quantum limit (SQL) 422
standing wave 307, 310, 319
static
– contact angle 1288
– friction 994
– friction force 1398
– friction torque 1398
– mode 460
– mode AFM 753
– random access memory (SRAM) 405
– wetting process 1285
steady sliding 1428
steady-state sliding 1043
Stefan equation 121
stellar shape 647

STEM electron-beam-induced CVD 100
step edge 260
step-and-repeat (S&R) 131
step-and-stamp imprinting lithography (SSIL) 131
stick–slip 968, 972, 976, 977, 1428
– friction 1034
– mechanism 993
stiction 1337, 1439, 1443, 1470
– electrostatic attraction 1442
stiffness 1167
– continuous measurement (CSM) 1158
– nuclear 1091
– torsional 748
Stillinger–Weber potential 1015, 1025
stimulated emission depletion (STED) 593, 618, 837, 854
stochastic optical reconstruction microscopy (STORM) 593, 618, 837, 858
Stoney's formula 461
strain 1079, 1445
– energy difference 1157
– graphene 374
– shear (torsional) 434
– tensile (axial) 434
stray capacitance 756
streptavidin (STA) 289
– adsorbed 714
– adsorbed to silicon 713
– binding 714
– biotin wear 1388
– coating 1384
stress 1131
– corrosion cracking (SCC) 1449
– distribution 1126
– fiber 1074, 1094
– gradient measurement 1462
– intensity 1106
– intrinsic 1153
– measurement 1461
– relaxation test 1446
– residual 1153, 1160, 1459, 1471
– shear 1078
– –strain ratio 1084
– tensile 1105, 1116, 1447
– thermal 1526
stretch-sensitive ion channel 1095
Stribeck curve 966
structural
– characterization 262
– integrity 1086, 1102
– lubricity 995
structured illumination microscopy (SIM) 858
stuck micromirror 1400
subcellular structure 633, 1086
subcuticular inclusions 1281
sublattice 307
submicron particle 1390
substrate 260
– conformal imprint lithography (SCIL) 130
– curvature measurement 1447
– strain 1079
subwavelength
– grating 133
– polarizer 136
sulfur–gold bond 150
supercapacitor 384
superconductivity 782, 796
– magnetic levitation 727
– vortices 796
superconductor–normal metal–superconductor (SNS) 96
supercritical solution, rapid expansion 338, 341, 353
superhydrophilic
– pinning anchor cells 1274
– surface 1285
superhydrophilicity 1265, 1269, 1291, 1294
superhydrophobicity 1265, 1291, 1295
superimposed wax 1268
superlattice 277, 311
– nanowire 277
– structure 260
superlubricity 995, 1042
superoxide dismutase (SOD) 1562
superparamagnetic
– limit 289
– NP 646
superresolution microscopy 847
superresolution optical fluctuation imaging (SOFI) 858
supersaturation 255, 256
supported lipid bilayer 812
supramolecular
– assembly 31
– chemistry 702
surface
– acoustic wave (SAW) 544
– anti-adhesive 1293
– area 1570
– bare 1033
– charge 1573
– charge density 944
– chemistry 148, 1265
– curved 147
– diffusion 148
– energy 1142
– energy measurement 1191
– enhanced Raman scattering (SERS) 678
– -enhanced Raman spectroscopy 644
– force 1439
– force apparatus (SFA) 870, 938, 1014, 1393, 1422, 1451
– free energy measurement 1199
– friction 1033
– interaction energy 1439
– micromachining 66
– modification 269, 1573
– nanometer-scale mechanical property 1019
– oxide 264
– phonon mode 1016
– plasmon resonance (SPR) 623
– potential 944, 1146
– potential measurement 876, 1234
– profiler 1394
– protection 1180
– roughness 872, 952, 978, 1392
– rubbing/sliding 1450
– stress 459
– structure 263, 978, 1268
– structuring 1268
– temperature simulation 1026
– tension 493, 1089, 1090
– topography 988
– -to-volume ratio (S/V) 491, 644, 658
– transfer chemistry 148
surfactant 253, 259
suspended beam 1121
suspended micro-channel resonator (SMR) 604
switching
– energy 284
– speed 284
synchrotron radiation circular dichroism spectroscopy (SRCD) 1576
synthetic vesicle 1390

T

talin 1093
Tannerella forsythia 834
tapping mode (TM) 731, 811, 872, 879
– etched silicon probe (TESP) 739
μTAS 488, 490, 1344
tellurium 259
temperature
– control 1290
– dependence of friction 999
– effect 1227
template 253, 261
– assisted self-assembly (TASA) 79
– effect 1277
– manufacturing 81
tensegrity 1086
tensile 459
– stress 1105, 1116
– stress relaxation 1447
terrestrial plants 1268
test
– (ATE) development 1494
– accelerated 1530
– environment 1171
– specimen 1446
tetrahydrocurcumin (THC) 341
tetrahydrofuran (THF) 221, 341
tetrathiafulvalene (TTF) 31
theranostic 645
therapeutic 645
– device 696
therapeutic nanodevice 693
– assembly approach 700
– utility 698
therapeutics 1382
– commercial 701
thermal
– activation model, lifetime 816
– decomposition 336, 349, 353
– drift 771
– effect 998
– expansion coefficient 223, 463, 1527
– fluctuation 989, 1084
– mixing 514
– noise method 594
– oxide 54
– processing 1103
– sensor 274
– therapy 661
thermal conductance
– quantized 275
thermal conductivity 273
thermo
– electric figure-of-merit 275, 277
– electric power (TEP) 372
– electrics 275
– gravimetric analysis (TGA) 220
– lubric 999
– mechanical noise 788
– phoresis 623
– plastic polymer 499
– pneumatically actuated valve 507
– responsive NP 343
– rheology 633
– setting polymer 499
– stat 1016
theta (θ) condition 949
thin film 1028
– deposition 53
thiol 470
– monolayer, self-assembled 471
– terminated SAM 149
thiolated 821
third harmonic generation (THG) 280
third-body molecule 1039
– frictional force 1039
third-order response 860
three-dimensional (3-D) 58, 87, 114, 151, 174, 197, 249, 363, 408, 489, 559, 636, 663, 712, 725, 769, 834, 854, 1004, 1079, 1106, 1270, 1357, 1498, 1505, 1577, 1593
– force measurements 757
– wax crystals 1273
– wax platelet 1276
through
– -silicon via (TSV) 1488, 1519
– -thickness cracking 1156
– -wafer etching 1507
time-resolved emission dynamics 284
time-temperature superposition (TTS) 634
tin plant 1283
tip
– apex 1003
– coated 1031
– deformable 1019
– elasticity 771
– functionalization 599
– -growth 210
– jumping 1039
– mount 729
– preparation in UHV 990
– preparation method 730
– Pt-Ir 730
– radius effect 738, 1161, 1229
– sliding 988
tip–surface
– interaction 1036
– interface 1021
– potential 993
tissue engineering 662
titanium oxide (TiO_x) 149
T-junction configuration 538
Tomlinson model 897, 987
– finite temperature 998
top–bottom ball grid array (TB-BGA) 1508
top-down 317, 349, 578, 1615
– approach 336
– fabrication 1404
– lithographical nanopatterning 415
– method 305, 1332
topographical
– and recognition imaging (TREC) functionalization 810, 821
– imaging 731, 810
topography measurement 734
toroid structure 706
torsional resonance (TR) 879
torus model calculation 759
total internal reflection fluorescence (TIRF) 832, 857
touch-mode pressure sensor 398
touch-point pressure 398
traction force microscopy (TFM) 1080
transcription 588
– factor 1082
transducer 458
transfer
– printing process 1500
– ribonucleic acid (tRNA) 710
transforming growth factor (TGF) 1564
transient cell adhesion 1087
transistor 153, 282
– graphene 379
– high-frequency 380
– laminated n-channel 154
– single-electron 379
– structure 324
– tunneling 380
transmission electron microscopy (TEM) 75, 88, 196, 255, 263,

303, 372, 665, 726, 908, 1269, 1366, 1430, 1448, 1469, 1566
– electron-beam-induced CVD 98
– environmental chamber 265
transmission x-ray microscopy (TXM) 1577
transport
– ballistic 267
– diffusive 267
– mechanics of nanoconstruct 649
– of ions in matter (TRIM) 94
– properties electrical 267
– properties nanowire 267
transversal piezo-resistive effect 755
transverse magnification 846
trapping site 1442
triangular cantilever 739, 748, 754
tribo
– -apparatus 1450
– charging 1442
– chemical reaction mechanism 1396
– chemistry 1041
– meter 1353
– polymer (TP) 1431
tribological
– C_{60} 1051
– characterization of coating 1154
– performance of coating 1168
trichome 1283
tricresyl phosphate (TCP) 1050
trimethylaluminum (TMA) 181
trinitrotoluene (TNT) 472
tubular nanoparticle 657
tumor 596
– targeting 650
tungsten 730
– sphere 989
– tip 730
tuning fork AFM 992
tunnel field-effect transistor (TFET) 325
tunneling
– current 727
– detector 731
turbo-molecular pump (TMP) 99
turgor pressure 834
turn-on voltage 269
tweezers, optical (OT) 626, 1079
two phase flow 495
two-dimensional (2-D) 88, 134, 182, 277, 303, 330, 363, 432, 503, 563, 782, 854, 882, 995, 1088, 1106, 1273, 1357, 1404, 1418, 1516
– electron system (2-DES) 782
two-photon
– 4-Pi fluorescence microscopy 854
– emission 860
type-2 diabetes 832

U

ultra high frequency (UHF) 419
ultra large scale integration (ULSI) 451
ultrafine particle (UFP) 1560
ultrahigh vacuum (UHV) 319, 564, 733, 772, 990, 1019
– -AVM 991
ultralow limit of analyte detection 661
ultra-low power consumption 449
ultrananocrystalline diamond (UNCD) 178, 1451
ultrathin DLC coating 1144
ultraviolet (UV) 52, 113, 175, 500, 820, 1266, 1404
– exposure 123
– light 285
– reflection 1290
Umklapp process 274
unfolded protein response (UPR) 1574
unified modeling language (UML) 697, 699
unit cell 1085
unloading curve 1023
unnatural AA (UAA) 709
unpaired electron 306

V

vacancy
– chain 1597
– diffusion 1445
van der Waals (vdW) 565
– attraction 743
– force 901, 942, 1390, 1392, 1439
– interaction 1391
– surface 789
Van Hove singularity 215
vapor 936
– deposition 336, 348, 353
– -grown carbon nanofilament (VGCNF) 199
– phase cocondensation 150
– -phase etching (VPE) 61
– –solid–solid (VSS) 257
vapor–liquid–solid (VLS) 210, 251, 255, 269, 288
– mechanism 255
vapor-phase epitaxy (VPE) 56
– organic (OMVPE) 1407
variable force mode 761
variable heavy–heavy (VHH) 712
vascular endothelial (VE) 825
– growth factor (VEGF) 352, 660
vascular plants 1270, 1279
velocity
– -dependent friction model 973
– effect 1225
– rescaling 1017
– strengthening of friction 999
– weakening 999
vertical
– -cavity surface-emitting laser (VCSEL) 135, 465
– coupling 729
– manipulation 793
– nanotube 1055
– nanowire array integrated nanogenerator (VING) 287
vertical-cavity surface-emitting laser (VCSEL) 135
very large-scale integration (VLSI) 52
very low density lipoprotein (VLDL) 345
vibrometry 1438
Vickers indentation 1104, 1109
vinculin binding site (VBS) 1093
virus-like particle 345
viscoelastic
– mapping 878
– solid 1083
viscoelasticity mapping 917
vital-sign monitoring (VSM) 1494
volatile organic compound (VOC) 468, 471
voltage-controlled oscillator (VCO) 409
vortex 796
V-shaped cantilever 740, 989

W

W tip 730
wafer acceptance test (WAT) 1494

wafer-level chip-scale packaging (WLCSP) 1492, 1495
Washburn equation 253
water
– loss reduction 1291
– plants 1297
– -repellent surface 1285
– treatment 335, 353
– vapor content 1171
wave interference 307
waveguiding 289
– nanowire 285
wax
– aggregate 1274
– chemical composition 1271
– compound 1272
– crystal 1267, 1274
– crystal morphology 1271
– ester 1274
– intracuticular 1268
– platelet 1274
– rodlet 1274
– self-assembly 1275
– superimposed 1268
– tubule 1274
wear 1013, 1231, 1425
– damage 1171
– depth 716
– detection 1237, 1245
– fullerene 1052
– map 1384
– mark 1400
– measurement 725, 1470
– of biomolecule 1382
– process 1168
– resistance 1167, 1170, 1196
– test 716, 1169, 1365
– tip 1039
– track 1034
– tribopolymer 1430
wearless friction 1041
Weibull distribution function 1529
wet etching 59
wettability 1268, 1275, 1283, 1291, 1360
wetting
– behavior 1285
– surface 1019
Weyl equation 307
Wheatstone bridge 466, 755
wire
– cantilever 738
– grating polarizer 134
– grid polarizer 132
woodpiles architecture 105
worm-like micelle 349

X

x-ray
– absorption fine structure (XAFS) 1576
– absorption near-edge structure (XANES) 1568, 1577
– diffraction (XRD) 253, 265, 1275
– fluorescence (XRF) 1577
– imaging 1453
– lithography 1404
– photoelectron spectroscopy (XPS) 502, 1143, 1190, 1244, 1254

Y

yield point 957
yield strength simulation 1025
Young–Dupre equation 1193
Young–Laplace equation 493
Young's modulus 91, 127, 461, 568, 595, 654, 745, 761, 834, 1000, 1072, 1394, 1459, 1464

Z

Z-DOL (fully bonded) 1360
zero-dimensional (0-D) 363, 564
zero-mode waveguide (ZMW) 517
ZGNR 321
– half-metallic 328
zigzag
– edge 304
– -edged nanographene 306
– -type nanotube 195
ZnO
– nanowire 261, 266, 284, 288

Recently Published Springer Handbooks

Springer Handbook of Nanotechnology (4th) (2017)
ed. by Bhushan, 1698 p., 978-3-662-54355-9

Springer Handbook of Electronic and Photonic Materials (2nd) (2017)
ed. by Kasap, Capper, 1450 p., 978-3-319-48931-5

Springer Handbook of Global Navigation Satellite Systems (2017)
ed. by Teunissen, Montenbruck, 1328 p., 978-3-319-42926-7

Springer Handbook of Model-Based Science (2017)
ed. by Magnani, Bertolotti, 1157 p., 978-3-319-30525-7

Springer Handbook of Odor (2017)
ed. by Buettner, 1151 p., 978-3-319-26930-6

Springer Handbook of Electrochemical Energy (2017)
ed. by Breitkopf, Swider-Lyons, 1016 p., 978-3-662-46656-8

Springer Handbook of Robotics (2nd) (2016)
ed. by Siciliano, Khatib, 2227p., 978-3-319-32550-7

Springer Handbook of Ocean Engineering (2016)
ed. by Dhanak, Xiros, 1345 p., 978-3-319-32550-7

Springer Handbook of Computational Intelligence (2015)
ed. by Kacprzyk, Pedrycz, 1633 p., 978-3-662-43505-2

Springer Handbook of Marine Biotechnology (2015)
ed. by Kim, 1512 p., 978-3-642-53970-1

Springer Handbook of Acoustics (2nd) (2015)
ed. by Rossing, 1286 p., 978-1-4939-0754-0

Springer Handbook of Spacetime (2014)
ed. by Ashtekar, Petkov, 887 p., 978-3-642-41991-1

Springer Handbook of Bio-/Neuro-Informatics (2014)
ed. by Kasabov, 1230 p., 978-3-642-30573-3

Springer Handbook of Nanomaterials (2013)
ed. by Vajtai, 1222 p., 978-3-642-20594-1

Springer Handbook of Lasers and Optics (2nd) (2012)
ed. by Träger, 1694 p., 978-3-642-19408-5

Springer Handbook of Geographic Information (2012)
ed. by Kresse, Danko, 1120 p., 978-3-540-72678-4

Springer Handbook of Medical Technology (2011)
ed. by Kramme, Hoffmann, Pozos, 1500 p., 978-3-540-74657-7

Springer Handbook of Metrology and Testing (2nd) (2011)
ed. by Czichos, Saito, Smith, 1229 p., 978-3-642-16640-2

Springer Handbook of Crystal Growth (2010)
ed. by Dhanaraj, Byrappa, Prasad, Dudley, 1816 p., 978-3-540-74182-4

Springer Handbook of Automation (2009)
ed. by Nof, 1812 p., 978-3-540-78830-0

Springer Handbook of Mechanical Engineering (2009)
ed. by Grote, Antonsson, 1576 p., 978-3-540-49131-6

Springer Handbook of Experimental Solid Mechanics (2008)
ed. by Sharpe, 1096 p., 978-0-387-26883-5

Springer Handbook of Speech Processing (2007)
ed. by Benesty, Sondhi, Huang, 1176 p., 978-3-540-49125-5

Springer Handbook of Experimental Fluid Mechanics (2007)
ed. by Tropea, Yarin, Foss, 1557 p., 978-3-540-25141-5

Springer Handbook of Engineering Statistics (2006)
ed. by Pham, 1120 p., 978-1-85233-806-0

Springer Handbook of Atomic, Molecular, and Optical Physics (2nd) (2005)
ed. by Drake, 1506 p., 978-0-387-20802-2